The Ontario Universities' Teaching Evaluation Toolkit:

Feasibility Study

A Ministry of Training, Colleges, and Universities Productivity and Innovation Fund Initiative

June 2014

Acknowledgements

We gratefully acknowledge the contributions of many people from the University of Windsor, Carleton University, Queen's University, and the University of Ontario Institute of Technology.

University of Windsor

Project Champion: Alan Wright

Project Lead: Beverley Hamilton

Joe Branget
Ken Cramer
Graham Fawcett
Michele Gnanamuttu
Phil Graniero
Betsy Keating
Erika Kustra
Peter Marval
Abby Nakhaie
Ashlyne O'Neil
Bethany Paper
Jennifer Patterson
Lisa Plant
Jessica Raffoul
Behdokht Rekabdar
Jennifer Stanton
Jennifer Stefanczyk

Carleton University

Project Lead: Joy Mighty

Julie Gregory
Patrick Lyons
Rhonda Newhook
Laura Ravelo

Queen's University

Project Lead: Jill Scott

Rylan Egan
Marcia Ingersoll
Stefan Merchant
Denise Stockley

University of Ontario Institute of Technology

Project Lead: Bill Muirhead

Kristen Boujos
Christine Cairns
Jaymie Koroluk
Dana Martin
Mark Naylor
Marj Rempel

External Consultants

Roger Bryan, KPMG
Gordon Joughin, Queensland University
Laura Winer, McGill University

For more information, contact Alan Wright at awright@uwindsor.ca

University of Windsor

Table of Contents

Executive Summary

This feasibility study (the first of three phases) sought to develop a framework for improvement-oriented formative and summative assessment of teaching in Ontario. It is intended to inform future developments in teaching evaluation in the Province, and to offer a well-contextualized understanding of what the goals of teaching evaluation ought to be, what the challenges are, and the kinds of initiatives and infrastructure that would best promote the evolution of a data-informed and inquiry-inspiring approach to evaluating and improving teaching.

Our institutionally-based project teams identified and examined leading teaching evaluation practices in use internationally, compared to those in use in the Ontario context, and identified a range of aggregate data and technical tool elements to be considered when moving forward.

Our review of effective teaching evaluation practices internationally identified the following recurring themes:

- The users of an evaluation system must have a shared understanding of quality
- Evaluations must be multi-faceted, using multiple types of data, approaches to gathering data, and methods for evaluating data
- Effective processes must have robust feedback cycles that are integrated into evaluation and instructional improvement programs
- The process of establishing and implementing a teaching evaluation system requires sustained, multi-level, consultative leadership
- Communications and dialogue are critical to the engagement of key stakeholders as is the adoption of change practices proven to be effective in complex systems

The Ontario teaching evaluation system exhibits a number of strengths, but also a significant number of gaps when compared to the recurring themes noted above. There are significant and regulated variations in teaching evaluation processes in Ontario, often bound by labour agreements and senate-approved policies. Because of the sensitivity of the information involved, tools created for provincial use must be customizable for individual institutions. Common standards should be established collectively: voluntary adoption of efficient tools is more likely to effect the desired change than mandatory imposition of a method.

While the study involves a design for a suggested suite of online tools to support teaching evaluation, the report concludes that the more critical aspects of planning and development lie in the procedural, cultural, and collaborative changes that would be required to inspire truly improvement-oriented approaches to evaluation across Ontario. This is more critically a challenge of inspiring engagement and creating a strategy for change management, rather than a primarily technical or tool-oriented exercise, though the tools can ultimately contribute to significantly better practice.

The report recommends starting with the establishment of a teaching evaluation consortium to promote effective teaching evaluation in the province of Ontario, including policies, practices, and principles as well as leadership for research and tool development. A first phase of projects for the consortium would include:

- developing and piloting a guided electronic teaching dossier prototype;
- a Ministry call for coordinated projects to enhance integrated evaluation and improvement and practice; and
- the development and piloting of tools that enable better analysis and visualization of SRI data by individual instructors, and at the departmental level.

I

Project Overview

Context

In 2012, the Auditor General of Ontario (AGO) undertook an audit intended to assess the "extent to which the Ministry and selected Ontario universities support, assess, and periodically report meaningful performance information on the quality of instruction provided to undergraduate students" (AGO, 2012, p. 274). The results indicated that while all Ontario institutions gather student ratings of instruction (SRI), the nature, methods, and use of these data vary considerably among institutions, informed by disparate labour agreements, policy contexts, and institutional cultures. The audit also identified common gaps in practice, including:

- reliance on SRI as single-point performance data;
- limited consistency in the SRI data gathered even within individual institutions;
- insufficient feedback on performance, and guidance regarding the use and interpretation of SRI data
- insufficient guidance regarding the use and interpretion of the results of SRI;
- some degree of inconsistency in the use of the SRI data for performance appraisal and personnel-related decision-making; and
- limited use of the SRI data for analytical and strategic purposes, or for planning of professional development, particularly in aggregate form.

Following the findings of this audit and the growing emphasis on quality enhancement and assessment in the university sector, the Ministry of Training, Colleges and Universities, in its July 2013 Productivity and Innovation Fund call sought a proposal for a project that would inform the development of a framework for "collecting, managing, and analyzing appropriate data to develop and strengthen the cycle of continuous improvement of teaching quality" in Ontario. The University of Windsor, with Carleton University, University of Ontario Institute of Technology (UOIT), and Queen's University, submitted the successful proposal, and the team was engaged to produce a feasibility study regarding a potential framework. This report provides the findings and recommendations that emerge from the study.

Purpose

The goal of this project, designed as a three-phase undertaking, is to develop a framework for improvement-oriented formative and summative assessment of teaching in Ontario. Ultimately, this framework will include the development of policies, a suite of online tools, and textual materials to support teaching assessment and improvement, framework implementation guides, as well as a core of provincial expertise, collaboration, and information that will position Ontario as a leader in evidence-driven teaching improvement practice.

The core of Phase 1 is the feasibility study, which seeks to answer the following questions:

- What is the case for producing a standard but

customizable suite of tools for the evaluation and improvement of teaching in Ontario Universities?
- What kind of teaching evaluation tools are most likely to facilitate the efficient, effective improvement of teaching in Ontario, and the engagement of faculty and administrators in teaching improvement?
- What are the basic requirements and technical specifications of such a system based on identified needs, the current regulatory climate, and identified best practices? What are further desirable elements?
- What are the conditions required for the adoption of a provincial framework and teaching evaluation tool suite in Ontario?

Deliverables

Phase 1 of this study includes the following deliverables:

- an environmental scan of current teaching evaluation practice and policy in Ontario universities;
- a report highlighting examples of national and international best practices in this area, including an assessment of what might be feasible and adaptable to the Ontario context;
- an analysis of the relative merits of third-party and custom-designed student ratings of instruction instruments;
- summary documentation from stakeholder consultation identifying challenges, opportunities, requirements, concerns, and possible approaches to full implementation;
- a recommended conceptual model and preliminary technical modeling for a suite of provincial teaching evaluation tools;
- a specifications document detailing an approach to the further development of a viable model, modeling some of the technical issues and design around tool development, in order to inform the second phase of tool and prototype development and testing across member institutions in Phase 2 and 3; and
- a conceptual model exploring how institutional aggregation of data can be shared amongst communities of interest. This conceptual model includes how SRI data analytical and visualization tools could be developed and tested.

II

Project Methodology

Researching the Case for Change

Phase I was undertaken by four teams, each housed at one institution: an effective practices team, an environmental scan team, a data team, and a technical planning team. Figure 1 provides an overview of the project plan. Mandates for each of the project teams can be found in Appendix A.

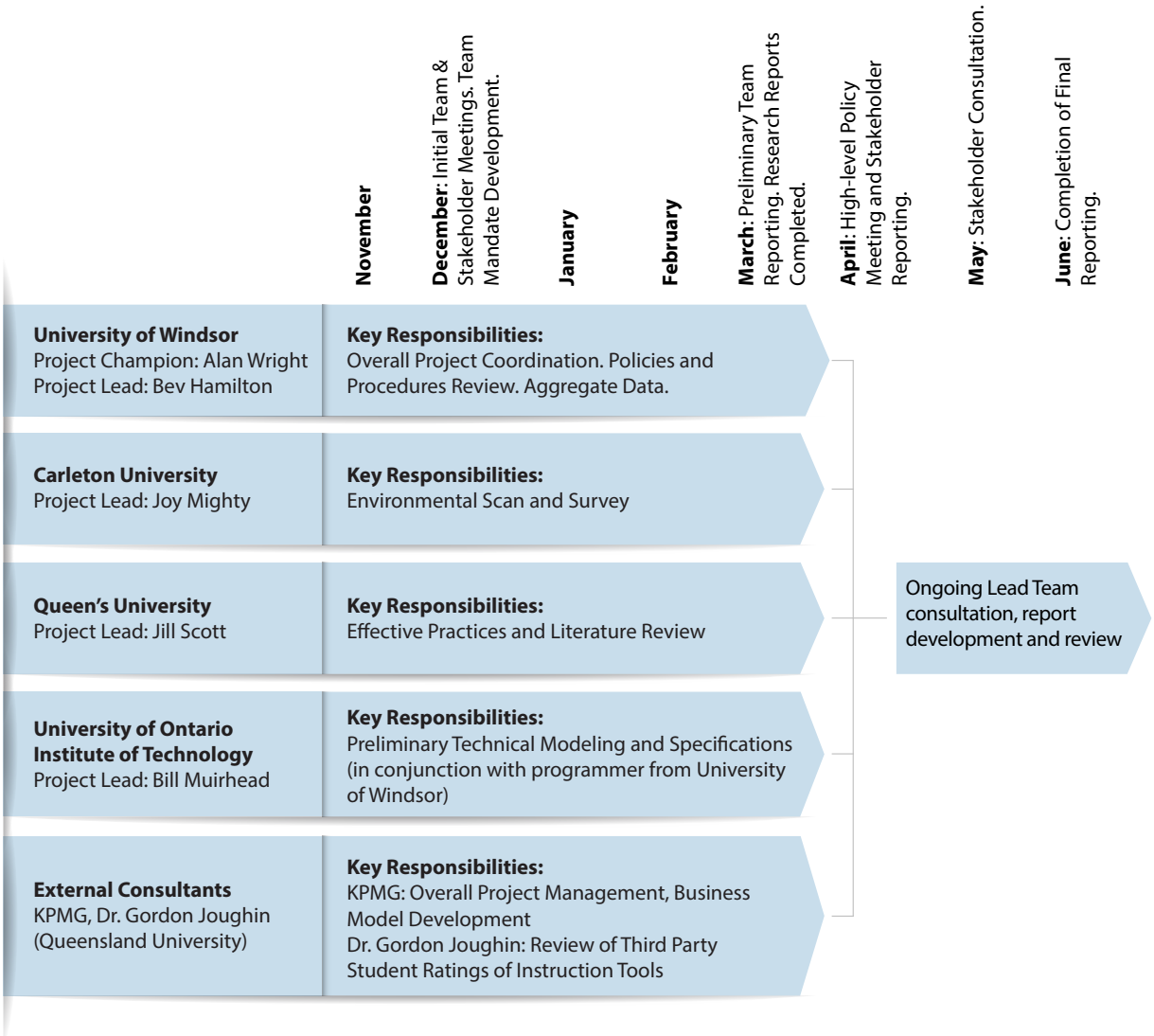


Figure 1: Project Methodology – An Overview

One unique feature of this project is the nature of the lead team, which is comprised of associate vice-presidents and vice-provosts, teaching and learning. This role is a relatively recent evolution in the Canadian context. At the institutional level, these positions facilitate a degree of integrated and systemic strategic planning with relation to teaching improvement that, heretofore, has been difficult to achieve (Wright, Mighty, Dyens, & Rogers, 2013). This project, one of the first extended inter-institutional initiatives undertaken by individuals in these positions, reflects their unique expertise and experience: leading institutional change, educational policy development and implementation, consensus building and stakeholder engagement, teaching improvement, institutional governance, and the implementation of technology in educational contexts. This background has been critical to the approaches taken in the project. Collectively, this team has decades of experience in the leadership of institutional change and the implementation of province-wide policy initiatives on the ground. The overall process drew on this experience in identifying core practices and assessing their implications, opportunities, and risks against institutional contexts.

> Our goal in pursuing this feasibility study has been to identify a pathway that will facilitate the adoption of demonstrably effective teaching evaluation practice at Ontario institutions, with the understanding that tools must be embedded in data-informed and improvement-oriented cultures in order to function transformatively.

Methodological Lens

This project has been approached with a fundamental orientation towards investment that, while documenting teaching quality, emphasizes improvement. Biggs (2001) distinguishes between "retrospective" and "prospective" quality assurance approaches. The latter captures our mission:

> **Retrospective QA** [quality assurance] looks back at what has already been done and makes a summative judgment against external standards. The agenda is managerial rather than academic, with accountability a high priority; procedures are top-down, and bureaucratic. This approach…is despite the rhetoric, not functionally concerned with the quality of teaching and learning, but with quantifying some of the presumed indicators of good teaching and good management, and coming to some kind of cost-benefits decision.
>
> **Prospective QA** is concerned with assuring that teaching and learning does now, and in future will continue to fit the purpose of the institution. It also encourages continuing upgrading and improvement of teaching through quality enhancement….While the proponents of retrospective QA talk as if they are concerned with educational quality in the sense of "fit for the purpose," the procedures adopted address "value for money," and are frequently counter-productive in the sense of providing rich teaching contexts and enhanced learning outcomes. (p. 222)

While universities can and should apply themselves to the process of demonstrating to the public that their teaching practices are consistent with accepted standards of practice, this project is an opportunity to approach the issue of quality, and continuous growth and development, as thoughtfully as possible. What do we mean by "quality" in teaching? Biggs (2001) argues that an approach that takes account of the systemic nature of educational quality, one that cannot easily be fragmented into discrete numerical components, asks the following:

- What theory of learning drives our practice?
- What mechanisms allow us to continually review and improve current practice?
- What can be done to remove impediments to quality teaching?

In this case, one would argue that quality emerges when an institution "has high level aims that it intends to meet, that teaches accordingly, and that continually upgrades

its practices in order to adapt to changing conditions within resource limitations" (p. 223). Ultimately, quality assurance, which includes teaching evaluation, must proceed based on clear statements of the intended character and qualities of an institution's educational practice, on the establishment of environments and activities that produce student learning, and on the development of effective assessment mechanisms that align accurately with the institution's stated outcomes.

This does not diminish the need for quality assurance that provides evidence of consistency with established mandates, of effective teaching, and an ongoing commitment to the improvement of learning. These are important benchmarks: this documentation, however, will not assure quality or a pathway to continuing and evolving quality at the institutional level, as these must emerge from a more internalized and systemic commitment to defining, striving for, and measuring programme impact and quality. Our goal in pursuing this feasibility study has been to identify a pathway that will facilitate the adoption of demonstrably effective teaching evaluation practice at Ontario institutions, with the understanding that tools must be embedded in data-informed and improvement-oriented cultures in order to function transformatively. The feasibility study therefore offers plans for both an evidence-based suite of tools, and a recommended approach to engaging Ontario universities and their stakeholders with an improvement-oriented teaching evaluation agenda in the coming years.

Clarifying Initial Terminology

Teaching

Vajoczki (2008) provides a usefully broad definition of teaching: "any activity which manipulates a student's environment in order to facilitate learning or behaviour change" (p. 5). The Canadian Association of University Teachers' *Model Clause on the Evaluation of Teaching Performance* (2007) suggests the following categories of activity:

- Giving courses; conducting seminars; guiding tutorials, laboratories, and studio work; supervising fieldwork, coaching and individual study projects;
- Preparing, grading, and correcting assignments, tests, and examinations;
- Guiding the work of teaching assistants, graders, markers and laboratory instructors;
- Guiding and evaluating students' individual work, such as theses and papers;
- Consultations with students outside of class or laboratory time;
- Participating in the development of teaching methods, programmes, or course content;
- Preparing course outlines, instructional material, laboratory exercises, and course notes; and
- Writing textbooks: textbooks may also be considered when evaluating a member's scholarship.

> [Teaching] includes the aims of a course, the methods of presenting the knowledge those aims embody, assessing students' achievement, and evaluating the effectiveness of the whole process. Professional teachers in higher education display certain salient characteristics. They possess a broad range of specialist teaching skills; they never lose sight of the primacy of their goals for student learning; they listen to and learn from their students; they constantly evaluate their own performance. They understand that teaching is about making it possible for students to learn; they succeed in integrating educational theory and shrewd classroom knowledge.
>
> (Ramsden, 2003)

While these provide a clear and well-considered description of activities relating specifically to the teaching of students, two further categories might reflect activities that many academics undertake in the interests of teaching, particularly as they evolve as educators:

- educational leadership through for example professional development (both pursuing and teaching); and
- the pursuit of educational innovation.

If we consider the traditional tripartite balance of academic practice (teaching, research, and service), both of these further criteria might fall into all or none of these categories, but they are critical contributions to the educational landscape and in particular to the capacity of institutions to evolve and change. These criteria are typically found as well in the teaching excellence criteria of teaching awards such as the well-established Ontario Confederation of University Faculty Associations (OCUFA) teaching excellence award and the prestigious 3M National Teaching Fellowships. Other researchers have extended the conception of teaching to capture the sense that teaching is not simply a set of practices and behaviours: Ramsden et al. (1995) characterize teaching as a reflective and intellectual practice requiring dynamic response to changing conditions and evolving student needs and characteristics. Further, teaching is not a static practice: the use of technology and its capabilities and capacities' are transforming the role of faculty, students, content and how interaction is occurring within learning environments, both face-to-face and online. Effective teaching is multi-faceted, responsive to context, and impacted by disciplinary differences as well as environment, instructors' goals, beliefs about teaching, strengths and preferences: there are many ways to be an effective teacher (Ramsden, 2003).

Teaching Evaluation

Though teaching evaluation is often confused with the process of gathering data about teaching, it is in fact a multi-stage process. Wolf (1987) identifies four main elements of evaluation:

- the systematic and thoughtful collection of data;
- a process through which these data are interpreted;
- the judgment of value; and
- a plan for action based on the first three steps.

Effective Teaching Evaluation Practice

For the purposes of this review, "effective practice" refers to faculty evaluation processes that allow institutions and instructors to meet their goals in ways that balance accuracy, cost, and other considerations such as legal or contract issues. Vajoczki (2008) identifies the provision of feedback to instructors about their teaching for improvement purposes, personnel-related decision-making, provision of information to students to inform course and instructor selection, curriculum and course re-design, and teaching research, as the most commonly acknowledged purposes identified for evaluating teaching.

> Effective practice in evaluation, while drawing from common practices and principles, should be understood as context specific, and that context must be explicitly articulated in a consultative fashion if the system is to function effectively.

For the purposes of hiring, promotion, and tenure, effective teaching evaluation should give robust, reliable information about the faculty member's effectiveness as a teacher. However, if the intent of evaluation is also the improvement of teaching quality, then planning and practices associated with teaching improvement must be incorporated and emphasized in its overall design. A wide range of stakeholders participate in and are impacted by the outcomes of teaching evaluation – instructors, administrators, institutional quality assurance units, faculty associations, parents, employers, the government. These groups have varying and sometimes conflicting interests and concerns within the overall evaluation system (Vajoczki, 2008).

Arreola (2007) argues, speaking of faculty evaluation more generally, the critical question of context – of institutional values and needs: it "involves the systematic observation of relevant faculty performance to determine the degree to which that performance is consonant with the values and needs of the educational institution" (p. xix). In other words, effective practice in evaluation, while

drawing from common practices and principles, should be understood as context specific, and that context must be explicitly articulated in a consultative fashion if the system is to function effectively.

Student Ratings of Instruction

"Student ratings of instruction" is one of a number of terms used to describe the instruments used to gather student feedback about courses or instructors in universities. They go by a variety of names, for example "student evaluation of teaching," "course evaluations," or "teaching evaluation forms." It is important not to assume that these ratings or evaluations are entirely directed towards teaching performance: in fact some of the data gathered focus on elements of course offerings. However, these terms are often confounded, and used interchangeably, perhaps reflecting an overall lack of precision in the planning and development of these procedures in universities. Some authors prefer the term ratings of instruction because they feel that students are not in a position to offer a complete evaluation of the many facets of instruction (Berk, 2014; Pallett, 2006), and student surveys do not form an entire evaluation but are instead one piece of information that is considered during the evaluation process. We agree with this position and will use the term student ratings of instruction throughout this review.

Assessment and Evaluation

Assessment is defined as data-gathering strategies, analyses, and reporting processes that provide information that can be used to determine whether or not intended outcomes are achieved. Evaluation uses assessment information to support decision-making, in this case with regard to personnel issues and instructional improvement (Foundation Coalition, 2001). These terms are often conflated, potentially reflecting limited engagement with the complexities of the evaluation construct. It should be noted that these terms are used quite differently in different jurisdictions. In the UK, for example, the term assessment is used to refer to marking and grading of student work and evaluation is concerned with judging the quality, impact, and effectiveness of programmes of study, teaching, and the student experience in general (L. Anderson, personal correspondence, June 6, 2014).

III

A Review of Effective Teaching Evaluation Practices Internationally

The effective practices team identified four major recurring themes in the research on effective teaching evaluation:

- Shared understandings of quality
- Multi-faceted data and evaluation
- Robust feedback cycles: integration of evaluation and instructional improvement programmes
- Sustained leadership for education, engagement, and change.

> Above all, measurement is an act of selection. Our senses and information systems select but a tiny fraction of possible experience…. Through our mental models, we define constructs… and design systems to evaluate and report them. We conflate what is salient, tangible, and familiar with what is important. As we measure these things, they become even more real, whereas the remote effects of our decision, the unfamiliar and the intangible fade like wraiths…Often the mutual feedback of expectations and perception blinds us to the anomalies that might challenge our mental models and lead to deep insight.
>
> (Sterman, 2006, p. 510)

Effective Teaching Evaluation: Recurring Themes

Shared Understandings of Quality

In order for evaluation to take place – that is, the process of systematically collecting data, thoughtfully interpreting the data, making judgments of value, and creating action plans based on the first three – all of those involved must work from shared and explicitly articulated understandings of effective teaching and quality (Arreola, 2007; Hénard & Roseveare, 2012; OECD, 2007; Gravestock, 2011; Gravestock & Gregor-Greenleaf, 2008). These definitions must be contextual, evolving, and periodically reviewed (Devlin & Samarawickrema, 2010; Hénard, 2009; Hénard & Roseveare, 2012); and faculty must be involved in determining these shared understandings (Arreola, 2007; Berk, 2006) if the system is to gain wide and thoughtful acceptance, and particularly if teaching improvement is a core goal of the initiative.

There is considerable and relatively convergent research regarding characteristics of effective teaching which can form the basis of institutional dialogue (Braskamp & Ory, 1994; Chickering & Gamson, 1987; Feldman, 1989, 2007; Hativa, 2013a; Hativa, Barak, & Simhi, 2010; Murray, 1997). However, numerous scholars (Colbeck, 2002; Gravestock, 2011; Pratt, 1997; Ramsden, 2003) argue that definitions of teaching must foreground motivation and rationale as core elements of how faculty approach teaching in order to balance a more traditional emphasis on discrete actions and behaviours as the basis for understanding teaching effectiveness. In addition, there is a solid tradition advocating for the adoption of a holistic framework for evaluating academic work, where teaching, research, and service function as integrated areas of scholarly practice (Boyer, 1990). Institutional exploration of these fundamental approaches is critical to the establishment of evaluation programmes that reward thoughtful and growth-oriented teaching.

Multi-Faceted Evaluation

The vast preponderance of research on effective practice in teaching evaluation indicates that multi-faceted evaluation is more effective than student ratings of instruction alone (Arreola, 2007; Berk, 2009, 2013; Buller, 2012; Ghedin & Aquario, 2008; Hassna & Raza, 2011; McLean et al., 2008; Weschke & Canipe, 2010; Zakrajsek, 2006), particularly for summative decision-making. Multiple types and sources of evidence, and multiple instances of single forms of evidence, are especially critical because of the breadth and complexity of educational practice.

First, no one group of individuals (students, peers, instructor) can respond with accuracy to the range of questions and practices involved in effective teaching (Berk, 2014). It is well established that students, for example, are not knowledgeable enough to assess the currency or relevance of course topics and readings, instructor competence within the discipline, or course purpose (Table 1). Peers are not in a position to review an

Table 1: SRIs: What and What Not to Ask Students (Hativa, 2013b)

Items to Include	Items and Practices to Avoid
Course difficulty/workload	Questions unrelated to classroom teaching performance
Course organization, structure, and objectives	Course content and purposes
Fairness of grading/assessment	Suitability of selected readings
Overall rating of the course	Poorly phrased items
Overall value/quality of the course	Poorly scaled response options
Student recommendation of the course or instructor	Questions asking students to compare the instructor with other instructors
Instructor's overall effectiveness	Instructor's competency in his/her discipline
Instructor's clarity/communication skills	Instructor's knowledge of the material
Instructor's ability to inspire interest in the subject matter	Questions concerning campus values and purposes
Instructor's concern for student learning	Questions likely to be endorsed by almost all or almost none of the students
Customized items including the use of course website, in-class technologies	Ambiguous items
Student demographics, which may include year of study, major, expected grade, etc.	

instructor's rapport with students across the wide range of settings where instructors and students interact (emails, office hours, labs), or to understand how students perceive the instructor. Instructors are generally likely to have blind spots about their own practices (Centra, 2000).

Further, every source of data related to such a system has limitations, potential for bias, potential for manipulation (Centra, 1977), but also potential benefits. As Berk (2014) puts it: "Each source can provide unique information, but…also is fallible, usually in ways that are different from the other sources….What should you do? Draw on three or more different sources of evidence. The strengths of each source can buffer the weaknesses of the other sources, thereby converging on a decision about teaching effectiveness that is more accurate, reliable, equitable, and comprehensive than one based on any single source…this notion of triangulation is derived from a compensatory model of decision-making" (p. 88).

Teaching is messy practice: classrooms are complex human systems (Doll, 1993) and the actual effectiveness of teaching, its capacity to foster student learning, is not easily untangled from context, convention, and audience.

Teaching is messy practice: classrooms are complex human systems (Doll, 1993) and the actual effectiveness of teaching, its capacity to foster student learning, is not easily untangled from context, convention, and audience. Complex systems tend not to lend themselves to the generation of reliable, reproducible evidence (Sterman, 2006), or to the ready and direct uptake of the implications of evidence. It is therefore necessary to approach the challenge of evaluation through the collection of multiple types of data, a fundamental collective commitment to critical inquiry, and by examining issues from multiple perspectives.

A multi-faceted approach also has a better chance at teasing out the contextual and structural factors that impact teaching performance. Many aspects of teaching practice are not individually determined, but are shaped by the value system, reward structures, practices, decision-making, and regulatory environments of the broader institution (Hénard & Roseveare, 2012, Sterman, 2006; Arreola, 2007). In some cases, and without absolving individuals of responsibility for their actions, teaching quality improvement may require a broader focus in order to better see the systemic levers and tensions involved in current practice (Hénard & Roseveare, 2012). As Sachs (2012) puts it, multi-faceted data and evaluative practice allow for the exploration of "how structures, policies and practices are aligned and how they contribute to teaching quality" (p. 6).

Finally, the use of multiple data sources is critical to the credibility of teaching evaluation practice in universities. As numerous studies indicate, there is a widespread perception of the insufficiency of current teaching evaluation practices in universities and of an over-reliance on single source data from SRIs (Gravestock, 2011). These perceptions serve as opportunities for instructors and others to dismiss the implications of feedback. As Sterman (2006) demonstrates, resistance to policy change is a deep-seated challenge in complex systems, and perceptions that the data are insufficient for the decisions made with them exacerbates resistance: "Unless able to assess the reliability of evidence about complex issues on their own, and frequently excluded from the policy process, citizen noncompliance, and active resistance grow" (p. 506).

Recommended data sets vary, but may include:

- SRI data;
- peer observations (Berk, 2009; Chism, 2007; Devanas, 2006; Weschke & Canipe, 2010);
- peer review of course documentation, including course outlines, assignments, course handouts, etc. (Arreola, 2007; Devanas, 2006);
- self-evaluations which focus on reflective teaching orientation and focus (e.g., Teaching Perspectives Inventory, Pratt, 1998);
- self-evaluations which enable instructors to compare what they believe they are doing with what students perceive them to be doing (e.g., CLASSE; Ouimet & Smallwood, 2005);

- video recordings for review (e.g., Performance Assessment for California Teachers, n.d.);
- samples of student work;
- student focus group data;
- curriculum materials; and
- student performance data (Pratt, 1997).

Multi-dimensional data is fundamental to effective teaching evaluation: teaching dossiers are the most effective way to thoughtfully and systematically integrate and represent those data.

Robust Feedback Cycles: Integration of Evaluation and Instructional Improvement Programmes

The feedback cycle is critical to the effectiveness of evaluation as a component of teaching quality enhancement (Arreola, 2007; Gibbs & Coffey, 2004; Knol, Veld, Vorst, Driel, & Mellenbergh, 2013; Lang & Kersting, 2007; McLean et al., 2008; Piccinin, 2003; Theall & Franklin, 2001; Harvey, 2011; Wininger & Birkholz, 2013). When re-cast in terms of solution-focused goals, formative feedback from multiple evaluation efforts can successfully influence teaching effectiveness (Devlin & Samarawickrema, 2010). Teaching evaluation processes are an essential element of teaching improvement, provided they are integrated into iterative cycles of analysis, reporting, action, and feedback (Harvey, 2011), a pattern that also enhances student satisfaction and engagement.

Teaching evaluation is more effective if integrated in a highly intentional manner with professional development and instructional improvement programming (Arreola, 2007; Steinert et al., 2006). Short training courses have limited impact on practice whereas sustained in situ training, peer consulting, student assessments, and intensive programmes can lead to improvement in the quality of teaching and student learning (Gibbs & Coffey, 2004; Prebble et al., 2004). Faculty development programmes have a role to play in the feedback and improvement cycle and include the following key features: the use of experiential learning, provision of feedback, effective peer and colleague relationships, well-designed interventions following principles of teaching and learning, and the use of a diversity of educational methods within single interventions (Steinert et al., 2006).

The value ascribed to teaching is affected by institutional culture, and faculty uptake of training and improvement of teaching can be positively influenced by a strong teaching development culture (Gibbs & Coffey, 2004; Healey, 2000; Knight & Trowler, 2000; McLean et al., 2008; Richardson, 2005). SRI data require contextualization: materials can include interpretive guides, comparative means, written narratives by faculty members, and consultations with administrators, peers, mentors, or educational developers (Buller, 2012; Clayton, 2012; Gravestock & Gregor-Greenleaf, 2008; Hativa, 2013a). Ideally teaching evaluation items should also align with institutional and sectoral understandings of quality and learning outcomes as determined by external regulatory bodies and other agencies (Hénard, 2009).

Sustained Leadership for Education, Engagement, and Change

Leadership and engagement are key to effective evaluation and in particular to the implementation of integrated teaching evaluation systems. As Hénard and Roseveare (2012) point out, leadership involves those in positions of authority, but also those most capable of instituting change within and across departments, whether that is due to position, expertise, credibility with peers, or a combination of all three. Establishing these multi-level, cross-unit teams requires meticulous attention to clarifying and coordinating roles and responsibilities.

Many effective models have involved faculty advisory boards, course-leads, or other leadership model at the faculty level (Arreola, 2007; Weschke & Canipe, 2010). Faculty-nominated curriculum leaders have been

> The primary difficulty in establishing successful faculty evaluation and professional development programs is not so much a technical one of developing the right questionnaires and procedures. Rather, the real problem lies in getting large numbers of intelligent, highly educated and independent people to change their behaviour. If we recognize this fact and deal with it openly from the beginning, we have a much greater chance of establishing a successful programme.
>
> (Arreola, 2007, p. xxiv)

instrumental in supporting curriculum initiatives and evaluative efforts (e.g., Cape Peninsula University of Technology in Hénard & Roseveare, 2012). Administrators and programme directors within departments can be effective in leading systematic review, assessment, and re-design in collaborative teams that include part-time, full-time, or online faculty (e.g., Universitat Oberta de Catalunya, Spain in Hénard & Roseveare, 2012; and Walden University in Weschke & Canipe, 2010). Systematic student engagement and input into evaluation systems can improve those systems, and increase student faith and participation in data gathering activities (Hénard & Roseveare, 2012).

Comprehensive and multi-directional communications and consultation strategies are critical (Arreola, 2007; Gravestock, 2011). In its implementation of online SRIs Dalhousie University incorporated institution-wide communication of the implementation to all stakeholders, using digital displays, sticky notes, and a wide variety of institutional online and print communications. In-class encouragement by instructors about the importance of the evaluation efforts as well as in-class opportunities to complete online evaluations were encouraged (Kiceniuk, 2012). Communications and coordinating practices must work out solutions to the traditional disconnects among centralized administrations and discipline-specific cultures (Hénard & Roseveare, 2012). Alderman and Melanie (2013) highlight the following lessons learned from the REFRAME project: (a) executive support (time, finances, human resources) is critical to successful outcomes; (b) widespread change should be implemented incrementally, systematically, and collaboratively; (c) the focus should be meeting stakeholder needs rather than constraining or managing them enables agency and promotes buy-in; (d) initiatives must build on the work of others and attending to culturally sensitive institutional nuances; and (e) initiatives should incorporate a multidimensional approach to data sources and collection within an overarching framework with elements that meet the needs of stakeholders and inform the outcomes.

> A long-term, dynamic, systematic implementation can prevent "outright rejection by faculty members and shape a consistent policy that serves the community as a whole"
>
> (Hénard, 2009, p. 6)

Finally, building cultures that value teaching and developing expertise in evaluation, and evidence-based inquiry and decision-making, appear to have a positive impact on faculty engagement with teaching evaluation and on the quality of decision-making (Arreola, 2007; Gravestock, 2011; Gravestock & Gregor-Greenleaf, 2008; Lee, 2007). Successful models identified throughout this study reflect sustained, responsive leadership, inquiry-based approaches, and an emphasis on improvement cycles both with regard to teaching and the evaluation system itself. Most successful models of teaching evaluation appear to be elements of integrated and highly embedded systems of teaching enhancement.

When implementing new initiatives, the importance of a long-term, dynamic, systematic implementation phase can prevent "outright rejection by faculty members and shape a consistent policy that serves the community as a whole" (Hénard, 2009, p. 6). To expand such an endeavour to the provincial level is ambitious: consolidation of varied initiatives in a coherent provincial policy would require a sustained, long-term, non-linear, collaborative

effort across disciplines (Alderman & Melanie, 2013; Hénard, 2009).

Teaching Evaluation: Core Practices

While the findings above provide critical foundations for the planning, design, and implementation of a teaching evaluation framework, it is important to understand the standard, proven elements that typify good practice in teaching evaluation. These include the following:

- Student ratings of instruction: face-to-face and online
- Peer review of teaching
- Self-evaluations
- Curriculum evaluations
- Teaching dossiers

This section provides a brief overview of how each of these works: a more detailed discussion of each practice can be found in Appendix B.

Student Ratings of Instruction

Student ratings of instruction are the most popular form of faculty appraisal in North America, often used to inform summative decisions (Berk, 2009). Although some authors claim that SRI data lack validity (Langbein, 2008), the majority of studies find that students are capable of making valid ratings of instruction (Beran & Rokosh, 2009; Hativa, 2013a; Marsh, 2007; Theall, 2002). While SRI research is replete with contradictory findings, various practices have been found to be effective: the use of items where students are able to make accurate judgments including questions about demeanour in the classroom, student experience, and student perceptions (Buller, 2012; Pallett, 2006; Theall, 2002); implementation designed in such a way that it creates student perceptions of SRI credibility and value (Joughin & Winer, 2014); and

C A S E S T U D Y

The Implementation of Student Ratings of Instruction at McGill University: A Sustained Leadership Approach*

McGill University implemented online student ratings of instruction (SRI) in 2006. The shift yielded many pragmatic benefits: reduced administrative workload; improved data security, access and processing times; and reduced environmental impact and costs. But the impact has been much more fundamental: McGill's team, headed by Dr. Laura Winer, has, as she puts it "tried to use SRIs as a tool for change, not just as a recording and retrospective tool." The intent is to inspire a more inquiry-based relationship with these data.

Implementation has been incremental, smooth and gradual, and reflective of effective change practice. A proactive approach to communications and strong institutional support were critical to this process. There was some strong initial resistance to the online system from faculty, especially concerning the risk of lower response rates. Staff from Teaching and Learning Services went to as many department and faculty meetings as possible to listen to staff concerns, explain the advantages of the system and use staff feedback to improve the system. Provosts and vice-provosts gave their full support to the system and refused to allow any exceptions for faculty not wanting to participate. Faculty now accept the system as a normal part of their working environment. Currently, all courses with an enrolment greater than 5 students are required to use the system.

The approach balances centralized control with flexibility, and has implemented gradual change to better reflect current practice. When the system first moved online, there were no changes from the paper surveys: the team worked towards a more standardized questionnaire that still left room for departmental and individual inquiry. The current model has four standard questions, but space for up to 25. It also accounts for evolving practice, for example, team teaching, and

de-emphasizes lecturing given changing instructional practice at the institution. All questionnaires include at least one opportunity for qualitative feedback, but faculty are free to add more if they wish.

Faculty and students have been involved throughout implementation. Ongoing stakeholder consultation and input have resulted in changes to the system: many units now choose, for example, to leave the SRI response window open until after exams, something that both students and faculty requested.

Regular data analysis has helped to establish transparency, engagement, and faith in the system. Reporting includes means by faculty, course level, and class size: Dr. Winer indicated that this information enables faculty to assess factors that truly impact scores: this practice has been effective in combatting the kinds of enduring myths that frequently undermine these processes. In one instance, faculty believed that the students responding online were more likely to be on the extremes than those reporting on paper versions of the instrument, based on the evidence of the student comments they were receiving. The team incorporated an examination of rating ranges into the following year's process, demonstrating that although the comments were more extreme, the rating ranges were not statistically different. This approach facilitates greater transparency, ensures fairness, and models an inquiry-based approach to thinking about these data.

Based on experience consulting with faculty and department heads, the Teaching and Learning Services unit authored a detailed guide to interpreting and reporting SRI results, which includes statistical evidence of factors associated with statistically significant differences in SRI results: this information, based on their own data, allows for a high degree of transparency in the reporting and analysis of instructor performance. While this approach has not resulted in uniform engagement with these data, many department chairs greatly value the SRI feedback and use it to good effect in faculty development. Efforts are also being made to demonstrate to students the impact of their ratings, in order to affirm the importance of their contributions.

As a model for change, McGill's approach reflects many aspects of the effective practices identified in our study: it is underpinned by iterative feedback loops and strong, sustained leadership and communication; and has, as its basis a profound commitment to building a consensus around the nature and evolving characteristics of effective teaching. Although this particular facet of McGill's practice is focused on SRI, it should be noted that this is part of a multi-faceted approach to teaching evaluation: SRI data is one element of the teaching dossier approach employed across the University. It has been, as Dr. Winer notes, a long, but positive process. Engaging in culture change of this magnitude is much more complex than simply changing the medium of instrument delivery.

*Comments are based on an interview with Dr. Laura Winer, Teaching and Learning Services, McGill University.

systems which facilitate faculty discussion and reflection upon the ratings (Penny & Coe, 2004; Winchester & Winchester, 2011).

Serious concerns with SRIs have also arisen in the literature including the use of SRI surveys that have not been validated (Arreola, 2007; Berk, 2013; Marsh, 2007), the lack of evidence linking SRI results to student learning (Berk, 2014; Hativa, 2013a), and little evidence to suggest that SRI ratings change over time or are used effectively by faculty to improve their teaching (Beran & Rokosh, 2009; Beran, Violato, Kline, & Frideres 2005; Lang & Kersting, 2007; Wininger & Birkholz, 2013). Furthermore, a variety of factors have been shown to impact SRI scores: class size, instructor likeability, course difficulty, course level, faculty, delivery method, and age of instructor (Clayson, 2009; Galbraith, Merrill & Kline, 2012; Hativa, 2013b; Langbein, 2008; Patrick, 2011; Slocombe, Miller, & Hite, 2011; Sullivan, Polnick, Nickson, Maninger, & Butler, 2013). Therefore, any interpretation of SRI results by faculty and

administrators must be performed thoughtfully and carefully (Buller, 2012; Hativa, 2013b; Struyven, Dochy, & Janssens, 2010; Theall, 2002).

Universities in Canada and internationally have implemented online SRIs. Although response rates have proved problematic in some implementations (Ravelli, 2000; Thorpe, 2002; Hativa, 2013b), Dalhousie University cites references indicating that students prefer online evaluations (Kuhtman, 2004 in Donovan et al., 2006), and provide more thoughtful comments when using user-friendly online tools (Ravelli, 2000). Providing information that clarifies the purpose of the evaluation supports student engagement in the process (Gravestock & Gregor-Greenleaf, 2008; Kiceniuk, 2012). While effective transitions have generally involved sustained and well-managed investment of resources and high degrees of stakeholder engagement (Joughin & Winer, 2014), cost savings from online teaching evaluation have been shown to be as much as 50% (Bothell & Henderson, 2003). One university saved more than 480,000 sheets of paper per year for questionnaire completion alone (McGill University, as cited in Winer, 2013). Some evidence suggests that mean scores from online teaching evaluation are lower than pencil-and-paper scores, and comparisons between these two versions of evaluation may be questionable (Hativa, 2013b). It is not clear if this finding is generalizable at this point, and so transition to online ratings of instruction must incorporate a stage of statistical analysis of outcomes at specific institutions.

Peer Review of Teaching

Peer review of teaching is a systematic, reflective process through which teaching colleagues offer instructors feedback about their teaching for either formative or summative purposes, based on multiple forms of data (Chism, 2007). The process can involve a variety of information gathering approaches, including discussion, classroom observation, review of teaching dossiers, instructional materials and curriculum plans, assessment practices, and evidence of student outcomes. Although not as commonly used as student ratings, peer evaluation of teaching is gaining popularity. Numerous studies indicate that peer review can provide accurate input to teaching evaluation in areas where students are less

C A S E S T U D Y

Establishing An Integrated Culture of Peer Evaluation of Teaching at the University of Wollongong*

Peer evaluation of teaching has a long history at the University of Wollongong (UOW). In 1989, a university learning and teaching course became mandatory for all new faculty members. The course included peer evaluation as an integral part of the learning process. One important assignment in this programme was for faculty to teach a concept and have that teaching videotaped and then critiqued. The second portion of the course paired new faculty members with existing faculty members who had been identified as good teachers. New faculty would observe, and be observed by, the experienced faculty member. Because this course has been in place for so long at UOW, almost all faculty members have been through the peer evaluation process, and peer evaluation has become embedded into the University's culture.

The use of peer review at UOW has increased in recent years due to the success of a pilot project in the Faculty of Medicine. Fairly new to UOW, the Faculty of Medicine had little SRI data to inform early teaching evaluations. The development of a peer evaluation system allowed for the collection of more evidence to inform promotion and tenure decisions. In order to be a peer reviewer, faculty would have to fulfill three conditions: i) the dean must recommend them; ii) they must complete a course in peer evaluation; and iii) they must have been themselves the subject of a peer review process. These three conditions were put in place so that the reviewers would have credibility as well as an understanding of the peer review process from both perspectives. This programme has since become University-wide

and generally follows identified effective peer review practices, including:

- The process is voluntary.
- Pre-observation meetings determine the focus of the observation.
- Peer reviewers are trained in how to be effective peer reviewers.
- Post-observation meetings include a draft of the report so that the observed teacher can make any necessary clarifications or justifications.
- The evidence is valued by administrators and used for decision-making purposes.
- The faculty member who was reviewed controls use of the reports.
- Reports are stored at the Centre for Teaching and Learning and are confidential.
- Clear guidelines and procedures are in place and are followed.
- To increase reliability, at least two peer reviews must be submitted.

The peer evaluation process was introduced with little resistance from faculty, and has been fairly successful both because of the long established culture of having one's teaching observed by peers and also because of the University's adherence to known effective practices. Peer reviewers can also count their reviews as service to the University for their own tenure and promotion decisions. The voluntary nature of the programme seems to have increased faculty buy-in to the process and to have minimized some of the stresses and risks associated with having peers observe and evaluate your teaching.

[*] This information comes from an interview with Dr. Gordon Joughin, who is currently at the University of Queensland but worked closely with the peer review process while he was at the University of Wollongong, Australia. The process implemented at the University of Wollongong serves as an excellent case study of effective peer evaluation processes being put into practice.

likely have the expertise to provide accurate feedback, such as course content, goals, and design, pedagogical content knowledge, evaluation practices, or ethical standards of practice (Cohen & McKeachie, 1980; Keig & Waggoner, 1994).

Peer review of teaching has the potential to be an effective means of sharing ideas, improving collegiality and enacting meaningful improvements. However, these benefits only seem to occur if faculty are involved in the entire process, if the process is demonstrably valued by administrators and if appropriate time for discussion and reflection is provided (Boerboom et al., 2011; Hansen et al., 2007; Iqbal, 2013; Snavely & Dewald, 2011; Trujillo et al., 2008; Wellein, Ragucci, & Lapointe, 2009). Despite the potential for improvement in teaching practices, some faculty feel that the peer review exercise is a pointless "ticking the box exercise," that their peers are not competent evaluators of teaching, and that reliability concerns with the activity have not been sufficiently addressed (Centra, 1993; Chamberlain, D'Artrey, & Rowe, 2011; Goodwin & Stevens, 1996; Iqbal, 2013).

Self-Evaluations

The practice of self-evaluation can contribute to raising instructors' awareness of individual teaching styles, enhancing reflective capacity and assisting them to set goals for improved practice. Whether the impetus is the improvement of practice or personnel decisions, engaging in the process of self-evaluation can influence faculty performance, including performance outcomes, motivation to teach, satisfaction with teaching, and likelihood that they will engage in teaching development activities (Amundsen et al., 2005; Marsh & Roche, 1993; Roche & Marsh, 2002). Self-evaluations should be conducted in such a way that they are constructive, designed to improve teaching, relevant to more formalized evaluation efforts and precede student ratings (Fang, 2007; Weschke & Canipe, 2010; Zakrajsek, 2006). Self-evaluation is a critical component of the instructional improvement cycle, enabling systematic, staged adoption and refinement of new approaches (Weimer, 2002). Acceptance of the value of self-evaluation as a component both of improvement-oriented evaluation

C A S E S T U D Y

Situating Instructional Practice Through Guided Reflection: The Teaching Perspectives Inventory*

The Teaching Perspectives Inventory (TPI) is an empirically-informed instrument which quantifies five common perspectives on teaching. It has been used as a means for institutional change and as a tool for professional development. The TPI reinforces the idea that there are many ways to teach and many situations that require different teaching practices.

Dr. Dan Pratt, University of British Columbia (UBC), developed the instrument following three observations of and experiences with peer evaluation that highlighted, for him, how different perspectives on quality teaching can impact peer evaluation. It is important to note that student and peer evaluation of teaching are part of the renewal and tenure process at UBC. During the first experience, a peer evaluator asked Dr. Pratt whether he believed students were getting their money's worth; this question, with its negative, learner-as-consumer implication, reflected how different his conception of teaching was from the evaluator's. The second experience occurred while he was teaching in China – simply, he learnt that elements of his teaching were ineffective due to cultural context. In the third example, Dr. Pratt explained that health sciences is an area where there is a clear formula for what is considered effective teaching, unlike many other disciplines. Observing these contextual and personal variations in diverse social, cultural, and physical settings provoked his study of this phenomenon.

Dr. Pratt and his students observed and interviewed 254 different instructors to learn about teaching goals and accompanying methods as well as their personal theories of teaching over the course of five years. Five perspectives of teaching emerged from their study: Transmission, Apprenticeship, Developmental, Nurturing, and Social Reform. Dr. Pratt and his colleague John Collins developed an instrument that would enable instructors to assess their perspective, and made it available online in 2000. The TPI contains 45 items on a 5-point scale and yields high internal consistency with stable comparisons across a variety of educational settings. Since then, more than 300,000 people have taken the TPI.

Although the TPI is not an evaluative tool, it is useful in the evaluation process. The instrument is self-reflective; a discussion tool for people to talk with others about teaching and its multiple approaches as well as with prospective evaluators about teaching prior to an evaluation; and a tool for a pre- and post-assessment of teaching. The TPI is used in some institutions as part of a systems approach to evaluation. It can also provide insights for those evaluating across disciplines: there are clear patterns of divergent response across disciplinary profiles. The TPI can support understanding of these differences during an evaluative process.

The TPI is one of a range of effective tools for enhancing instructors' and evaluators' reflective capacity regarding teaching, and is the kind of instrument that can be easily integrated into processes for teaching dossier development. It can be used as a component of multi-faceted documentation of teaching practice, as a tool for the development of shared understandings of quality that also reflect the diversity of teaching practice, and as an element of feedback and consultation regarding teaching evaluation. Its use can also inform more nuanced leadership in the drive to improve teaching quality and teaching evaluation.

* Comments are based on an interview with Dr. Dan Pratt, Professor & Senior Scholar, 3M National Teaching Fellow, Centre for Health Education Scholarship, University of British Columbia.

and of evaluation for personnel purposes has grown over the last two decades (Centra, 2000).

Curriculum Evaluations

A component of effective teaching and learning is the design and implementation of curriculum, course content, and student learning contexts (Hénard & Roseveare, 2012). Evaluating and improving the curriculum is part of quality teaching. Effective practices include:

1. Considering the range of student experiences that are part of an educational course (Palloff & Pratt, 2008; Hénard & Roseveare, 2012).
2. Using cyclical feedback mechanisms that measure established competencies against achievement outcomes (Dunn, Morgan, O'Reilly, & Parry, 2004; Palloff & Pratt, 2008; Smith, Herbert, Robinson, & Watt, 2001).
3. Undertaking curriculum review as a component of peer review activities (Chism, 2007).
4. Differentiating the unique aspects of online environments (Sullivan, Polnick, Nickson, Maninger, & Butler, 2013; Taylor & Maor, 2000; Tobin, 2004; Weschke & Canipe, 2010).

Teaching Dossiers

According to the Canadian Association of University Teachers (2007), a "teaching dossier is a summary of an academic's major teaching accomplishments and strengths. It is to an academic's teaching what lists of publications, grants, and academic honours are to research," (p. 2). Teaching dossiers consist of a range of quantitative and qualitative data, often including a record of teaching responsibilities, SRI data, written feedback and comments from students, supervision responsibilities, a teaching philosophy, descriptions of pedagogical approaches employed across a broad range of contexts, peer observation reports, records of innovative practices and their impact, evidence of involvement with curriculum renewal or design, and student work samples (Wright & O'Neill, 1995).

As Seldin (1991) points out, teaching dossiers serve a dual purpose: 1) allowing for the collection and representation of hard evidence of teaching effectiveness for decision-making and evaluative purposes; and 2) providing an effective framework to facilitate reflection about areas of teaching that need improvement. The teaching dossier is also directly relatable to the tripartite requirements of the promotion and tenure process (i.e., teaching, research, and service), which may serve to increase faculty buy-in. However, institutional emphasis on each requirement varies; therefore, a flexible model that is comprehensive yet adaptable would have the most

C A S E S T U D Y

Teaching Dossiers: A Balanced and Supportive Approach to Faculty Engagement at the University of Windsor*

The teaching dossier is a useful evaluative framework: it employs multiple data sources to enable formative, reflective, and dynamic evaluation, consistent with the complex nature of teaching. The dossier offers an opportunity to develop a sophisticated awareness of how individual teaching philosophies are situated in relation to student learning and outcomes, and to elucidate pedagogical choices and performance. While some institutions mandate dossier completion, Dr. Wright advocates a voluntary approach with strong cultural mechanisms incentivizing participation, in particular emphasizing the value of gradual, consultative approaches to shifting practice in this area.

There is a role for a central authority in establishing some consistency in practice: however, a degree of flexibility is critical to reflective practice. The University of Windsor has an optional teaching dossier process for promotion and tenure decisions, and offers a well-delineated guide to the development of the dossier with a number of required

components, allowing for both flexibility and greater consistency among dossiers.

Formal introduction, support, and enticement are essential to the establishment of a strong teaching dossier development tradition. The University's Centre for Teaching and Learning provides ongoing support for dossier development, as well as an annual and over-subscribed one-week intensive Teaching Dossier Academy (TDA), which aims to enable every participant to leave the week with a rough draft of the dossier in hand. Participants take TDA for a range of reasons, from those who are seeking their first university positions, to those undertaking the process for professional growth, to those preparing for promotion and tenure, to those taking their first steps towards the submission of external teaching award applications. There is also a TDA stream for educational developers, for whom the dossier is generally a critical professional document. Successful features of TDA include mentoring from an educational developer, peer consultation, small groups, and reflection on practice. There is broad faculty uptake and anecdotal feedback is positive. Approximately 20% of Academy enrolment is external to the University. The TDA is often a gateway to greater involvement with instructional improvement on campus: participants enroll in order to complete their documentation for promotion or tenure, but the reflective process inspires greater overall engagement with teaching improvement.

One area that is particularly challenging is ensuring that administrators and promotion and tenure committees are well-equipped to parse teaching dossiers and render decisions. Locally defined criteria are key, but systematic processes providing a range of criteria to capture diversity and context are also important.

Academic administrators at all levels can have an impact on the adoption and development of effective evaluative practices. Raising awareness and dialogue among these groups is critical to improving practice. Like all evaluation practices teaching dossiers work best when there is buy-in, and buy-in requires intrinsic and extrinsic rewards: in order for teaching dossiers to become fully integrated into institutional practice, their benefits at the individual and collective levels must be much better understood.

[*] Comments are based on an interview with Dr. Alan Wright, Vice-Provost, Teaching and Learning, University of Windsor. Dr. Wright has more than 25 years of experience with the teaching dossier movement and has published extensively in the area.

utility. There are many possible technological avenues that might be pursued in establishing a more integrated approach to dossier-based evidence and data collection: this is an area for further research and evaluation.

Evaluation as a Programme

Regardless of the specific practices involved, effective teaching evaluation involves systematic and thoughtful planning, design, implementation and administration, data interpretation, reporting, and judgment. Culturally, it involves values, consultation, engagement, communications, negotiation, decision-making, governance, and reward structures. Good evaluation programmes are not additive: they are woven into the fabric of institutions (Arreola, 2007).

There is substantial agreement among scholars regarding the critical features of evaluation programmes. Two widely accepted and often-cited models are Seldin (2006) and Arreola (2007). Seldin (2006) emphasizes relevance (match to institutional goals), comprehensiveness, sensitivity (ability to distinguish between effective and non-effective teaching), freedom from contamination (e.g., exclusion of faulty instruments or items), reliability, and acceptance by stakeholders. Arreola (2007) emphasizes explicitly articulated institutional value structures and the integration of evaluation with improvement programmes. Grounded in the Canadian context, Vajoczki (2008) and Gravestock (2011) each undertook extensive reviews of the international literature on comprehensive, integrated, and aligned teaching evaluation programmes, identifying similar core elements (Table 2).

National Initiatives: Evaluation Frameworks

One emerging trend in teaching evaluation is the attempt to establish national or sector-wide frameworks for the evaluation of teaching. Three examples, provided below, reflect some fundamental principles worth considering. The projects evolved through the collaborative work of faculty and institutions, rather than

Table 2: Review of Well-Aligned Teaching Evaluation Programmes

	Arreola (2007)	Seldin (2006)	Vajoczki (2008)	Gravestock (2011)
Purpose	• Clarity of purpose • Alignment with an explicitly articulated institutional value structure	• Relevance	• Identified goals • Identified purposes	• Institutionally aligned goals and purpose
Definitions and Criteria	• Clearly defined criteria	• Sensitivity		• Definition of teaching effectiveness • Identified evaluation criteria and standards
Roles and Responsibilities	• Sustained, systematic faculty involvement			• Clear roles and responsibilities • Clear and transparent governance
Systematic Collection of Multi-faceted Data	• Variety of data sources	• Comprehensiveness	• Reflective of breadth of practice • Multi-dimensional data sources	• Multiple forms of evidence • Multiple evaluation mechanisms
Feedback Cycle	• Meaningful feedback • Integration with faculty development programmes		• Integrates consultation and feedback	• Support and training
Validity and Reliability	• Accuracy, validity, reliability	• Validity • Reliability	• Valid, reliable, and stable instruments	
Communications and Transparency	• Provide detailed information to faculty • Strong communications plan	• Acceptance		• Effective stakeholder communications
Other	• Facilitative reward structure		• Works within a context of a supportive teaching culture	

emerging as externally imposed mandates. They involve the use of multiple forms of data, only one of which is student ratings of instruction, and employ dossier-like approaches to articulating practice. They also involve high levels of integrated institutional practice, and the intentional development of training and expertise plans to ensure effective use of data. In two of three cases, the emphasis is on teaching quality at the programmatic and institutional levels, rather than at the individual level. In practice these two are quality assurance processes with a strong multi-faceted emphasis on teaching quality, rather than systems for the evaluation of individual teachers.

The DOCENTIA Programme (Codina & Jimenez, 2008)

Traditionally, the Spanish approach to university teaching evaluation focused on teaching qualifications. Between 2002 and 2007, seven public Catalonian universities collaborated with their regional quality assurance agency to develop a system of evaluation intended to focus on teaching competence, rather than qualifications. The direct motivation for the initiative was a regional agreement to link salary increments to teaching evaluation in the region, an atypical practice in Spain. The system is based on portfolios, student feedback, and a report from the instructor's supervisor.

The system was subsequently adopted by ANECA, the National Agency for Quality Assessment and Accreditation of Spain. It evaluates four criteria: the fit between a university's requirements and instructor practice, student satisfaction, teacher's effectiveness in developing students' competencies in specific areas within a context, and teaching innovation. Regional quality assurance agencies enact standards and guidelines to verify procedures universities are using to assess teaching staff competence, and certify that the procedures have been appropriately applied. The universities design and put into practice institutional procedures for evaluating teaching competence. Administrators, students and instructors carry out the assessment. A 2008 report indicated that the programme, which by that time included 62 out of 74 Spanish universities, had been well-received by universities; The review also indicated that the programme had raised awareness of teaching quality and characteristics of effective teaching, in part because of its strong consultative basis. However, both faculty and administrators expressed concern about the labour intensity of the process. Stakeholders found reviewing dossiers to be a complex task.

The Australian Teaching Standards Framework

The Teaching Standards Framework (TSF) was first developed at Macquarie University for its own internal auditing purposes in 2009, with financial support from the Australian Learning and Teaching Council. This project was expanded with a pilot at 11 universities in 2012. The use of the TSF is not mandated: it provides a tool enabling institutions to comply with the reporting requirements related to the teaching area as defined in Australian Tertiary Education Quality and Standards Agency threshold standards (Sachs, 2012).

The TSF offers departments and institutions a tool for comprehensively assessing and reporting on teaching quality. The approach is holistic, examining every aspect of teaching, from management structures, to policies and practices, to curriculum design and learning support. The tool is built around six standards involving (1) teaching structures and practices, (2) institutional service and support, and (3) how curriculum produces quality learning outcomes and student learning experiences. Each theme is reviewed against seven focus areas: management, responsibilities, planning, resources, policies and procedures, practices, outcomes, and monitoring and evaluation (TSF User's Guide, n.d.). The TSF then provides criteria (statements of principles) and performance indicators (actions undertaken) for each of these 21 subsections. At the institutional level, the online questionnaire is supplemented by a report regarding the six standards, and also providing institutional context, areas of excellence, identified areas for improvement and timelines to address them. In subsequent reports the institutions should also provide updates on progress on previously identified areas for improvement. One aim of the tool is to create a fuller "organic" image of what is occurring at teaching at the programme or institutional level (Sachs, 2012), so that quantitative and qualitative data – from policy documents to student and stakeholder feedback, student performance data, teaching and learning research and development projects – can be read in relation to one another (TSF User's Guide, n.d., p. 3). In principle this is a nested system: reports at the programme level are visible and should inform reports at the departmental level, which should inform reports at the faculty level, which should inform reports at the institutional level. Because there are different types of universities, appropriate and relevant quantitative

C A S E S T U D Y

Agile and Consultative Process Re-Design: The Queensland University of Technology REFRAME Project

In 2011, Queensland University of Technology (QUT) established a review of its teaching evaluation practices in response to recommendations made by the Australian Universities Quality Agency. The institution identified stakeholder concerns: clarity of purpose and principles, the degree to which practice reflected the dynamic and complex nature of teaching, over-reliance on SRIs, lack of validity and reliability, and over-emphasis on accountability and marketing rather than instructional improvement. As a result, QUT launched the REFRAME project.

REFRAME refers more to the process of establishing the framework than to the framework itself. Undertaken through the collaborative efforts of QUT's Learning and Teaching Unit, relevant reporting and technology departments, and user-led stakeholder engagement, this process involved 13 working parties, collaboration with approximately 500 students, survey pilots with approximately 100 faculty members and 6,600 students, and iterative cycles of consultation and design with university governance bodies and technical staff (Alderman & Melanie, n.d.). Because it sought to integrate attention to stakeholder concerns with the research literature on teaching evaluation, the project demanded a flexible approach to design and development (Alderman & Melanie, 2012): the development team adopted an 'agile' project management approach, derived from the information technology industry, which foregrounds people, communications, the product, and flexibility as core values (Layton, cited Alderman & Melanie, 2012). Stages of discovery, design, and delivery framed the process, which was supported by extremely detailed communications plans.

The evaluation process now involves all academic staff in a user-led, research-based approach to evaluating their teaching: they develop a personal evaluation strategy that draws on multiple sources of data annually, with a focus on student learning. QUT provides a range of automated surveys and optional evaluation strategies that can be launched at different points in the term. These include formative and course-end surveys, exit surveys students who withdraw from a course, instructor-customizable surveys, peer review, and instant response surveys, as well as access to a variety of existing data to inform reflection on curriculum. The existing data is critical as it takes into account institutional, faculty, programme, and course-level reports with hundreds of lines of evidence. The most significant change was handing agency back to academics whereby the personal evaluation strategy was required and teacher survey evaluations were optional. The Kaleidoscope project is an effort to generalize the REFRAME approach: to create an adaptable set of guidelines for institutions undertaking similar change according to their own culturally specific contexts (Alderman & Melanie, 2013).

Key lessons learned include:

- senior administrative support is critical to successful outcomes;
- widespread organizational change requires systematic, collaborative approaches;
- a stakeholder approach is critical to meeting stakeholders' needs;
- complex initiatives should deliberately build on the work of others; and
- tools should reflect both research and stakeholder needs (Alderman & Melanie, 2013).

In terms of the characteristics of effective practice identified by our study, this model demonstrates a commitment to multi-faceted data collection and use, commitment to the development of shared understanding of quality, the necessity for distributed leadership, and a clear commitment to embedded feedback loops. The iterative stakeholder engagement design model described here is highly consistent with what is understood about effective change practices in complex systems.

measures must be established each time the tool is used. As a 2011 report on the TSF states, it is not intended as a complete assessment tool, but as a framework within which assessments can be made: "The most important part of an assessment exercise is the discursive narrative the TSF gives rise to, which provides a full account of where a University is in a particular field, and its long-term development plan" (Sachs, 2011). At some institutions, access to programmatic reports of this nature is also provided to teaching and learning centres for the purpose of programme improvement and strategic planning (Joughin & Winer, 2014). A sample of the hard-copy TSF forms (the system has now moved online) can be found in Appendix C. Universities in Australia also undertakes individual teaching evaluation for personnel decision-making, generally based on a dossier approach.

The UK Experience (Universities UK, 2008)

In the 1990's, the UK established an extensive programme of teaching quality assessment across all jurisdictions. It was conducted on a subject-by-subject basis by teams of academic assessors who reviewed a self-assessment document produced at the department level, usually followed by visits. The review team provided input and recommendations, subsequently included in institutional reports, and summarizing judgments provided to the departments. By 2001, almost all departments in the UK had been assessed: only a small number had been found unsatisfactory, and a review identified that the inspectorial nature of the process had produced a variety of counter-strategies to disguise perceived problems within the system, rather than the intended goal of frank, collegial exchange about areas requiring improvement. One intended use of the reporting was to inform student choice of universities, and evidence mounted that this was not happening.

This process allowed the sector to derive a sense of baseline standards and confirm that standards were comparable across the country. However, it was extremely resource intensive both for the quality assurance agencies and institutions. In 1997, the Dearing Committee concluded: "given that the vast majority of outcomes have been satisfactory, we are not convinced that it would be the best use of scarce resources to continue the system in the long term" (Universities UK, 2008). The functions of the programme were absorbed into the Quality Assurance Agency for Higher Education in 2001. Ultimately the systems were integrated so that teaching quality and the student learning experience are now focal points in programme quality assurance.

Currently, teaching and learning comprises one component of the UK Quality Code for Higher Education, which is the definitive reference outlining "what higher education providers are required to do, what they can expect of each other, and what the general public can expect of them" (Quality Assurance Agency for Higher Education [QAA], 2012, p. 3). Each chapter is accompanied by a series of indicators reflecting sound practice in that area, intended to prompt reflection and inspire procedures, regulations, and practices to demonstrate compliance with these expectations. The recently adopted chapter on teaching and learning offers greater integration of standards related to modes of instructional delivery, work-based and experiential learning, and a previously separate section on disabled students. The emphasis in this work is on documenting the learning opportunities that higher education providers make available to students at the programmatic level.

The Code reflects an understanding of the complexity of teaching and learning, and the many factors that influence both. Those factors may have varied impacts on different individuals, and that must be taken into account in designing instruction, learning support and learning environments. It emphasizes that there is great variety in effective teaching and effective approaches to empowering student learning. The critical themes are:

- inclusive learning through promotion of equality, diversity, and equal opportunity;
- higher education providers, staff, students and stakeholders working in partnership; and
- learning facilitated by enthusiastic and capable staff (anyone who is involved in teaching and learning) through teaching and other types of support for learning, formal and informal.

While the Code provides a suite of indicators based on these core themes (see Appendix C), their evaluation is based on how institutions individually articulate their strategic approach to teaching and learning. The

review is subject-specific but also takes into account crosscutting themes such as citizenship, academic and digital literacies, and ethical behavior. The Code outlines a wide variety of possible data sources, noting that evaluation takes place at different levels, from individual teacher to course instructors, through to senior management. Overall, programme review is intended to foster collegial discussion and reflection based on multi-faceted data. Teaching improvement is taken into account as an element of improving programmes and student learning overall.

The UK Professional Standards Framework (UKPSF) for teaching and learning in higher education is generally incorporated into this process as one tool for assessing the degree to which programmes have achieved several of the indicators in the Code. However, it is the also primary method of evaluating individual teachers in the UK: it is used to benchmark instructors' teaching qualifications. All instructors are expected to have met the minimum requirements for Descriptor 1 (Associate Fellow). First proposed in 2003, its development was a joint effort among several higher education agencies and the university sector involving significant consultation. Specifically, its aim is to function as:

- an enabling mechanism to support the professional development of staff engaged in supporting learning;
- a means by which professional approaches to supporting student learning can be fostered through creativity, innovation, and continuous development;
- a means of demonstrating to students and other stakeholders the professionalism that staff bring to the support of the student learning experience; and
- a means to support consistency and quality of the student learning experience (Appendix C).

The UKPSF is comprised of a set of standards, including dimensions of practice, which outline areas of activity undertaken, the core knowledge needed to carry out those activities, and the professional values someone performing those activities exemplifies (QAA, 2012). See Appendix C for the full document. As in the Australian system, this model integrates a collective form of teaching evaluation to a strong degree into its programmatic quality assurance. Individual evaluation is based on qualifications and training, not on measures of performance, but does use a national set of benchmarks.

Although Canadian conventions comprehensively emphasize teaching evaluation as a practice focused on the practices and contributions of individuals, it is wise to keep in mind that this is not the only approach internationally. In the UK, for example, teaching evaluation is primarily an element of programme review, where the emphasis is on assessing the student experience across a programme, rather than assessing the performance of individuals for promotion and tenure purposes, or even for the purposes of individual professional development (C. Popovic, personal communication, December 13, 2013). At the individual level, instructors are evaluated based on their attainment of the UKPSF benchmarks, often attained through accredited professional development programmes. Australian quality assurance practices also emphasize the collective documentation of effective teaching at the programme and institutional level with an emphasis on programme improvement. Both systems use a dossier-like approach, which incorporates a breadth of qualitative and quantitative data and attempts to capture the full "system." Although this is not a study of quality assurance, it seems worth considering whether, in terms of the eventual development of a teaching evaluation framework, the critical factor is really to more effectively evaluate individual instructors, or to develop more effective ways to evaluate and improve the collective quality of instruction students receive as a more holistic phenomenon. It should be noted, however, that both the Australian and UK systems are populated by faculty and administrators who have undertaken quite extensive training in teaching in higher education, a level of training that undoubtedly enhances faculty ability to engage in these practices in an informed way.

Summary of Findings

From a review of the literature and successful programmes, it is clear that the following practices are critical to effective teaching evaluation:

- Shared understandings of quality
- Multi-faceted data types and evaluation

- Robust feedback cycles: integration of evaluation and instructional improvement programmes
- Sustained leadership for education, engagement, and change.

Gravestock's (2011) review of core elements in well-aligned teaching evaluation programmes, created as a foundation for a national study of teaching evaluation, and our review of international findings, are highly consistent, and together offer a highly productive basis for identifying necessities for effective teaching evaluation (Table 3).

Our findings reflect the fundamentals identified by Gravestock and others. None of these are discrete, easily achievable interventions: they are not the kind of practices that can simply be "bolted on" to existing programmes. Given the complexity of the activities and the multiple purposes and intents of stakeholders, everyone involved must come to understand teaching evaluation as an integrated set of practices, most effectively based on well-researched principles and instantiated within a specific cultural context.

None of these can simply be "bolted on" to existing programs. Given the complexity of the activities and the multiple purposes and intents of stakeholders, everyone involved must come to understand teaching evaluation as an integrated set of practices, most effectively based on well-researched principles and instantiated within a specific cultural context.

In the Canadian context we tend to think of teaching as entirely an individually determined set of practices and principles, and evaluate accordingly. Other jurisdictions do not see this in the same way, and there is certainly room for further study of the implications of these practices for our context. An international study by Gibbs, Knapper, and Piccinin (2007) provides persuasive evidence that it is at the departmental level, rather than at the individual level, that there is greatest traction for teaching improvement intervention, and this may have some bearing on future directions for the practice and use of teaching evaluation. A recent OECD-Institutional Management in Higher Education (IMHE) (Hénard & Roseveare, 2012) project on quality teaching lends support to this direction for further study. Analyzing the roles of faculty members, departments, university administrations and the state in 29 higher education institutions across 20 OECD and non-OECD countries, it identified three major supports for quality teaching: (a) institution-wide and quality assurance policies, (b)

Table 3: Elements of Well-Aligned Teaching Evaluation Programme

Elements of a Well-Aligned Teaching Evaluation Programme - Gravestock's (2011)	Recurring Themes in our Review of Effective Practices
• Identified evaluation criteria and standards • Definition of teaching effectiveness • Institutionally aligned goals and purpose	• Shared understandings of quality
• Clear roles and responsibilities • Clear and transparent governance	• Sustained leadership for education, engagement, and change
• Multiple forms of evidence • Multiple evaluation mechanisms	• Multi-faceted evaluation
• Support and training • Effective stakeholder • communications	• Robust feedback cycles: integration of evaluation and instructional improvement programmes

programme monitoring, and (c) teaching and learning support. Faculty evaluation is just one of the 21 dimensions suggested for enhancing quality teaching (see Appendix D). A review of international quality assurance practices related to teaching and learning would be of use in determining an integrated and impactful approach to the development of a provincial teaching evaluation framework.

IV

The Ontario Context

In order to evaluate the case for producing customizable tools for the evaluation and improvement of teaching in Ontario universities, it is important to understand the current conditions and context in Ontario: the state of current practice, how data are collected and used, the regulatory context, emerging directions already in play, and a sense of factors that might pose barriers or opportunities for change. These factors play a role in how tools and frameworks should be designed to reflect the current context and address gaps in practice, and in understanding how best to approach such a major change initiative in what is clearly a sensitive area of practice.

The environmental scan team undertook a multi-faceted review of current teaching evaluation in Ontario. This included a review of stakeholder literature on the topic, stakeholder consultation, study of institutional web documentation, and two online surveys (one focusing on policies and administrative responsibilities related to teaching evaluation, and a second focusing on teaching and practices related to teaching and learning improvement at respondents' institutions). In addition, the team analyzed SRI forms collected from most institutions, except those changing their forms. An additional component of the study, undertaken by a separate team, involved an exploration of the commonalities and variability of the policy and labour contexts, which inform teaching evaluation in Ontario universities.

Stakeholder Consultation and Review

There is considerable interest in the collection of more usable data to improve teaching quality: in addition to a major review of teaching evaluation practices undertaken by Gravestock and Gregor-Greenleaf in 2008[1], stakeholder groups including the Council of Ontario Universities (COU), the Higher Education Quality Council of Ontario (HEQCO), the Ontario Confederation of University Faculty Associations (OCUFA), and the Ontario Universities Student Alliance (OUSA) have either engaged in research and reporting in this area within the last five years or are currently undertaking related study.

In a report published by HEQCO, Gravestock and Gregor-Greenleaf (2008) note that teaching evaluations (primarily student ratings) are, by and large, used for summative purposes related to personnel decisions about hiring, tenure, and promotion. While confirming the research that SRI tend to provide valid and reliable data within a limited range of items, they identify five key barriers to their effective use: persistent myths and misconceptions about variables affecting SRI validity; unclear definitions of effective teaching; insufficient education about uses; validity for students, faculty, and administrators; poor presentation and contextualization of evaluation data; and inconsistent and inequitable policies and practices regarding implementation and administration of course evaluations.

[1] Formal research on teaching evaluation in the Canadian context is extremely limited. The work of Gravestock and Greenleaf (2008) and Gravestock's subsequent dissertation on teaching evaluation and its collective agreement context in Canada are foundational work in this area, and are therefore a critical component of our scan.

Generally, they suggest, current practice evokes a positive administrative view of teaching evaluations, a negative faculty view, and an ambivalent student view. To address these concerns, Gravestock and Gregor-Greenleaf highlight a need to make the process and results of SRIs and their interpretation more transparent and accessible. They argue that this is required for the appropriate use of evaluation results in personnel decisions to promote further engagement by faculty in course development initiatives, and by students in course selection and evaluation.

> Our findings suggest that no matter the reliability and validity of the evaluation instruments themselves, the policies, processes, and practices at an institution determine the degree to which evaluations are an effective measure of teaching quality.
>
> (Gravestock & Gregor-Greenleaf, 2008)

There are a number of concerns that Gravestock and Gregor-Greenleaf (2008) highlight in relation to this latter finding. Foremost, it is widely agreed that students are not in a good position to evaluate course content and therefore clear that surveys should not include questions of this type. Second, items pertaining to teaching effectiveness are only valid and useful if conceptualized in accordance with the findings of educational research, and with (shifting) institutional goals and practices as well as student demographics. Third, if one of their stated purposes is to inform teaching development and/or effectiveness, then teaching evaluation items must be designed to solicit comparable, multi-dimensional and longitudinal (i.e., contextualized) types of data and/or be supplemented with other data sources that can provide such measures.

In summary, Gravestock and Gregor-Greenleaf (2008) recommend "that institutions improve [their] efforts to assist students to become better evaluators of teaching and…better train [faculty and] administrators in the interpretation and use of ratings data" (p. 36). In other words, it is necessary but not sufficient to ensure the validity of SRIs by setting clear and consistent goals, definitions and survey items. SRIs must also be promoted and made more amenable to the needs and abilities of students, faculty, administrators, and Ontario institutions as a whole.

The 2012 Annual Report of the Office of the Auditor General of Ontario promotes a standardized approach to SRIs at the departmental, institutional, and provincial levels. They also recommended the use of online tools to facilitate aggregation and subsequent analysis of results, the identification and promotion of "best" practices, the provision of clear guidelines for tenure and promotion, implementation of teaching development initiatives, as well as determination of the relative impacts of various resources on teaching and learning and of the adequacy of higher learning in preparing students for the workforce.

Faculty perspectives are represented in a document produced by OCUFA in response to the Commission on the Reform of Ontario's Public Services, also known as the Drummond Report (2012). In this document, OCUFA argues that given widespread institutional variance regarding teaching evaluation processes, and that they speak only to students' feelings of satisfaction and enjoyment, SRIs are questionable measures of either teaching or institutional quality (see also Kelly, 2012).

Since 2006, OUSA has published policy papers related to teaching quality (2006), quality and accountability (2008), and student success (2010). In these documents, OUSA argues that SRIs do not adequately assess students' learning experiences – that is "[w]hat students...take away from the courses in which they are registered" (p. 9) – and that, often in accordance with collective agreements, SRI results are not made public, hindering students from making informed course selections. Of the recommendations put forth in this document, one stands out: that Ontario universities submit accountability plans to HEQCO which would include, for example, institutional teaching philosophies, identified areas for improvement, available learning supports, and designation of a senior position of leadership in teaching and learning.

OUSA maintains the need for comparable measures of quality and accountability in a 2008 policy paper. Specifically, in this document they argue for measures

that are multi-dimensional, transparent and accessible to all stakeholders, managed in cooperation with students, and flexible and responsive. Again, like OCUFA (2012), OUSA (2008) makes a case for processes of accountability that reward improvement, are relatively cost efficient, and are conducted at arm's-length. HEQCO is touted as an optimal agency to help with such matters, provided they consult with key stakeholder organizations to ensure construction of a system-level framework to collect and analyze metrics of quality.

There are a few common perspectives on SRIs that emerge from all of the above-cited documents. First, while students are likely the best evaluators of how something is taught (i.e., course delivery), they are not qualified judges of what is taught (i.e., course content). Therefore, valid and reliable evaluations must be multi-faceted, and designed and distributed in ways that recognize that teaching is a multi-dimensional, variable practice that requires multiple methods of assessment in order to help offset "uncontrollable factors (such as class size, age and gender of instructor) that impact [evaluation] scores" (Kelly, 2012, p. 11; see also, Murray, 2005). Truly formative feedback may also require enforcement – or at least encouragement – of faculty engagement with institutional supports for pedagogical development (Kelly, 2012).

Second, from the student perspective, it is essential that the impact of SRI results on chosen teaching methods be made public in order to increase the utility of SRIs (Kelly, 2012). Kelly further notes the utility of mid-term evaluations for informing instructors whether or not immediate changes to the course would be beneficial to student learning. Additionally, if the intention is to use information garnered from mid-term teaching assessments for formative purposes, providing students with guidance about how to offer detailed, constructive, narrative feedback on teaching and learning would be indispensable.

Third, from a faculty perspective – particularly untenured faculty – the summative aspects of teaching evaluations are the most contentious (Kelly, 2012). This likely explains those provisions of collective agreements restricting their use and accessibility and, as such, pose challenges to administrators and students (and researchers, as per the 2012 Annual Report of the Office of the Auditor General of Ontario, cited above). Additionally, the central role of teaching evaluations in tenure and promotion decisions may explain reluctance among (some) faculty to experiment with new teaching methods.

A less common, but increasingly timely, issue expressed in the above-noted sources concerns the use, or not, of online methods for conducting teaching evaluations. As Kelly (2012) notes, while online evaluations are found to have lower response rates, they are also found to generate "more descriptive [and] detail[ed feedback]... than paper-and-pencil" formats. Another important issue that deserves further research is "the use of [student evaluations of teaching] as a measure of institutional accountability for excellence in learning outcomes and teaching excellence" (Kelly, 2012, p. 1). In relation to this suggestion, it is useful to note that while standardized data may be especially conducive to the generation of aggregate and exemplary practice data, by definition, they do not account for cross-disciplinary and/or cross-institutional differences in teaching approaches and definitions of teaching effectiveness.

Stakeholder documents reviewed tended to emphasize adequate funding and increasingly stretched resources as critical issues in the quality of teaching and learning at Ontario universities. While they advocate for more effective teaching evaluation in ways which mirror the key findings of our literature review on effective practices, it is clear that they see the mission of teaching improvement as a systemic challenge with many complexities. At the same time, the above synopses of perspectives on teaching evaluation practices and policies emphasize the diversity in conceptualization of various aspects of the phenomenon of teaching, including its quality, effectiveness, development, and evaluation.

Stakeholder Feedback on the Teaching Evaluation Toolkit Project

In December 2013, the research team met with representatives from HEQCO, COU, OUSA, COED, and the MTCU to review the preliminary vision and project plan for the Teaching Evaluation Toolkit. The wide-ranging discussion reflected the following:

- the importance of ensuring that institutions

and administrations provide sufficient time and resources to support individual faculty members in their pedagogical practice;
- the importance of ensuring faculty association consultation regarding teaching evaluation, as consultation with institutional administrations (and institutional web materials) may not really capture important aspects of the faculty perspective;
- the existence of ongoing projects in teaching evaluation through a number of stakeholder organizations, reflecting interest and concern regarding teaching evaluation practices in Ontario;
- the challenges of evaluating teaching as practices broaden and evolve;
- concerns with the usability and interpretation of data: what is publishable, what is not, and how this impacts student experience, change practice, and educational development planning;
- the importance of an inquiry-based approach to one's teaching;
- concerns about how the focus on SRI data for tenure and promotion makes these processes "high stakes" in Canada;
- the need for a more balanced, multi-faceted, improvement-oriented approach;
- the need for greater emphasis on student experience and collective programmatic assessment of teaching practice, rather than on a single-focus approach to human resources decision-making; and
- agreement that the "landscape" of teaching is very messy, constantly changing and evolving, and that effective teaching documentation and evaluation tools must have the capacity to adjust and evolve as practice does.

There were marked differences among stakeholder perspectives: issues such as student access to data, faculty versus administrative perspectives on teaching evaluation, the potential for increased expectations without increased resources, and the high-stakes use of SRI data in personnel decision-making were clear points of tension. There was consensus that multi-faceted teaching evaluation that reflected the complexities of teaching and how it evolves would be of value, and no expressions of serious concern beyond those expressed above. A summary of high-level findings was distributed to stakeholders in May with a request for feedback.

An Environmental Scan of Teaching Evaluation in Ontario

This section provides an overview of current practice in Ontario, based largely on institutional website review and survey data. A total of 16 universities responded to the Policies and Administrative Responsibilities survey (80% response rate) while 100% of universities responded to the Teaching Evaluation Practices survey. This method was supplemented by in-depth telephone interviews with key personnel at a small sample of Ontario universities.

Purpose for Teaching Evaluation

Ontario universities generally articulate two purposes for teaching evaluation: personnel decisions and teaching improvement. Teaching evaluation also serves programmatic and institutional accountability functions, both in discipline-specific and professional school accreditation, and to a degree in provincial quality assurance procedures through the *Quality Assurance Framework*. Because of its implications for personnel decision-making and professional reputation, and widespread perceptions of the problematics of the single-focus SRI approach, teaching evaluation tends to be a "hot-button" topic on many campuses. Most agree SRIs are important for teaching improvement (formative) and personnel decisions related to hiring, promotion, tenure, renewal, and awards (summative) – this trend translates into a positive administrative view of teaching, a negative faculty view, and an ambivalent student view (Gravestock & Gregor-Greenleaf, 2008).

Primary Practice: Student Ratings of Instruction

Survey data confirmed that student ratings of instruction remain the dominant source of teaching

evaluation data used by universities in Ontario: 82% of respondents reported that these ratings are mostly used for summative evaluation, while 29% identified their use for formative evaluation as well. Only one respondent indicated that SRI is not used at the institutional level, although it can be assumed that individual instructors use student ratings if they so desire. Our findings are largely consistent with Gravestock and Gregor-Greenleaf's (2008) study of teaching evaluation practices in North American post-secondary institutions: variability is the main theme that emerges from both that document and our study. They note that while some teaching evaluation procedures are governed by collective agreements between faculty and their universities, others are not, and while some are designed to collect qualitative comments, others are designed to collect only quantitative types of data. There are also institutional similarities to be noted: more often than not, teaching evaluations are conducted in-class rather than online and generally at the end of term. Moreover, all institutions' SRIs are conducted with limited faculty involvement and designed to ask "about course content…[and] teaching behaviours of the instructor" (p. 21). Given the emphasis on this approach at Ontario universities, we have provided a detailed analysis of this institutional practice, divided into four key dimensions of SRI: design, implementation, interpretation, and reporting.

Instrument Design

Information gathered about the design of SRI instruments revealed that most forms were designed through joint input from faculty, faculty associations (or unions), senate committees, and/or university administrators. Those currently in use were adopted at various times ranging from the 1980s to 2013. Only one identified a regular and mandated approach to SRI review and revision, and the last revision of SRIs ranged from no revision since adoption in 1990 through to 2013. Noticeably, some reviews take place without implementing revisions, a testament to either the perceived durability of the instrument already in use or the challenges of achieving consensus about changes. Most locally developed instruments appear to be pilot-tested within the institution: only four of the universities identified an approach to establishing the validity and reliability of their SRI instrument.

In the majority of cases (94%) institutions report the use of a common SRI, but in practice these may be customized at the faculty or departmental level. Departments may also exercise the freedom to change or not to change the instrument in use as they see fit. One university uses a core of common questions with a bank of over 200 other items from which departments, faculties and individual faculties can select additional items. Externally standardized forms, such as the Student Evaluation of Educational Quality (SEEQ) developed by Marsh (2007), are typically assumed to be both valid and reliable based on initial research done by their developers and their wide adoption at other institutions. However, when local adaptations are made, such instruments are not necessarily re-tested for validity and reliability.

Questions on the forms range from 8 to 47 in number. Eighty-three percent (83%) of respondents confirmed that their forms consisted of both scaled and open-ended questions focusing on the instructor and on the course, although they are not always explicitly organized into these two categories. In fact, some forms have all the questions listed without categorization. Others, such as those modeled after the SEEQ, are organized by key elements, including learning, enthusiasm, organization, group interaction, breadth, assessment, and assignments.

Survey participants identified the most common characteristics in questions about the instructor as "rapport with students/interaction" (91%), "accessibility" (89%), and "enthusiasm for the course" (83%). The most common themes identified in questions about the courses include "workload" (100%), whether they would recommend the course to others (100%), their enthusiasm for the course (100%), and course difficulty (100%). Most (80%) of the respondents indicated that their student rating instruments have at least one "overall" or global question which asks students to rate either the instructor or the course for overall effectiveness. Interestingly, in one case it was revealed that the collective agreement does not allow global questions to be used for summative purposes, only for broader evaluative purposes. There is, in some cases, a degree of confusion about questions related to "course" and questions related to "instructor": is the workload in a course a function of the instructor, or the course?

Table 4: Key Themes Addressed by the SRI Forms for Instructor and Course Evaluations

	Instructor Evaluation	Course Evaluation
Organization	64%	73%
Clarity	80%	70%
Enthusiasm/Stimulation of Interest	78%	44%
Rapport with students/Interaction	91%	27%
Responsiveness	78%	44%
Instructor accessibility	89%	33%
Assessment	62%	62%
Exam/Grading fairness	56%	56%
Materials	33%	89%
Activities	20%	100%
Course difficulty	40%	100%
Workload	33%	100%
Recommend to others	50%	100%
Enthusiasm for the course	83%	33%
Student self-assessment of learning	43%	86%
Overall effectiveness	80%	60%
Other	100%	0

Items appear at times under course evaluation, and at times under instructor evaluation in various instruments (Table 4).

Fifteen institutions also uploaded their SRI forms for review. These contain a total of 289 questions, distributed as follows: 200 (69%) are instructor-specific, 66 (23%) are course-specific and 23 (8%) ask for demographic information about the student respondents. Analyses of the student rating forms found them to be consistent with the information reported by survey respondents and interviewees. Appendix E contains summary tables of the surveys.

In addition to the core and global questions summarized above, some institutions give instructors (73%) and/or departments (67%) the option of adding other questions from a bank of questions. In three cases the fact that the instructor and/or department exercised this option is explicitly stated on the final form that students are asked to complete. Several instruments ask students for demographic and other information presumably to help contextualize and interpret their responses. Examples include questions about students' gender, year in programme, major, number of courses completed to date, expected grade for the course, and whether the course is required or an elective. A few instruments ask students to indicate whether they wish to give permission to send their responses to the open-ended questions to the instructor's chair.

Implementation

Survey responses confirmed that there are significant variations in the implementation of student ratings of instruction at Ontario universities. While the majority of universities (approximately 71% of those responding to the survey) evaluate the teaching of pre-tenure, tenured, and sessional instructors once per course, others do

so less frequently, sometimes once annually, and in the case of one university, once every three years. In at least one case departments determine which courses will be evaluated in a given year. Similarly, 71% of the respondents indicated that their institutions evaluate all courses every time they are offered, except for graduate courses and courses with enrollments lower than five. The remaining respondents indicated that only courses taught by "instructors" are evaluated or that either individual instructors or departments determine what courses should be evaluated. There are also variations with regard to the evaluation of graduate courses: in some places they are not evaluated, and in others they are evaluated through different instruments.

Given uneven expertise in parsing teaching evaluation data, formats that foreground statistically valid data, exclude statistically insignificant comparisons, and promote the growth of critical acumen in the examing of data may be critical to improved decision making.

SRI forms are distributed and completed on paper (75%), online (83%), and a few (17%) use mobile technologies. As can be seen from these results, many institutions are using both paper and online modes, typically because they are evaluating their online courses online. Only four universities use completely online systems. A few institutions (33%) use a third party to implement their student rating of instruction.

It is also clear, from the review of teaching evaluation information on university websites, that in many cases responsibility and coordination of teaching evaluation tends to be distributed among numerous units, including the Registrar's Office, Information Technology Services, the Office of the Provost or Vice-President Academic, Deans' and Departmental offices, and in a few cases, Centres for Teaching and Learning and Quality Assurance offices. While this model reflects the complexity of the activity and the many areas of responsibility it touches upon, it may also limit the potential for agile and responsive change to teaching evaluation practices on Ontario campuses.

Ten of sixteen universities identified low SRI response rates as a challenge they face with SRI implementation. Five identified tensions around faculty perceptions of SRI or resistance to changing instrument items. One-quarter of respondents are currently working to resolve challenges with the transition to online SRIs, and three identified lack of standardization as a concern.

Analysis

While some aspects of the SRI process are highly regulated, others exist in a kind of 'grey area' of practice based more on convenience than on strategic policy formulation. For example, the ways in which SRI data are reported or contexualized may be determined based primarily on what makes sense for personnel management and decision-making, or in some cases simply by what appears, to those calculating the data, to be appropriate and accurate. This may be of particular concern in cases where merit pay or other comparative decision-making may be affected by over-zealous approaches to the calculation of teaching evaluation scores: calculating scores to the third decimal place and then ranking them is statistically unsound, but has been known to occur (Joughin & Winer, 2014). These are critical questions: what is reported, its statistical validity, identified norm groups, how the data are stored and organized, and the contextual information that accompanies data may have a profound impact on institutional decision-making, on instructors' capacity to draw legitimate conclusions from their data, and on instructor and stakeholder confidence in the legitimacy and salience of the data.

Moreover, given the uneven expertise that instructors and decision makers bring to their examination of teaching evaluation data, formats that foreground statistically valid data, exclude statistically insignificant comparisons, and promote the growth of critical acumen in the examination of these data may be essential to improved decision-making. Neither our Ontario data nor a reading of international literature suggest that universities are adopting consistently thoughtful and informed approaches to these challenges (Abrami, 2001; Theall & Franklin, 2001; Hativa, 2013a; Gravestock & Gregor-Greenleaf, 2008). Eighty-two percent (82%) of

survey respondents reported that their universities do not use teaching evaluation data in the aggregate to examine student perceptions of teaching over multiple years, to look for patterns, to identify educational development needs, or for departmental, faculty, and institutional planning.

One critical aspect of practice that is rarely well articulated at a formal level is how actual determinations are made about teaching quality. Gravestock (2011) provides a useful overview of the formal language related to teaching quality in Ontario collective agreements and policy documents, demonstrating that while documents relating to personnel decisions in Canadian promotion and tenure guidelines outline a fairly consistent range of teaching activities, detailed definitions of the characteristics of effective teaching are much less common (See Appendix F). There is considerable variety in the requirements for tenure and promotion submissions, and the nature of decision-making, the norms and processes applied, for example, remain opaque. Gravestock's (2011) review of promotion and tenure documentation uncovered no direct references to teaching evaluation rubrics and few references to training for promotion and tenure committees, although some universities do provide unofficial documentation to support administrative decision-making. Numerous universities also acknowledge the salience of disciplinary differences in determining the characteristics of effective teaching. Given all of these characteristics, it is very challenging to envision the establishment of a common set of competencies for teaching across the province, although that is clearly a valuable long-term goal.

The capacity for informed judgment is critical and difficult (if not impossible) to standardize because of the complexity of teaching. For the most part universities have not developed systematic training in teaching evaluation for administrators, promotion and tenure committee members, or other critical members of the campus community, although individual departments often do have guidelines and criteria for determining what is considered to be an acceptable quality of teaching. However, as the AGO demonstrated through review of promotion and tenure cases, decisions risk idiosyncrasy, and may vary from committee to committee, and individual to individual. While departments, faculties, and institutional promotion and tenure committees may have agreed-upon standards to inform decision-making, actual variations may be considerable. Committees are dealing with highly complex representations of individual performance, and multiple committees at multiple levels may treat and evaluate the data in various ways (Hativa, 2013a).

Most universities have not developed systematic training in teaching evaluation for administrators, promotion and tenure committee members, or other critical members of the campus community

It is important to distinguish between tools used to gather evidence for summative purposes, and the actual evaluation of teaching. While the former is the subject of considerable regulation at institutions, the latter, which relies upon informed judgment, is potentially one of the most difficult challenges that must be addressed. It is this challenge that appears to underlie the attitudes of university instructors who view student ratings with skepticism and distrust (Gravestock, 2011). Stakeholder documents and respondents to the survey report that many faculty members, in some cases inaccurately, perceive that promotion and tenure committees use only student ratings to assess their teaching performance and since, in their view, such ratings are emotional judgments by persons not qualified to evaluate teaching, they discount the results, a finding confirmed by Gravestock (2011) and evident in other research on faculty perceptions of teaching evaluation as well (Beran et al., 2005). Faculty members' concerns are exacerbated by perceptions of ineffectiveness in instrument design and implementation.

Reporting

Who

SRI data are generally reported to individual faculty members and administrators, although in a small number of cases course or instructor information is available only to senior administrators. The data are generally provided to promotion and tenure committees through the department and instructor, though in some cases this is

Figure 2: Who has access to teaching evaluation data?

Response	Chart	Percentage
Individual Faculty Member		100%
Department Head		58%
Faculty Office		58%
Students		8%
Centre for Teaching and Learning		0
Provost's Office		50%
Institutional analysis		8%

with the understanding that they remain the property of the faculty member and are not available for review by a third party. In some cases administrator access is precisely the same as the individual faculty members, while at others administrative access is only to required questions and not to any supplemental questions added at the departmental or individual level (Gravestock & Gregor-Greenleaf, 2008). In still other cases department heads only have access to aggregate data. In general, the degree of faculty, student, and administrative access to and use of instructor-specific SRI data is limited and varies across the Ontario sector. In a few cases, the office of institutional analysis and the registrar's office also have access because they are responsible for implementing and analyzing the data. There were no institutional responses indicating that centres for teaching and learning had access to these data. Apart from the individual faculty member,who, as one survey respondent put it, "owns the data" in many cases, the only other persons who have access to the data are the department head and/or dean (58%) and the Provost (50%). Most survey respondents reported that their institutions use such data primarily for performance reviews, promotion, and tenure (Figure 2).

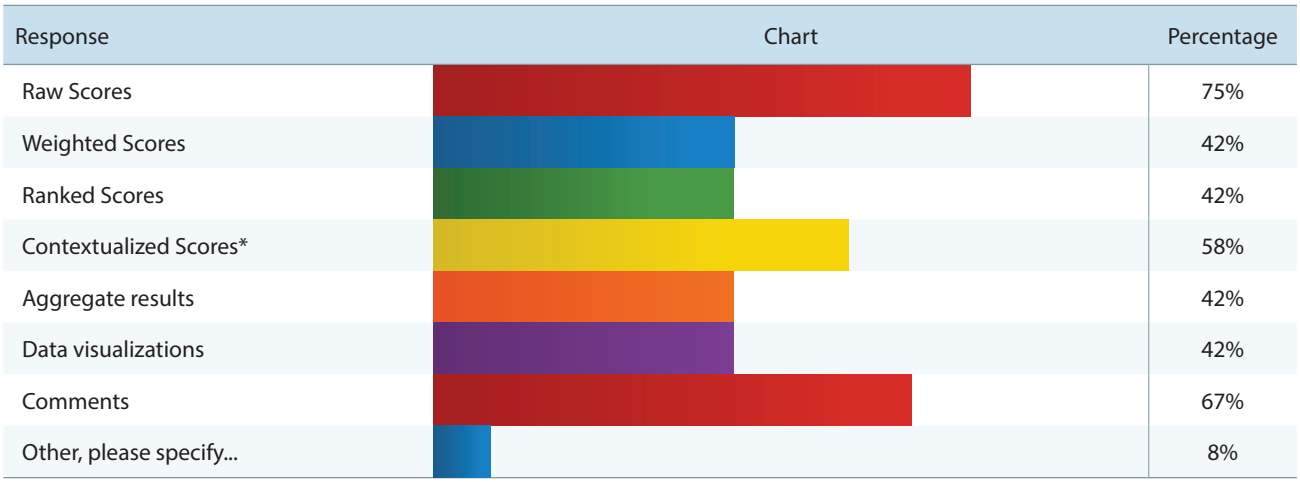
Student access is a particular point of contention according to stakeholder feedback from the Ontario University Student Alliance (Stakeholder Consultation Meeting, 2013). Students appear to have full access to SRI data at only one respondent university, and only one university identified "student course selection" as a use to which these data are put. A number of universities provide aggregate data such as average scores at the faculty level. Others make instructor and course-level data available to students only if instructors release those data. At some institutions, students produce an "anti-calendar" which contains SRI data, but these data are published only with the permission of the instructor (Gravestock & Gregor-Greenleaf, 2008).

While the Freedom of Information and Protection of Privacy Act (FIPPA) is often cited anecdotally as the rationale for these limitations, this is not accurate. As an OCUFA update on FIPPA in 2013 noted, "It may seem counter-intuitive that sensitive employment information—such as employee financial or health information, peer review assessments of faculty members, or student evaluations of teaching—is not subject to the protection of privacy provisions of the Act and therefore that there is no statutory restriction on the Employer's ability to disclose such information. Rightly or wrongly, however, that is the clear implication of court decisions to date" (OCUFA, 2013). Universities may, however, be reluctant to take on this particular battle, and enforcing a change in policy on institutions without groundwork to create a more constructive teaching evaluation culture may be counterproductive. As they are not regulated by law, access to these data are typically limited by collective agreements or institutional policy and guidelines. In all cases reviewed, written comments from students are supplied solely to the instructor.

What

The most common teaching evaluation data that instructors receive are raw SRI scores (75%), comments (67%) and scores in the context of departmental means

Figure 3: What teaching evaluation data do instructors receive?



Response	Percentage
Raw Scores	75%
Weighted Scores	42%
Ranked Scores	42%
Contextualized Scores*	58%
Aggregate results	42%
Data visualizations	42%
Comments	67%
Other, please specify...	8%

* Scores in the context of departmental means or other contextualizing data

or other contextualized data (58%). Notably, 91% of respondents indicate that data from student ratings are not reported with a guide for interpreting the results. Some universities provide graphic representations of data, while others do not (Figure 3).

Teaching evaluation data tend not to be well integrated into systematic approaches for improving teaching or targeting areas for improvement, and are rarely used beyond the level of the individual instructor. Only half of universities reported that SRI data are used for institutional and strategic planning, and, as no centres for teaching and learning have access to the data, they are clearly not using them to inform programme planning or other initiatives. However, the use of individuals' data with permission, for example as a part of teaching dossier development and to drive self-reflection and inquiry, are commonly part of the one-on-one consultation and mentoring that would be typical of the work of teaching and learning centres. At an informal level this kind of information may inform decision-making in numerous ways.

Beyond the SRI: Multi-Faceted Data Collection

Beyond the use of SRIs, many institutions recommend a multi-faceted range of approaches such as peer review of teaching (67%), teaching dossier development (31%), classroom assessment techniques (43%), self-evaluation instruments (43%), and review of video-recordings (31%). Notably, these are all primarily used for formative evaluation, although teaching dossiers (50%), self-evaluation instruments (29%), and peer observations (20%) may also be used for summative evaluation (Table 5).

Table 5: Which of the following are used to evaluate teaching at your institution?

	Used formatively	Used summatively
Student ratings of instruction (SRI)	29%	82%
Peer observation of teaching	67%	20%
In-class surveys or other in-course classroom assessment techniques	43%	7%
Self-evaluation instruments	43%	29%
Review of video-recordings	31%	8%
Teaching dossiers	31%	50%

Table 6: Teaching Evaluation Evidence for Promotion and Tenure at Canadian Universities

Form of Evidence	Number and Percent of Institutions		
	Required	Recommended	Optional
Course evaluation data	29 (63%)	6 (13%)	5 (11%)
Teaching dossier	18 (39%)	7 (15%)	4 (9%)
In-class observations by peers	3 (7%)	9 (20%)	10 (22%)
Letters/testimonials/opinions from students	6 (13%)		14 (30%)
Letters/testimonials/opinions from colleagues	6 (13%)	4 (9%)	9 (17%)
Sample student work		3 (7%)	1 (2%)
Interviews with students		2 (4%)	1 (2%)

Gravestock (2011) provides a summary of common evidence required for teaching evaluation during promotion and tenure processes in Canadian universities (Table 6).

Instructors often incorporate data collected for formative purposes into dossiers to document effective teaching: one challenge brought to light by the survey is that institutions, even when they encourage multi-faceted data collection, often lack standard materials and procedures for those processes, which may make the evaluation of such evidence more difficult (Table 7).

Survey results indicate teaching dossiers and peer observations are mandatory at three and two institutions respectively. Teaching dossiers are in much more common use however: they are either mandatory or optional parts of promotion and tenure at about half of Ontario universities. Review of policy documents also indicated that at some universities teaching dossiers may also be required for some types of faculty members as elements of hiring or review, while optional for others.

Teaching dossier components across the Province tend to be consistent with the elements described in the Canadian Association of University Teachers guide to teaching dossiers (Shore et al., 1991), a touchpoint for teaching dossier development in universities nationally. Required and recommended elements of teaching dossiers are fairly consistent with this guide as well, as delineated in the table below reviewing documentation guidelines for teaching dossiers at 16 Ontario universities. This table reflects the enormous complexity of teaching responsibility at Ontario university, which must be taken into account in any robust teaching evaluation programme (Table 8).

Twelve universities offered insights into the challenges they face in implementing teaching dossiers. Nine identified a lack of consistency as a core problem. Nearly

Table 7: Are there standard processes and instruments used across your institution for the following?

	Yes	No
Student ratings of instruction (SRI)	94%	6%
Peer observation of teaching	38%	62%
In-class surveys or other in-course classroom assessment techniques	33%	67%
Self-evaluation instruments	27%	73%
Review of video-recordings	13%	87%
Teaching dossiers	56%	44%

Table 8: Reviewing the Most Common Elements of Teaching Dossiers at 16 Ontario Institutions

Categories/Approach	Mandatory	Recommended	Optional	Total
Teaching Philosophy	13%	81%	-	94%
Teaching Practices	13%	81%	6%	100%
Teaching/Professional Development	19%	69%	13%	100%
Self-Evaluation of Teaching and Student Learning	13%	75%	-	88%
Contributions				
Curriculum Vitae	13%	6%	-	19%
Teaching assignment(s)	31%	69%	-	100%
Student supervision	25%	56%	6%	88%
Teaching awards or nominations	6%	69%	13%	88%
Teaching-related activities	13%	69%	6%	88%
Teaching-related publications	13%	75%	-	88%
Curriculum/programme development/revision	19%	56%	13%	88%
Grants	13%	50%	6%	69%
Course syllabi	25%	38%	-	63%
New course proposals	-	19%	6%	25%
Colleague mentoring	-	31%	13%	44%
Community outreach	-	19%	13%	31%
Future plans for developing teaching skills and/or future contributions to teaching	6%	38%	13%	56%
Invitations to teach or contribute curriculum to other institutions or departments	6%	25%	6%	38%
Academic advising	6%	19%	13%	38%
Independent study/reading course supervision	-	25%	-	25%
Committee membership	6%	56%	13%	75%
Introducing/use of technology	-	38%	-	38%
Teaching materials	13%	69%	6%	88%
Teaching workload	6%	6%	-	13%
Availability to students	-	6%	-	6%
Identification of student difficulties and encouragement of student participation	-	25%	-	25%
Developing successful internship programme(s)	-	6%	-	6%
Using general support services to improve teaching	-	25%	-	25%
Other kinds of invitations such as a media interviews	-	13%	-	13%

Categories/Approach	Mandatory	Recommended	Optional	Total
Feedback				
Unsolicited letters from students, and colleagues	6%	69%	13%	88%
Solicited letters from students, alumni, and employers of former students	13%	50%	-	63%
Reports from employers of students (e.g., in a work-study or cooperative programme)	-	13%	-	13%
Peer evaluations based on visits to the classroom	6%	56%	6%	69%
Colleague evaluations based on analysis of course documents and materials	6%	44%	6%	56%
Student evaluations of teaching	38%	50%	-	88%
Exchanging course materials with a colleague from another institution	-	6%	-	6%
Conducting research on one's own teaching	-	13%	-	13%
Involvement in an association or society concerned with the improvement of teaching (e.g., STLHE)	-	19%	-	19%
Interview data collected from students	-	6%	-	6%
Written comments from those who teach courses for which a particular course is a prerequisite	-	13%	-	13%
Statements from colleagues from other institutions	-	25%	-	25%
Requests for or acknowledgement of advice received by a committee on teaching	-	13%	-	13%
Documentary evidence of the effect of courses on student career choice	-	19%	-	19%
Statement about teaching achievements from administrators	-	19%	-	19%
Alumni ratings or other graduate feedback	-	13%	-	13%
Appendices				
Annual reports	6%	-	-	6%
Multiple course summary	-	25%	-	25%
Course evaluation reports from the institution	19%	56%	6%	81%
Samples of other evaluations completed by students (e.g., formative/ summative examinations)	6%	63%	13%	81%
Examples of student achievement	-	63%	13%	75%
Student test scores	-	25%	6%	31%
Course(s) status (required/elective)	-	19%	-	19%
A record of students who select and succeed in advanced courses of study in the field	-	31%	6%	38%
A record of students who elect another course with the same instructor	-	13%	-	13%

Note: this table reflects the enormous complexity of teaching responsibility at Ontario university. This must be taken into account in any robust teaching evaluation programme.

half identified lack of understanding of what should be in teaching dossiers or how to read them. While only one respondent is currently piloting an online tool, a much greater number were seeking solutions that would provide stronger guidance, more consistency of format and contents, and a more efficient and engaging tool for instructor use.

Teaching and learning centres are perceived as sources of support for formative teaching evaluation and teaching development, and appear to play little role in summative evaluation. Thus, although SRI data do not appear to play a key role in the planning of teaching improvement strategies at Ontario universities, the range of formative, improvement-oriented teaching evaluation practices described above reflects a core time and resource commitment of centres across the province. The facilitation of reflective, data-driven documentation of practice at universities across Ontario probably does inform planning and decision-making regarding teaching improvement initiatives, but evidence of how this occurs is a subject for further and important research. What seems to be lacking, however, is a systematic approach to drawing conclusions and establishing strategic planning (both for individuals and for the institution) based on a sythesis of multiple forms of data.

A Note on Regulatory Contexts

As should be clear from the descriptions of current practice above, the regulation of teaching evaluation at Ontario universities is detailed, varied, and falls both within collective agreements and within Senate bylaws and policies. Gravestock (2011) notes that teaching evaluation is regualted by collective agreements at 12 universities, by institutional policy at two, and by both at four. Regulated elements at various universities include approval and revision of the instruments used for summative evaluation purposes; procedures for the implementation of summative evaluation; types of data that must or may be included in files for hiring, promotion, tenure and performance review; formal functions the data serve at the university (e.g., personnel review and decision-making, student access to inform course selection); definitions, frameworks, and processes used for identifying "satisfactory" performance; ownership and rights to data; and reporting contents. The degree of regulation reflects sustained concern about the use and effect of these data on instructors' professional lives and on the quality of teaching at universities: the regulatory context itself and what it reflects about attitudes towards and history of teaching evaluation are core barriers to rapid, across-the board change in teaching evaluation practice in Ontario.

SRI data are often intended for both personnel decision-making and for teaching improvement. Because the former can be highly sensitive and have serious financial and legal implications, it tends to be the focus of regulatory practice, and in many cases the implications of these regulations for teaching improvement may not be taken into account. Gravestock (2011) notes, for example, that only two Ontario universities refer directly to a centre for teaching and learning in their promotion and tenure guidelines and documents, reflecting this disconnect. In some cases decisions about procedure, reporting, and file management are made based on what is convenient for hiring, promotion, and tenure processes. These arrangements may not be conducive to their use for other purposes such as instructional improvement.

The State of the Province: Discussion

As a lens for the assessment of the current state of teaching evaluation in Ontario, we will draw again on Gravestock's (2011) framework for a comprehensive teaching evaluation system (p. 218). It should be noted that in nearly all cases, there are examples of strong practice at specific universities: as Gravestock and Gregor-Greenleaf noted in 2008, the overall picture is one of variability, but it is not without its strong exemplars (Table 9).

Table 9: The Ontario Context Compared to Gravestock's Model for Aligned Teaching Evaluation Systems

Framework Element	The Ontario Context
Evaluation goals and purpose are established and align with institutional goals	Strengths • Without exception, all universities in Ontario have **a history of actively seeking to evaluate the quality of teaching** at their institutions both for summative purposes (hiring, tenure and promotion, review) and formative (improvement-oriented) purposes. • Institutions and stakeholders across the province have indicated **increased concern and interest in improving teaching evaluation**, particularly in connection with accountability and provincial quality improvement and productivity initiatives. There are significant differences among stakeholder views of what constitutes "improvement." Gaps • **Limited evidence of systematic evaluation of teaching for pedagogical or programme improvement**: there is a strong perception at the indiviual level that these data are primarily intended for regulatory purposes, and that they are basically "locked away in a drawer" by faculty members after the fact. The use of SRI for accreditation purposes may have a programme-improvement orientation in some processes. • **Overemphasis on summative evaluation**, and insufficient attention to formative evaluation for the purpose of developing and enhancing teaching.
A clear understanding of faculty responsibilities is involved	Strengths • **Some institutions have well delineated procedures and practices**, as well as focused committees for review of teaching documentation (Gravestock, 2011). Gaps • **Inconsistent and potentially inequitable policies and practices regarding implementation** and administration of SRIs. • Lack of clarity about **responsibility for the instructional improvement component of teaching evaluation.**
Teaching effectiveness is defined	Strengths • Some strong examples in place in the province (Gravestock, 2011). Gaps • **Lack of clarity** in Ontario university definitions generally, if provided at all. • Stakeholders express concerns regarding **whether definitions reflect varied and evolving teaching practice in Ontario.**

Framework Element	The Ontario Context
Evaluation criteria and related standards of performance are articulated	Strengths • **A small number of recent, substantive, well-conceived studies of aspects of Canadian and Ontario teaching evaluation practice** which can provide a strong foundation for the development of policy and practice (See, for example, Gravestock, 2011; Gravestock & Gregor-Greenleaf, 2008; Vajockzi, 2008). Gaps • **Little evidence of robust, well-designed rubrics** or other decision-making tools at the institutional level, and considerable evidence of **decentralized practices which are difficult to see or assess**. • Limited aggregate data use makes it **difficult assess whether standards used are statistically accurate.**
A range of evaluation mechanisms are used and multiple forms of evidence are sought	Strengths • Use of a **range of other teaching evaluation tools and processes**, primarily for formative purposes though some, such as teaching dossiers and peer observations, can also be included in summative evaluation. Many institutions have expertise in the support of these practices, generally housed in centres for teaching and learning. • **SRI instruments** across the province, while highly varied, do have **considerable overlap in items**, offering the potential that eventually a small number of common items might be achievable. However, **cross-institutional comparisons are not methodologically valid**. Gaps • Institutions continue **to rely on student ratings of instruction as a single source of data**, or, as Gravestock (2011) puts it, on student ratings of instruction and instructor self-report. Most Ontario universities do not require the triangulation of data recommended in the literature. • **SRI instruments are of varying quality**: some SRIs employ questions outside what is commonly understood to be students' scope of accurate judgment (Hativa 2013b), and there is a growing concern that evaluation questions do not reflect the full range of teaching practices in Ontario universities (online, hybrid, flipped classrooms, service and experiential learning, etc.). • **The structure, design, implementation, and use of teaching evaluation tools is inconsistent** across Ontario institutions, or even within them. Practices can vary from department to department, or depending on the type of position (faculty, sessional, teaching faculty position, etc.), resulting in variations in data which can impact their usability and the consistency of decision-making based on them. • Stakeholders exhibit a **lack of confidence in SRI instruments**. In many cases **perceptions of lack of validity and reliability are not borne out by research**, though in practical terms SRIs would have to be better and more rigorously assessed at individual institutions to clarify the legitimacy of these concerns. Thus in practice, there needs to be a greater emphasis on the analysis of data for validity, reliability and to assess their general usefulness for institutional improvement.

Framework Element	The Ontario Context
Clear and transparent governance and decision-making structures are established	Strengths • Despite widespread perceptions of poor decision-making structures, review of collective agreement and promotion and tenure practices in the province suggest **that institutions have taken varied approaches to foster diligence in the review of teaching** in the province (Gravestock, 2011). • It is clear that **stakeholders within institutions have at various points worked in good faith to reach consensus** about how to proceed with regard to teaching evaluation. Gaps • Apparent lack of stakeholder confidence in current evaluation systems, which they perceive as too bureaucratic, inaccurate, unfairly affected by various intervening variables, or (among students) as ineffectual. • Variation in regulatory contexts from institution to institution: attempting to constructively mandate a common approach across the province at this time could be extremely challenging. • See training, below.
Sufficient support and training for all involved in the review process	Gaps • **Insufficient education about uses, goals, validity for students, faculty, and administrators.** • Limited training in evaluation of teaching performance data and/or teaching dossiers by assessment committees, department heads/chairs, deans, and other administrators. • **Clear evidence that myths and misconceptions regarding teaching evaluation practices persist** despite considerable evidence-based research.
Effective and consistent communication to all relevant constituents is ensured	Gaps • **Lack of common standards for access to evaluation data** for stakeholders. • **Uneven presentation and contextualization** of evaluation data. • **Inefficient documentation, data storage and search tools** result in a limited ability for decision makers to review data and documentation of teaching practice either for individuals or in a more aggregate way for strategic planning purposes. • Ontario universities **lack consistent standards for the reporting of teaching evaluation data**, and generally do not provide effective contextualizing documentation to support faculty, administrator and committee use of data. • **Lack of feedback loop** for instructors. • **Lack of connection between teaching evaluation data and teaching improvement practices** at universities.

The Ontario teaching evaluation system exhibits a number of strengths, but also a number of significant gaps. These require attention and growth if Ontario is to become a leader in improvement-oriented teaching evaluation. We face challenges across all of the identified themes (shared understandings of quality, multi-faceted evaluation, robust feedback cycles, sustained leadership for education, engagement, and change) identified in our effective practices research. While institutions may have some of the formal elements in place, they may face challenges in implementation, in the expertise and awareness to use these systems effectively, in the capacity to integrate teaching evaluation and improvement practice, and, overall, with a sense of ennui and disengagement with the process among administrators, faculty, staff, and various stakeholders. While stakeholders acknowledge that teaching evaluation should be better, resources are limited and demands on time and attention many. Voluntary adoption of a more labour-intensive approach without a compelling reason to do so is unlikely, and mandatory adoption without sustained efforts to change the culture

is likely to result in minimal compliance without the kind of dialogue and engaged inquiry that is critical to teaching improvement (Hannan & Silver, 2000). There is clear concern with teaching evaluation practice, among stakeholders of all kinds, and even a degree of consensus around some of the changes that should occur, like the use of multi-faceted data, improved transparency, faculty and administrator training, and a greater focus on teaching improvement. There is also considerable concern whether teaching evaluation data is now, and will continue to be, used wisely; that access to it not become an opportunity for punitive comparisons; that the use of data be valid and constructive, and that teaching evaluation practices not become an onerous burden on faculty, students, or administrators. Building the better mousetrap in this case is simply not enough.

These challenges go far beyond the question of a better-designed SRI instrument, beyond the adoption of better online tools, or the provision of a collection of "tips and tricks" for better practice. The challenges here are systemic and human: they involve the dispelling of myths, engagement and negotiation of individuals with varied and often conflicting interests; the coordination of practices across multiple units with varied perceptions of the purpose of the activities involved; awareness raising; and above all, culture change. Faith in our SRI-driven teaching evaluation practice is uneven, and many are reluctant to invest the resources that would be required for wholesale adoption of a new approach. The development of new tools is the easy part: what is considerably more difficult is creating a compelling, system-wide belief that these changes are necessary and possible, and that the efforts will be worth it.

It is worth it. Continuing to use and accept a system that is clearly and, by any standard of empirical evidence, insufficient for the purposes to which it is put simply reinforces the belief that teaching, and the effort involved in good and great teaching, is not truly valued by universities. To allow decisions to continue to be made based on insufficient data using insufficient decision-making structures reinforces patterns of cynicism and disengagement that damage the aspirational culture of universities. It also reinforces belief in the pre-eminence of research over teaching in universities, a pattern that has become increasingly incompatible with the demands and needs of the university sector as a whole. Good teaching must be correctly valued, and then correctly rewarded. Problems in teaching – individually and in broader patterns – need to be identified, explored, and mediated in order to give teachers the best chance to improve. The system through which we value this practice speaks volumes about the true value we place on it. A system which does not provide people with the capacity to represent their own work with nuance and to articulate their strengths and needs is not one which inspires growth, leadership, or commitment for the vast majority of people.

More than that, it is necessary. As the Province shifts towards a greater emphasis on quality and differentiation, universities' capacity to assess teaching performance and to integrate those data both into performance review practices and into fully integrated, strategically targeted teaching improvement practice will become increasingly critical to their capacity to compete and thrive. Further, universities' capacity to effectively and strategically differentiate will both require and produce more complex, nuanced data: differentiation will lead to divergent pedagogies, contexts, and teaching practices. A multi-dimensional approach to assessing teaching across the university is one component of the necessary data requirements that institutions will need if they are to succeed within the evolving and increasingly competitive university context. And it's not just gathering that information, or even effectively reporting it, that is the true game changer: it's the distributed capacity to read it, analyze it, critically assess findings based on it, and to plan, evaluate, and refine initiatives based on it. Understanding ourselves through our data is something few members of university communities truly excel at. So although universities, mostly still working with their traditional approaches to decision-making and data collection, may not yet see the crucial need for better data and wisdom around those data, they need to. Teaching evaluation is a case in point, and a promising opportunity to change thinking in the Province, in part because it is such an intractable and embedded problem. The shape of growth in institutions is shifting, and we are reaching the limits of sustainability through enrolment growth (HEQCO, 2013): in the next phase, quality enhancement, differentiated mandates and responsiveness to changing markets and policy

These challenges go far beyond the question of a better-designed SRI instrument, beyond the adoption of better online tools, or the provision of a collection of "tips and tricks" for better practice. The challenges here are systemic and human: they involve the dispelling of myths, engagement and negotiation of individuals with varied and often conflicting interests, the coordination of practices across multiple units with varied perceptions of the purpose of the activities involved, awareness raising, and above all, culture change. Faith in our SRI-driven teaching evaluation practice is uneven, and many are reluctant to invest the resources that would be required for wholesale adoption of a new approach. The development of new tools is the easy part: what is considerably more difficult is creating a compelling, system-wide belief that these changes are necessary and possible, and that the efforts will be worth it.

demands, and the capacity to assess, identify and implement targeted improvement in instructional practice will be critical to competitiveness.

Change Process Considerations

There is considerable literature exploring the slow pace of change in universities and the challenge of leading change in this sector. As Fullan and Scott (2009) put it, many universities are typically "change averse" (p. 33). Among the challenges they identify are: unresponsive administrative structures; unaligned decision-making, accountability, and reward systems; limited change implementation strategies; and approaches to leadership selection that are at times unsystematic. In addition, universities are not monolithic: many who live and work within them view their core allegiance as disciplinary, and certainly antithetical to any centralized managerial culture (Silver, 2003). When one adds to this the multiple layers of conflicting interests involved in practices like teaching evaluation, the need for a sustained, gradual change leadership is clear. There are several critical considerations.

1. **Work on the basis of what is known about effective change practice**. Heath and Heath (2010), for example, provide a well-researched but accessible articulation of the fundamentals. These include:
   - identifying, investigating and duplicating anomalous "bright spots" in contexts where most are not thriving;
   - scripting only the critical steps in order to maintain flexibility;
   - engaging with people at the emotional as well as intellectual level;
   - breaking down changes into small, doable parts;
   - cultivating growth mindsets and identities; and
   - changing the situation and context in order to change behaviours.

2. **Understand that the provision of more and better information does not guarantee better decision-making**. As Diamond et al. (2014) demonstrate thoroughly in their recent review of literature about decision-making and the provision of information in the UK higher education context, in conditions of uncertainty (either too much, or too little, information) people tend to fall back on heuristics or mental short cuts in their decision-making. Further, they are generally largely unaware that the decisions they are making are based on insufficient information. This means that the creation and implementation of enhanced data collection, and analysis and reporting tools, in and of itself, is unlikely to produce change.

3. **Recognize that universities are complex systems made up of multiple interdependent subsystems, and that individual stakeholders within those subsystems function and make decisions based on their own understanding of their context, their interests, and what they perceive will protect the needs and rights of those within those contexts** (Hannan & Silver, 2000). New initiatives must therefore take into account multiple perspectives and move stakeholders gradually towards new understandings of what is in their interests (Sterman, 2006). Further, interventions that do not take this into account are likely to result in unexpected and unintended side effects (Sterman, 2006 ; Senge, 1990; Meadows & Wright, 2008).

> Gathering that information, or even effectively reporting it, isn't the true game changer: it's the distributed capacity to read it, analyze it, critically assess findings based on it. To plan, evaluate, and refine initiatives based on it.

4. **Recognize that externally imposed and mandated accountability measures without a plan for multi-level engagement are likely to result in minimal compliance, and will not have "begun to address the improvement or maintenance of the quality of academic work"** (Martin cited in Gosling & d'Andrea, 2001, p. 127). Strategic planning for post-secondary change must engage at a minimum with leadership, governance structures, technologies, the interests and concerns of individual instructors, and other stakeholders and interest groups in a sustained and responsive fashion in order to create the necessary cultural changes to truly integrate data-driven inquiry into every day decision making and practice (Whitehead, 2013).

5. **Understand that significant change requires resources:** time, tools, people, and money. Moreover, investment is a critical message in establishing that teaching evaluation is a priority practice.

V

Recommendations

Our study identified four clear and compelling practices that impact effectiveness across multiple components of teaching evaluation programmes: the importance of multi-dimensional data, strong feedback loops, shared understandings of quality, and leadership/engagement/change management. Evidence from the Ontario context indicates needs in all of these areas. In establishing our proposal, we also considered what stakeholders identified as their perceptions of their needs; whether a possible direction offered optimal, evidence-based traction for teaching improvement; the nature of university change processes; the importance of respecting collective agreements and institutional governance; and opportunities to enhance productivity and efficiency.

Establishing a provincial framework to guide teaching evaluation would articulate a clear set of values, principles, and purposes for teaching evaluation, and assist universities in identifying areas of their practice that need improvement. It could also provide needed traction in making those changes happen as well as engage institutions in important dialogue about how and what kinds of teaching fit into their strategic missions. Given the premise of a common framework against which institutions would map their own context-specific approach, it could provide a degree of consistency while also respecting institutions' unique mandates and cultures: such a framework could offer sufficient flexibility to articulate standards for teaching and methods of evaluating them based on strong evidence and fit with context.

However, given (1) the change considerations listed above; (2) the evidence of effective practice elsewhere and our current context; and (3) the steering team's experience of institutional change processes, our view is that the mandatory imposition of frameworks, tools, and guidelines is not the place to start. This is a process that in the end will thrive or fail based on the engagement of those it affects. Nothing in the literature suggests a proven "silver bullet" approach: much of the practice focuses on shared understandings, agreed upon definitions, and informed, context-specific judgments. These require sustained, consultative engagement and customization at the institutional (and in some cases at more granular) level. The design of better tools, or even the imposition of a well-informed policy framework, can help, but will not by itself result in the depth of change sought.

We therefore recommend a phased, incremental, inquiry-based approach that **engages in a sustained fashion with stakeholders at a range of institutional levels.** This approach should involve **building tools**, **establishing momentum and informed engagement**, and working towards the kind of **data- and dialogue-driven consensus** that will ultimately form the basis for a common teaching evaluation framework and a new era of effectively valued teaching in the province of Ontario. This requires both institutional and province-wide change and dialogue.

We **propose the establishment of a teaching evaluation consortium**, possibly organized as a virtual centre for excellence, which would coordinate projects; expand the provincial research base; promote informed thinking and practice of teaching evaluation; liaise with international initiatives, the college sector, and

stakeholder groups about teaching evaluation; and lead knowledge transfer, expertise development, and ongoing dialogue about the nature and dimensions of effective teaching and its documentation in the Province of Ontario. **The consortium would lead and coordinate the following preliminary projects responding to urgent needs within the university sector**:

1. **The development and piloting of a guided, customizable electronic teaching dossier prototype.** The proposed teaching dossier includes the development of other tools for data collection and analysis, including surveys, peer review procedure manuals, reflective guides, guides to the analysis of SRI data, as well as tools for the evaluation of the teaching dossiers produced.
2. A funded Ministry call for coordinated projects to **enhance inter-unit strategic collaboration for the purposes of enhancing the use of teaching evaluation to improve instruction**.
3. The development and piloting of **tools that enable better analysis and visualization of SRI data by individuals and by departments**.

Building Capacity and Engaging Stakeholders: The Ontario Teaching Evaluation Consortium

Rationale: There is considerable evidence of institutions and Ontario stakeholder groups researching and working for change in teaching evaluation practice. At present these efforts are essentially invisible to one another, and we are not benefiting from the knowledge exchange that might otherwise be possible. The preliminary suite of projects, described above, focus on core elements where Ontario practice is not well aligned with effective practice: the impact of these projects will be considerably greater if they function as a coordinated unit and are able to benefit from each other's work and learning. Because effective teaching evaluation involves highly integrated elements of practice, each of these projects needs to at the very least be aware of the emerging findings and outcomes of the others.

It is clear that teaching evaluation is a challenging mandate at universities. If culture change is to occur, we must nurture leadership and expertise, at multiple levels, in institutions: lateral connections with like-minded leaders are of significant benefit to individuals whose efforts may otherwise isolate them. The Consortium will also produce a series of research-based practical guides to promote fundamentals and raise awareness of research-based approaches that have been shown to be most effective. These may provide improved traction and easily accessible information for those seeking to move institutional agendas forward.

A teaching evaluation consortium and its evolving expertise will be the most credible source for the eventual development of a teaching evaluation framework for the Province, as well. The Consortium speaks particularly to the need for **sustained, informed, multi-level leadership**. One of its core mandates will be the coordination of institutional and inter-institutional inquiry and development in teaching evaluation across the province. This mandate can incorporate work with individuals in many different roles across the university sector.

Description: The Ontario Teaching Evaluation Consortium (OTEC) will enhance the culture of improvement- and inquiry-based teaching evaluation across the province by:

- promoting effective teaching evaluation in the province of Ontario, including the establishment of policies, practices, and principles; effective documentation; reporting; mediation; decision-making; and the use of teaching evaluation data for strategic planning of instructional improvement programmes;
- leading research initatives in what is currently an under-researched area in Ontario;
- coordinating and disseminating the findings of research and development initiatives around teaching evaluation in the province;
- establishing networks of collaborative

leadership in the province;
- improving efficiencies across the university sector by coordinating the development of joint and complementary initiatives across the province; and
- developing and creating adaptable first-generation tool sets that facilitate the adoption of improved and more consistent practice at Ontario universities.

OTEC will take a global and cross-sectoral view of teaching evaluation, seeking to learn from international and college-sector evidence and to understand how those findings do or do not apply to the Ontario university context. OTEC will support the expansion of successful pilots and initiatives to other Ontario institutions, and in general promote ethical and effective practice in teaching evaluation with the ultimate goal of establishing Ontario as an international leader in system-wide teaching evaluation.

OTEC will function most effectively with multi-stakeholder leadership to facilitate communication and negotiation among groups with diverse interests, and must involve OCUFA representation. The majority of the membership of OTEC will be draw from faculty and staff involved with teaching evaluation in the province. It would also draw board of directors representing multiple stakeholders from, for example OCUFA, HEQCO, MTCU, the Quality Council, and the Ontario Student Alliance. Student input is a critical element of progressive policy and practice in teaching evaluation. Based on this organizational structure, OTEC will also become a source of critical expertise about the policy and procedural contexts and stakeholder interests that must be addressed in navigating a way forward.

Benefits: As independent bodies, consortia invite the involvement of stakeholders from multiple levels and roles. They can provide strong knowledge management, ensuring the capture and the dissemination of learning from multiple projects and groups in order to truly establish an Ontario-specific body of expertise. They allow for a multi-partisan approach, and enable high degrees of consultation and incentivized change if appropriately structured. They allow for cross-institutional collaboration and network development, and for the development and management of an ongoing agenda for change in the province. A consortial approach also facilitates potential province-wide licensing agreements should that be deemed appropriate at a future date, as well as large-scale research and development collaboration nationally and internationally.

Risks: If these projects are to build momentum and change thinking in the province, a consortial approach is the best opportunity to ensure a high degree of dialogue among universities. However, establishing and maintaining that dialogue will require support for the consortium's work at institutions across the province. A parallel process of engaging senior administration with regard to this important mandate, enabling them to inform the establishment of the consortium, is critical, but ultimately the independence of the consortium from any specific stakeholder interests is also fundamental. In addition, strong and sustained input and collaboration with OCUFA and OUSA are critical to establishing fruitful dialogue that leads to real change in the province. Drawing on effective practice in the Australian context, the consortium might also seek discipline-specific cross-communication, for example through associations of deans in various subject areas, as well as with provincial organizations such as the Council of Ontario Educational Developers, a provincial organization of educational developers. The development of a strong teaching evaluation culture in Ontario is a long-term project: funding for the consortium must take this into account. Three initial projects to be coordinated through the Consortium follow.

Project 1: Building Tools for Multi-faceted Documentation and Reflection on Teaching: The Electronic Teaching Dossier

Rationale: Effective and broad-based use of teaching dossiers is the **best developed and most researched approach to using multi-faceted data to document teaching effectiveness**. It is also a standard practice in a number of jurisdictions with more sophisticated and holistic approaches to teaching evaluation. In terms of **balancing evaluation for personnel decision making and for instructional improvement**, the dossier is a strong tool for promoting reflective practice: the development of the dossier is frequently an opportunity for feedback and teaching dialogue in and of itself.

The teaching dossier is the second-most used teaching evaluation practice in the province, but in nearly all cases, instructors must produce and submit hard-copy documents, which are then copied and circulated, a considerable investment of resources. Survey respondents indicated interest in an electronic dossier development tool: the tool therefore has a natural market. A small number of universities are experimenting on a small-scale with standard e-portfolio tools for the purpose, which, though adequate to the task, do not provide the ease of use, contextualized guidance, or uniformity of format of a dedicated tool. A custom tool for teaching dossiers is **a directly applicable, easy-to-use tool that will significantly improve the ease and efficiency of building, maintaining, publishing, and sharing teaching dossiers**. As it evolves, it could function much like the common curriculum vitae (CV) in terms of providing a uniform approach to documenting teaching across the province, and allow for the creation of dossiers that, if the tool is widely adopted, could be portable from institution to institution with minor revisions to accommodate differences in institutional requirements.

The development of this tool has **considerable potential to enhance productivity and improve practice across the university sector** as it will provide a strong degree of mediation to instructors on an as-needed basis, potentially reducing the amount of one-on-one consultation required for dossier completion. In addition, the coordinated provision of guided materials for teaching dossier completion reduces duplication of efforts at participating institutions. Over time, the system may produce broad acceptance of web-based dossiers, reducing the need for material duplication and allowing for stronger, more representative multimedia data use.

Description: The electronic teaching dossier will be **a customizable electronic record of an instructor's practice**. It will include tabbed sections for each of the elements of teaching dossiers required or recommended by Ontario institutions (such as a teaching philosophy, description of teaching practices, course syllabi, student evaluation of teaching data, teaching-related publications, etc.). Documents and **multiple forms of institutional data will be importable** in a range of formats, or **can be produced directly within the dossier**. The e-dossier will be multimedia compatible: instructors will be able to include video and photographic evidence as well as written documentation. The electronic teaching dossier will also be exportable as a word-processing or PDF document. Instructors will be able to give other access to their dossier if they wish to do so. Visualizations from the SRI data visualization tool (Project 3) will be easily importable into the e-dossier.

In addition to its greater ease of use as a "bespoke" teaching dossier tool, a critical difference between this tool and a generic dossier generator is that it will be **framed by an extensive compendium of materials to support the instructors' skills in gathering, analyzing, reflecting on, and writing about their teaching data.** In effect, we are proposing an electronic guide that provides just-in-time support to instructors as they develop their dossiers. In the teams' extensive experience, teaching dossier development is a critical opportunity to rethink and reframe teaching experience. The guide will provide questionnaires, tools for data gathering, reflective prompts, and example materials from actual faculties' dossiers, with materials optimally licensed from many of the best in the field, for example, Pratt's *Teaching Perspectives Inventory* (http://www.teachingperspectives.com/drupal/), materials from Chism's *Peer Review of Teaching*, and CLASSE (http://nsse.iub.edu/_/?cid=211), the classroom-sized version of NSSE, and Angelo and Cross' (1993) *Classroom Assessment Technique*.

The tool will also **provide guidance, potentially including video-based instruction, for those reading and reviewing dossiers for both decision-making and teaching improvement purposes**. Chism (2007) provides a productive range of approaches to both formative and summative dossier evaluation, noting that agreed upon criteria and approaches (potentially including narrative response, discussion formats, checklists, and rubrics) to ensure consistency in feedback and decision-making. **Sample evaluation rubrics** can be provided, though of course all evaluative materials will be at the determination of the individual institution. It will also include ways to contact educational developers or others at the institution with expertise in this area and the responsibility to assist people in dossier development. Dossiers will remain the sole property of their authors: instructors can either publish versions for distribution or permit individual access to the dossier online.

In order to ensure consistency with individual institutional requirements, **many features of the dossier will be customizable at the institutional level**, for example the mandatory components of the dossier, access rights, and the content and provision of guide materials. All institutions will be able to create links to their own policy and guideline materials within the Guide. The tool will be institutionally brandable but could also ultimately be available for open, individual use without institutional involvement.

The proposed plan: We propose a Ministry call for a small number of institutional teams willing to take part in an iterative design and trial pilot. One institution will function as lead, with a technical team engaged in the actual tool development, while institutional teams will beta test and then pilot the e-dossier with small cross-disciplinary faculty teams. Institutional teams should include an administrator familiar with promotion and tenure procedures, an educational developer, several faculty members willing to work with the prototype, and technical support staff. If possible, there should be a senior-level administrator with a teaching and learning portfolio and/or a representative from the quality assurance office associated with the project at each institution. The project will produce a functional prototype, and an assessment of the benefits of moving to an RFP for full development or continuing to work with the existing prototype. Lead representatives from this project would also liaise with the coordinating team for the longitudinal and aggregate data project, so that tools developed in that project capture options for SRI data reporting within the e-dossier. The team will also collectively explore and develop institutional planning for the promotion of dossier use. Further details about the envisioned e-dossier can be found in the technical report in Appendix G.

Benefits: Although designing a teaching dossier will always require serious intellectual effort, the e-dossier tool will make teaching dossiers **easier and faster to develop, use, maintain, and review**, and will make them both **better and better understood**. It also provides instructors with a multi-media storage repository for all of their teaching-related artifacts, which, used over time, will **enable instructors to more efficiently document their practice**.

The guide materials offer instructors the kind of **support and mediation** they would receive if they attended an intensive teaching dossier academy, an opportunity many people do not have, and which they would now be able to **access on their own schedule**. The tool will also allow instructors **access to their dossiers from any location**. This cost-effective model reduces duplication of effort, but also allows for individual institutions to customize suiting their own needs. Given widespread adoption the e-dossier tool could function in a manner similar to the Canadian Common CV, allowing for an essentially common format and interface across the Province. The project also has significant value as a support for graduate students as they seek to document and promote their growing professional skills.

This project is in keeping with the proposed incremental, phased approach: this is a project with limited technical infrastructure that could significantly impact practice across the province, and one with readily understandable practical value to the average faculty member or administrator. It is a step forward, but one they can still "see themselves" in. This project has significant implications for propelling institutional practice towards a greater emphasis on the **use of multi-faceted data** and decision-making, and also for **teaching evaluation practices that connect evaluation with instructional improvement**. Governmental engagement with this initiative will **strongly communicate a new emphasis on improvement-oriented evaluation in the Province**.

Consistency among dossiers has often been noted as a challenge in evaluating dossiers. This tool will produce a significantly more consistent product while still allowing instructors a degree of flexibility in the structure and design of their final product. In the long run, a common but customizable tool specifically designed to meet the requirements of Ontario universities allows for improved uniformity in dossier development both at single institutions and for those moving among institutions. This consisteny may also facilitate informed decision-making for hiring and promotion.

Risks: Teaching dossiers are **labour-intensive** and the expectation that every faculty member at every institution and at every career stage will develop one is, at the present time, unrealistic. Although teaching dossiers

are mandatory or recommended at approximately two-thirds of universities, a key challenge of this project is in working towards broad adoption of the tool, which will require expert knowledge of the right champions in specific institutions, multi-institutional development, and broad-based promotion of the project. Consortial promotion of the tool at multiple levels will be a key lever here. Although this project does not pose significant technical challenges, it does require a strong level of user-interface design and customizability to ensure that the tool is efficient, effective and enjoyable to use.

2. Building Momentum, Knowledge, and Consensus: Integrating Evaluation With Instructional Improvement

Rationale: This project focuses on **one of the most critically needed areas of improvement among Ontario universities**. Research demonstrates that **in order for SRI and other teaching evaluation data to impact teaching practice, instructors must receive the data in well mediated ways**, talk about their data with others, and receive useful information about how to move from feedback to action. Currently, our survey data indicates, **Ontario universities are not "closing this loop" at all effectively**, meaning that there is every chance that we are simply not getting value for the teaching evaluation we do in terms of improving instruction. One challenge is that the institutional units involved in implementing teaching evaluation often do not have significant interaction with those producing the reports, who in turn may have little to do with those working to improve instruction on campus. Further, all of these groups may not have sustained connections with the department heads and deans who may be the first line of communication about teaching evaluation data. It is little wonder that three of the stronger examples of change in teaching evaluation practice have brought responsibility for these activities together within teaching and learning centres, and while institutions take different views about that arrangement, the fact remains that **effective practice requires much greater coordination among these units as well as innovative approaches to offering feedback both efficiently and respectfully** given the challenge of resource allocations. This project speaks to the need for **improved feedback loops** and **better connections between teaching evaluation and teaching improvement practice,** and may also significantly contribute to the **establishment of shared understandings of quality**, minimally at the institutional level, but potentially more broadly.

Description: We propose a Ministry call for coordinated projects to enhance strategic collaboration on campuses among units gathering and reporting on SET data, centres for teaching and learning, quality assurance offices, information technology units, and relevant senior administrators: in short, all those involved with the reporting of teaching evaluation data and its potential use by faculty to improve their teaching. The call might include a range of "effective practice" examples from other institutions to broaden and inspire creative thinking at the institutional level. Criteria would include projects that involve:

- initiatives that integrate improved feedback loops into SRI reporting and provide tie-ins between SRIs and instructional improvement programming;
- improvement of institutional reporting to individual faculty members and departments about SRI data;
- institutional consultations to establish explicit teaching quality statements, definitions, and criteria;
- design and publication of professional development materials on understanding and making judgments based on multi-faceted teaching data;
- the establishment of mechanisms use of institutional teaching data for strategic planning in teaching and learning; and
- initiatives to explore the potential programme-level teaching evaluation.

Projects could involve technology-driven solutions, but that would not be mandatory. At the cohort level, a leadership team from among the projects, in partnership with Consortium representatives, would work together to coordinate findings from the studies to produce a **report and guide on connecting teaching evaluation to instructional improvement**, based on all the studies. Such documents, as they evolve, can form **core elements of an eventual teaching evaluation framework** for the

Province. All tools designed during the project would be made available for provincial use.

Benefits: This project **addresses a fundamental structural challenge** in teaching evaluation which universities must solve if practice is to improve: a basic lack of coordination among those who manage teaching evaluation, and those who manage instructional improvement. Secondly, it **builds leadership capacity and real expertise** in institutions, while allowing for a broader dialogue among institutions. Thirdly, these projects will **offer evidence that teaching evaluation is not perceived purely as a personnel decision tool** in the province: in fact, **they will directly contribute to changing that**. This is probably the element of teaching evaluation practice that can most profoundly impact teaching improvement practice. Successful projects could have significant impacts on institutional cultures and perceptions. In general, universities are not currently effective in this area: a collective, efficiency-oriented and value-added approach would have significant advantages. As is clear from our research, projects related to teaching evaluation require sustained commitment and engagement with the communities involved: these projects offer a starting point for establishing these critical dialogues.

Risks: It is critical that the project teams consult carefully with their faculty associations, are knowledgeable about collective agreements and policy as they pursue these projects, and that the individual privacy rights of instructors are respected. It may be of benefit to launch these with a joint event exploring change management, the goals of improvement-oriented evaluation, and other matters related to optimizing the success of the projects. In general, campuses will not be culturally ready for "enterprise solutions" to this challenge: targeted, impactful pilots will provide proof of concept, paving the way for more extended initiatives.

3. Building Tools to Inspire Data-driven Dialogue: Visualization and Analysis Tools for SRI Data

The aims of this project are:

- to **improve individual faculty members' ability to explore the patterns and implications of their own SRI results**, and report and represent those data in more effective ways for promotion and tenure; and
- to create **effective ways for department heads to analyze programmatic SRI data**, identify patterns, and set future directions for programme and instructional improvement.

Rationale: Generally, SRI data have been used more for purposes of evaluation than actual analysis of student perceptions of teaching (Üstünlüoğlu & Güngör-Culha, 2012). SRIs are only one source of data about instructional activity at universities. Although limited in scope, however, they are ubiquitous: institutional SRIs produces millions of data points every year which universities internationally currently put to limited use (Hénard & Roseveares, 2012). Only 18% of Ontario universities indicated that they were using these data in the aggregate or for instructional improvement.

We can accomplish much more with the data SRIs provide. We can make it possible for instructors and department heads to parse and explore these data using **powerful analytical and visualization tools** which enable effective inquiry and representation of student perceptions of teaching and courses at the course, instructor and programmatic level.

We can create tools which allow instructors to study and reflect on their own SRI data more effectively: to view graphic representations of their own SRI data across multiple years or multiple types of groupings, such as every instance of a given course, of every instance of a course over a given size, or with a focus on specific items or validity-tested clusters of items. Ludlow (1996, 2005) demonstrates that **this approach can have positive effects on faculty members' engagement with these data**, their willingness to make changes to practice, and their capacity to effectively represent their work for promotion, tenure, and performance review.

Departments can use the data in more varied and comprehensive ways, for example, to draw out patterns and commonalities that might inform decision making (AGO, 2012), enable more targeted instructional development intervention, and potentially inform curriculum and course re-design efforts (Joughin & Winer, 2014).

Institutions can also use aggregate data to more carefully parse variations and commonalities in response patterns in order to **more accurate identify norm groups**, to **identify and address bias**, and to **inform decision making** that takes these limitations of student feedback data into account (Joughin & Winer, 2014; Hativa, 2013a) Finally, we can use aggregate data **to combat myths about SRI on our own campuses**: at one Canadian institution, SRI reporting always includes means for large enrolment classes and other disaggregations that are believed to "make a big difference" in teaching scores. In most cases, it doesn't, but the provision, year after year, of the evidence that there is no significant difference is a powerful way to dispel such myths (Joughin & Winer, 2014). And in cases where it does make a difference, this level of transparency allows for a greater degree of trust across the board. In order to make any of these possible, we must provide faculty and administrators with **better tools and expertise** so that they can parse and evaluate these data more effectively and carefully.

At present, these practices may seem rather foreign to most institutions' practices. They do, however, have many antecedents. At the informal level, educational developers working with faculty on teaching dossiers have for years encouraged them to analyze their data in this manner manually, a painstaking and tedious process but often worth the effort (Ludlow, 1996, 2005). IDEA, an American third-party SRI provider, offers users a range of ways of examining institutional aggregate data, as does Explorance Blue, a Canadian competitor. **In other jurisdictions, aggregate data usage at the programmatic level is common (Joughin & Winer, 2014): critically, these evaluative practices always involve multiple forms of evidence.** However, as the majority of universities are continuing to use paper and pencil evaluation systems, we need a tool that allows us to explore the possibilities of this approach based on the existing "back end" tables of data produced from those hard copies. As well, many institutions currently lack the "data wisdom" to apply these tools effectively: **this project would enable us to establish expertise and momentum, and to work towards changing faculty and institutional ideas about the potential uses of SRI**. As a consortial project, there is considerable potential also for the **development of a critical mass of academics and staff engaging in collaborative inquiry** regarding extended approaches to SRI use, a field with considerable potential which is under-theorized and under-researched. The "research angle" offers a strong draw for faculty engagement, as well.

One reason for identifying this approach is that while SRI instruments across the province appear to be in varying compliance with generally agreed-upon standards of SRI design, and while there is significant variation at times even within institutions, for the most part there is a degree of overlap in the SRI data universities gather. **SRI instruments are highly contentious and often highly regulated: changing them is a major political hurdle.** In terms of SRI development, there are numerous strong examples of third-party tools (See Appendix H), and also a fair degree of consistency among experts regarding appropriate items and design: **producing a document identifying recommended practices**, **providing guidance on approaches to assessing the validity and reliability of SRI instruments**, **or negotiating a consortial license for the use of a strong third-party tool which universities could choose to adopt**, might be a **less politically challenging approach to improving SRI instruments in the immediate to short term**.

More to the point, **mandating change to the forms will not produce better engagement with the data**. The approach we advocate **targets aspects of practice where intervention has the best chance of leading to improvement**: the use and visualization of data is a more powerful and less contentious lever for change that instrument re-design.

Description: The goal of this project is **to build a set of data analysis and visualization tools for use by individual instructors and by department heads**. For individual instructors, the tool would enable them to aggregate, disaggregate and annotate their collected (multi-year) data in a variety of ways. For example:

- Creating histograms that clump all instances of the same course together.
- Examining their individual item scores across all courses to identify patterns of strength and weakness.
- Tracking scores or subscores in a course taught repeatedly, over time.

- Comparing their course to similar courses (by size, course level, require or non-require status, etc.).
- Identifying "first-time" offerings.
- Disaggregating SRI data by student demographics.

Visualization is a critical tool here: it is nearly impossible to identify patterns from raw SRI scores. These tools can be of real value to faculty, both in formulating specific arguments about the nature and quality of their teaching for various kinds of performance review, and as the basis for ongoing critical inquiry into practice, growth, and areas for further development. Ludlow (2005) provides evidence that **this approach can have a positive impact on faculty engagement with teaching evaluation**.

There are many ways in which programme chairs would benefit from better and more analyzable SRI data reporting. Evaluation of teaching falls to programme chairs in more than one way: firstly in terms of **instructor performance review**, in particular for new faculty and for faculty undergoing promotion and tenure review. In the latter case, programme chairs are called upon to write about instructor teaching, often working with data that are not easily usable for analysis. The tools above would significantly improve this situation. However, there are other areas of potential significance. As Gibbs, Knapper, and Piccinin (2007) demonstrate, programme and department chairs are often the most critical lever for instructional improvement in universities. Using fairly technically simple but conceptually sophisticated tools, it is possible to **provide programme chairs with the ability to study visual representations of their programme's collective SRI data, organized in a variety of ways**. They could, for example, study the student response patterns by item in all first-year courses to identify areas of practice for collective inquiry and action, or seek insights into whether collective changes in practice were impacting students' perceptions of levels of interaction, course satisfaction, or sense of clarity about expectations. This data could also be used collectively: a group of instructors who all regularly teach the same course could study their collective data in the aggregate to identify overall patterns in student response and possible areas for course improvement.

What we propose is a funded Ministry call for a project to develop tools for analyzing and visualizing various types of aggregate SRI data and longitudinal individual data. The goal of the project would be **to build a prototype suite of data tools enabling more effective use of aggregate and longitudinal SRI data at the instructor and departmental level**. It might also explore other data sets that could be effectively and feasibly integrated into improvement-oriented analytical tools The lead team would draw on existing models from other jurisdictions, and confer with statisticians, business intelligence experts, educational developers, administrators, faculty associations, and university secretariats to build initial prototypes after consultation with teams from other universities. Each university team would then use the data modeling tools in a pilot study with iterative prototype redevelopment phases in order to assess the value of the tools. A further and important goal of the study would be the establishment of preliminary ethical guidelines for aggregate SRI data use. The project team could also liaise with the e-dossier team to identify promising approaches to modeling SRI data for dossier presentation, and with the Quality Council and institutional quality assurance offices with regard to potential implications for institutional programme review reporting. One aspect of the project will be **exploration of how SRI data should best be integrated with other forms of evidence in exploring teaching effectiveness**: one product will be a report outlining initial suggestions for effective practice in this area. Both the technical report and the aggregate SRI data reports (Appendices H and I) provide fuller explorations of the potential of SRI and of the proposed process for moving forward.

Benefits: This project directly **addresses the issues of the usability, validity, and purpose of SRI use** while refocusing the conversation, at least to a degree, on matters beyond personnel decision making. It provides faculty and department heads with more room to maneuver in presenting and exploring the nuances of teaching data, and **more opportunity to contextualize and effectively present this element of the evidence regarding their teaching**. The existing evidence suggests that opportunities to engage with these data more constructively can **have a positive impact on faculty engagement with instructional improvement and inquiry** into their own teaching (Ludlow, 1996, 2005).

At present Ontario universities have very little experience with using these data in this fashion: this project will be **a prime opportunity to expand dialogue and thinking about teaching evaluation data in the province**. Effective data analysis will also enable us to provide **clearer evidence of the validity and limits of SRI data on Ontario campuses**, on an institution-by-institution basis. This project may also have significant benefits in terms of improving reporting on programmatic teaching quality for quality assurance purposes.

Risks: There are some obvious potential risks involved in the wholesale adoption of aggregate SRI data analysis, risks that are in essence **a magnified version of the many challenges and tensions involved in all SRI activities**. Firstly, the data involved are highly sensitive, and impact people's professional lives in significant ways. Just because it is possible to calculate, for example "the ten worst SRI scores on campus" does not make it constructive, statistically valid, or ethical. Further, tools that facilitate comparisons **must be accompanied by mechanisms that limit the capacity for misinterpretation, bias, and the drawing of inappropriate conclusions**: a limit that can and will be designed into the system based on a strong level of consultation with faculty and administrators throughout the project.

It is clear that there are **ethical issues to be addressed** in the development of aggregate data tools: **we must find ways to establish practices that are collegial, appropriate, respectful, and beneficial**. Some preliminary fundamentals, however:

1. Practices must be in keeping with the ethical principles of the university as well as all policies and collective agreements.
2. Practices should be consistent with the stated purposes for which data has been gathered
3. Practices must be respectful of instructors as the actors within the field and in some cases as the owners of the data
4. Practices must emphasize and make clear the limitations of data and tools, and limit user capacity to draw invalid conclusions where possible.
5. Practices must be based on classification of data in terms of access rights, and must also respect the need for confidentiality. Where possible, data should be anonymized.

(Based on Slade & Pinsloo, 2013)

The **development of ethical guidelines for SRI data use** is a fundamental element of this project.

As well, **it must be clear from the beginning that these tools are not intended to allow comparisons for which there is no statistical basis, such as comparisons across institutions**. As Hativa (2013a) points out, differences in institutional populations as well as culture make such an activity unviable. Further, it is entirely counterproductive in terms of establishing a productive culture of teaching improvement in Ontario universities. There is among faculty **considerable distrust regarding institutional and provincial willingness to misuse these data**, data which, despite considerable evidence to the contrary, many faculty dismiss as invalid and unreliable. It is critical that **enhanced data use show a meticulous attention to emphasizing what is constructive, and helpful, and truly improvement-focused** as this initiative evolves. **Tools for personal use, or for use with educational developers by permission of the instructor, are likely to be a useful entry point**. It is particularly important that faculty members and educational developers drive the direction of this project, with strong degrees of consultation.

This is a project that requires **detailed attention to the regulation of SRI data use at various institutions**, a task that would be undertaken at each institution in consultation with its administration, faculty association and university secretariat. The tools would be built to accommodate customized levels of permission and access to data in order to ensure that institutions would remain in compliance with their own regulatory contexts. **They will also be built for implementation at the institutional, not sectoral, level.**

For more information on the potential and challenges of aggregate data use in Ontario context, please see Appendix I.

Teaching Quality and Teaching Evaluation: An Agenda for Further Inquiry

In its initial call, the Province sought a project that would inform the development of a framework and information management approach for "collecting, managing, and analyzing appropriate data to develop and strengthen the cycle of continuous improvement of teaching quality" in Ontario. Among the possible topics suggested were: developing and norming evaluation questions; focus groups with faculty and students; system selection; software installation; system integration with campus information systems; and stakeholder communication/ engagement. The outcomes of our project provide substantive and valuable direction in a significant number of these areas, and the recommended actions provide an integrated suite of initiatives that will strengthen and deepen our teaching evaluation culture in important ways. At the same time, it is important at this point to reflect carefully on the complex and nuanced connection between teaching evaluation and teaching improvement and to consider where emphasis should be placed in the long term if the fundamental goal is a cycle of continuous improvement of teaching quality in the Province.

> The quality of a student's learning experience is an ensemble production.

Research in international jurisdictions has foregrounded some significant differences between the Ontario approach and the approach taken elsewhere. Hénard and Roseveare's (2012) extensive OECD study of policies and practices internationally which foster quality teaching in higher education provides three critical lessons:

1. We must see teaching evaluation as part of a larger framework of commitment to quality of student experience and the ongoing enhancement of that experience.
2. The quality of teaching is perhaps better understood as one factor contributing to the quality of educational programmes: the greatest potential for change is seen at the programmatic level rather than the individual.
3. Quality teaching and its improvement only work as a part of an institutional strategy that has to be implemented at all levels (for example student services, policies, infrastructure) – all need to be involved.

The quality of a student's learning experience is an ensemble production. Beyond the people involved, who stretch far beyond the admittedly important primary relationship with an individual instructor, policy, infrastructural, and procedural contexts are an integrated system that impact how people teach and learn. Given this reality, if the intention is to inspire evaluation that improves instruction and learning, one has to wonder whether a narrow focus on individual teaching practice will really identify the most promising levers for effecting change in a given institution. As Hénard and Rosevaeare (2012) put it, "evaluating quality teaching needs to be seen within the broader institutional context, closely linked to quality assurance mechanisms and supported by the development of suitable measurement tools that are robust, reliable and meaningful" (p. 37). This is a much more complex, but potentially fruitful challenge prompting numerous questions: What are the 'borders' that define the teaching environment that impacts students? Who are the players? What institutional policies and procedures are impacting instructors' choices and actions? What kinds of physical, technical, and social infrastructure should be taken into account?

What the Ministry is seeking, in terms of a standard framework for the improvement of teaching quality, may be a better fit with an improved version of programmatic quality assurance, rather than an improved version of traditional North American approaches to teaching evaluation. Such an approach would emphasize, to a much greater degree, collective and collegial inquiry about programmes' instructional contexts and students' learning experience within those contexts, comparable to existing and emerging practices in the Australian and UK contexts. Our quality assurance framework is a relatively recent addition to the Provincial practice: so far it places a very limited emphasis on true, multi-faceted

explorations of the quality of the learning experience or indeed of teaching. In identifying this as a possible direction, we must emphasize the critical importance of using multi-faceted, context-rich data for these purposes, rather than relying on uni-focal sources such as SRI data and other quantitative measures of inputs and outputs. As we expand the practice of teaching evaluation to this broader and more complex arena, the same lessons apply: effective practice involves multi-dimensional data, strong feedback loops, shared understandings of quality, and strong, sustained leadership. So, although quantitative measures are important elements of an overall picture, effective models of practice integrate quantitative and qualitative data, as well as nuanced understandings of systemic impacts on the learning environment. These practices require considerable skill and discernment if they are to create strong representations of what, in the end, is a highly complex social practice (Henand & Roseveare, 2012; Sachs, 2012). These issues require further research and reflection.

The teaching evaluation frameworks that seem best to reflect this potential were built with strong, sustained engagement from multiple agencies, institutions and stakeholder groups, and are predicated on a system with strong requirements for teaching expertise within the professoriate and a high degree of collegial engagement in quality assurance. That is not our current context. Right now, the imposition of such frameworks, without the considerable advance work we are recommending to create cultures that adopt and internalize the role of teaching evaluation in teaching improvement, is likely to be received as a purely regulatory action, resulting, as Hannan and Silver (2000) put it so cogently, in resigned, minimal compliance with very limited impact on practice or perception. And, given that resources are already stretched, a relatively sudden collective transition to a more robust approach to teaching evaluation will be a difficult burden for institutions to bear.

Our recommended programme of next directions is a critical foundational phase for the development of the infrastructure, expertise, and predispositions required to adopt a more strategic and systemic approach to the evaluation of teaching quality. It will support the growth of a culture that is ready to move in this direction, enable us to build and test tools that will make the direction more institutionally feasible, and significantly enhance the practice of teaching evaluation in the Province so that stakeholders begin to value it and use it in ways that make the engaged adoption of a future framework significantly more likely. In fact the dossier approach we are recommending is the model most consistent with the approach to programme review for teaching quality used in other jurisdictions: over time, and with an appropriate mandate, the tools built for the individual level could be further developed to meet these needs. Depending on approach, this tool could allow us to evaluate individual teachers, but also promote reflection at the programmatic and institutional level, enabling institutions to explore their instructional improvement practices from a systemic level, to identify and target specific areas for growth.

It is worth noting that all of the most effective cases we studied took a slow, incremental approach that built the model and the culture simultaneously, reciprocally. If the aim is to truly support and inspire a culture of teaching improvement, which inevitably requires a culture of teaching inquiry, a phased approach is critical. There is much that can be accomplished, and a number of incremental phases that can put us closer to the establishment of a provincial framework. In the meantime, there is a great deal of work to do in discerning approaches to teaching evaluation that truly meet the needs of all stakeholders in the province. On a final note, there is one more important lesson to draw from Henand & Roseveare (2012) study: this shift of culture, perception and practice is only truly likely to take hold in Ontario if universities are incentivized at the provincial level. Clear, sustained, and consultative leadership that integrates expertise, concerns, and needs from across the university sector is required in order to make this vision a reality.

References

Abrami, P.C. (2001). Improving judgments about teaching effectiveness using teacher rating forms. *New Directions for Institutional Research*, *109*, 59-87.

Alderman, L., & Melanie, L. (2013). Kaleidoscope: Reframing learning and teaching evaluation through a shared stakeholder lens. *Australasian Higher Education Evaluation Forum (AHEEF) 2013*, Launceston, TAS.

Alderman, L., & Melanie, L. (2012). REFRAME: A new approach to evaluation in higher education. *Studies in Learning, Evaluation, Innovation & Development, 9*(1), 33-41.

Alderman, L. & Melanie, L. (n.d.) *Kaleidoscope: Reframing learning and teaching evaluation through a shared stakeholder lens*. Retrieved June 9, 2014 from https://www.eiseverywhere.com/file_uploads/638148a66ce2cff5f504aaffc08fe0d5_Session5-LAlderman-Kaleidoscope.pdf

Amundsen, C., Abrami, P., McAlpine, L., Weston, C., Krbavac, M., Mundy, A., & Wilson, M. (2005, April). *The what and why of faculty development in higher education: A synthesis of the literature. Paper presented at the American Educational Research Association (AERA), Montreal, QC.*

Angelo, T.A., & Cross, K.P. (1993). Classroom assessment techniques. San Francisco: Jossey-Bass.

Arreola, R. (2007). Developing a comprehensive faculty evaluation system: A guide to designing, building and operating large-scale faculty evaluation systems. San Francisco: Anker Publishing.

Auditor General of Ontario [AGO]. (2012). University undergraduate quality teaching. In Office of the Auditor General of Ontario *2012 Annual Report (pp. 273-291). Toronto: Queen's Printer for Ontario. Retrieved from http://www.auditor.on.ca/en/reports_en/en12/2012ar_en.pdf*

Australian Government. (n.d.). Teaching Standards Framework [TSF] user guide. Retrieved from https://teachingframework.edu.au/tsftool/users/tsf_guide

Beran, T., & Rokosh, J. (2009). Instructors' perspectives on the utility of student ratings of instruction. *Instructional Science*, *37*(2), 171-184.

Beran, T., Violato, C., Kline, D., & Frideres, J. (2005). The utility of student ratings of instruction for students, faculty, and administrators: A "consequential validity" study. *The Canadian Journal of Higher Education, 35*(2), 49-70.

Berk, R.A. (2014). Should student outcomes be used to evaluate teaching? *Journal of Faculty Development, 28*(2), 87-96.

Berk, R.A. (2013). *Top 10 flashpoints in student ratings and the evaluation of teaching*. Stylus Publishing, Sterling VA.

Berk, R. A. (2009). Using the 360° multisource feedback model to evaluate teaching and professionalism. *Medical Teacher, 31*(12), 1073-1080.

Berk, R. A. (2006). *Thirteen strategies to measure college teaching*. Sterling, VA: Stylus.

Biggs, J. (2001). The reflective institution: Assuring and enhancing the quality of teaching and learning. *Higher Education, 41*, 221-238.

Boerboom, T., Jaarsma, D., Dolmans, D., Scherpier, A., Mastenbroek, N., & Van Beukelen, P. (2011). Peer group reflection helps clinical teachers to critically reflect on their teaching. *Medical Teacher, 33*, 615-623.

Bothell, T.W. & Henderson, T. (2003). Do online ratings of instruction make sense? *New Directions for Teaching and Learning, 96*, 69-79.

Boyer, E.L. (1990). *Scholarship reconsidered: Priorities of the professoriate*. Lawrenceville, NJ: Princeton, Carnegie Foundation for the Advancement of Teaching.

Braskamp, L.A. & Ory, J.C. (1994). *Assessing faculty work*. San Francisco: Jose Bass.

Buller, J. (2012). *Best practices in faculty evaluation: A practical guide for academic leaders*. Jossey-Bass: San Francisco.

Canadian Association of University Teachers [CAUT]. (2007). *CAUT Teaching dossier.* Retrieved from http://www.caut.ca/docs/default-source/professional-advice/teaching-dossier.pdf

Canadian Association of University Teachers [CAUT]. (2007). *Model clause on the evaluation of teaching performance.* Retrieved from http://archive.caut.ca/pages.asp?page=385&lang=1

Centra, J.A. (2000). Evaluating the teaching portfolio: a role for colleagues. *New Directions for Teaching and Learning, 2000*(83), 87-93.

Centra, J.A. (1993). *Reflective faculty evaluation*. San Francisco: Jossey-Bass.

Centra, J.A., (1977). How universities evaluate faculty performance: A survey of department heads. Graduate Record Examinations (GRE) Board Research Report, *75-5bR,* Princeton, NJ.

Chamberlain J., D'Artrey M., & Rowe, D. (2011). Peer observation of teaching: A decoupled process. *Active Learning in Higher Education, 12*(3) 189-201.

Chickering, A.W. & Gamson, Z.F. (1987). Seven principles for good practice in undergraduate education. *Association Higher Education Bulletin, 39*(7), 3-7.

Chism, N.V.N. (2007). *Peer review of teaching: A sourcebook* (2nd ed.). Bolton, MA: Anker.

Clayson, D. (2009). Student evaluations of teaching: Are they related to what students learn? A meta-analysis and review of the literature. *Journal of Marketing Education*, *31*(1), 16-30.

Clayton, J. (2012). *Mentors and reflective frameworks: Internationalizing the local provision of professional development in e-Learning in New Zealand*. Proceedings from the Fifth Annual Conference on e-Learning Excellence in the Middle East: Sustainable Innovation in Education. Hamdan Bin Mohammed e-University, Dubai, United Arab Emirates.

Codina, E. A. & Jiménez, E.G. (2008). Evaluation of teacher competence in Spain: The DOCENTIA programme. In EUA (Ed.), *Implementing and using quality assurance: Strategy and practice – A Selection of Papers from the Second Annual European Quality Assurance Forum* (pp. 17-22). Brussles: European University Association.

Cohen, P.A. & McKeachie, W.J. (1980). The role of colleagues in the evaluation of teaching. *Improving College and University Teaching*, *28*(4), 147-154.

Colbeck, C. (2002). Evaluating faculty performance. *New Directions for Institutional Research, 114*. San Francisco, CA: Jossey-Bass.

Commission on the reform of Ontario's public services [Drummond Report]. (2012). Retrieved from http://www.fin.gov.on.ca/en/reformcommission/chapters/report.pdf

Confederation of University Faculty Associations [OCUFA]. (2012). *Analysis of the Drummond Report: Long on cuts, short on Insight*. Toronto: Ontario Confederation of University Faculty Associations.

Diamond, D., Roberts, J., Vorley, T., Birkin, G., Evans, J., Sheen, J., & Nathwani, T. (2014). *UK review of the provision of information about higher education: Advisory study and literature review*. Report to the UK Higher Eduction Funding Bodies by CFE Reseeach, Leicester, UK.

Devanas, M. A. (2006). Teaching portfolios. In P. Seldin (Ed.), *Evaluating faculty performance: A practical guide to assessing teaching, research and service* (pp. 111-130). Bolton, MA: Anker Publishing.

Devlin, M., & Samarawickrema, G. (2010). The criteria of effective teaching in a changing higher education context. *Higher Education Research & Development, 29*(2), 111-124.

Doll, W.E., Jr. (1993). *A post-modern perspective on curriculum*. New York: Teachers College Press.

Donovan, J., Mader, C.E., & Shinsky, J. (2006). Constructive student feedback: Online vs. traditional course evaluations. *Journal of Interactive Online Learning, 5*(3), 283-296.

Dunn, L., Morgan, C., O'Reilly, M., & Parry, S. (2004). *The student assessment handbook: New directions in traditional and online assessment*. London: Routledge Falmer.

Fang, B. (2007). A performance-based developmental model for online faculty. *Performance Improvement*, *46*, 17-24.

Feldman, K.A. (1989). The association between student ratings of specific instructional dimensions and student achievement: Refining and extending the synthesis of data from multisection validity studies. *Research in Higher Education, 30*(6), 583-645.

Feldman, K.A. (2007). Identifying exemplary teachers and teaching: Evidence from student ratings. In R.P. Perry & J.C. Smart (Eds.), *The scholarship of teaching and learning in higher education: An evidence-based perspective* (pp. 93-143). Dordrecht, The Netherlands: Springer.

Foundation Coalition. (2001). *Assessment and evaluation*. Retrieved from http://www.foundationcoalition.org/home/keycomponents/assessment_evaluation.html

Fullan, M. & Scott, G. (2009). *Turnaround leadership for higher education*. San Francisco: Wiley & Sons.

Galbraith, C., Merrill, G., & Kline, D. (2012). Are student evaluations of teaching effectiveness valid for measuring student learning outcomes in business related classes? A neural network and Bayesian analyses. *Research in Higher Education, 53*(3), 353-374.

Ghedin, E. & Aquario, D. (2008). Moving towards multidimensional evaluation of teaching in higher education: A study across four faculties. *Higher Education, 56*(5), 583-597.

Gibbs, G. & Coffey, M. (2004). The impact of training of university teachers on their teaching skills, their approach to teaching and the approach to learning of their students. *Active Learning in Higher Education, 5*(1), 87-100.

Gibbs, G., Knapper, C., & Piccinin, S. (2007). The role of departmental leadership in fostering excellent teaching. *Leadership Foundation for Higher Education, 13.* Retrieved from http://www.kuleuven. be/duo/_pdf/JournalClubs/Gibbs_Knapper_Piccinin_the_role_of_departmental_leadership_in_fostering_excellent_teaching.pdf

Goodwin, L.D. & Stevens, E.A. (1996). The influence of gender on university faculty members perceptions of 'good' teaching. In D.E. Finnegan, D. Webster, & F. Gamson (Eds.), *Faculty and faculty issues in colleges and universities* (pp. 27-286). Needham Heights, MA: Simon & Schuster.

Gosling, D. & d'Andrea, V. (2001*). Quality development: A new concept for higher education*. Proceedings from the Sixth QHE Seminar: The End of Quality? Birmingham.

Gravestock, P. (2011). *Does teaching matter? The role of teaching evaluation in tenure policies at selected Canadian universities.* (Unpublished doctoral dissertation). University of Toronto, Toronto, Ontario.

Gravestock, P. & Gregor-Greenleaf, E. (2008). *Student course evaluations: Research, models and trends. Higher Education Quality Council of Ontario*. Retrieved from http://www.heqco.ca/ SiteCollectionDocuments/Student%20Course%20Evaluations_Research,%20Models%20and%20Trends.pdf

Hannan, A. & Silver, H. (2000). *Innovating in higher education: Teaching, learning and institutional cultures*. Buckingham: Open University Press.

Hansen L., McCollum M., Paulsen S., Cyr T., Jarvis C., Tate G., & Altiere, R. (2007). Evaluation of an evidence-based peer teaching assessment program. *American Journal of Pharmaceutical Education, 71*(3), 1-45.

Harvey, L. (2011). The nexus of feedback and improvement. In C.S. Nair & P. Mertova (Eds.) *Student feedback: The cornerstone to an effective quality assurance system in higher education*. Witney Oxford, UK: Chandos Publishing.

Hassna, L.O. & Raza, S. (2011). An assessment of the relationship between the faculty performance in teaching, scholarly endeavor, and service at Qatar University. *Research in Higher Education Journal, 10*, 1-18.

Hativa, N. (2013a). *Student ratings of instruction: A practical approach to designing, operating, and reporting*. Charleston, SC: CreateSpace Independent Publishing Platform.

Hativa, N. (2013b). *Student ratings of instruction: Recognizing effective teaching*. USA: Oron Publications.

Hativa, N., Barak, R., & Simhi, E. (2001). Exemplary university teachers: Knowledge and beliefs regarding effective teaching dimensions and strategies. *The Journal of Higher Education, 72*(6), 699-729.

Hassna, L.O. & Raza, S. (2011). An assessment of the relationship between the faculty performance in teaching, scholarly endeavor, and service at Qatar university. *Research in Higher Education Journal, 10*, 1-18.

Healey, M. (2000). Developing the scholarship of teaching in higher education: A discipline-based approach. *Higher Education Research & Development, 19*(2),169-189.

Heath, C. & Heath, D. (2010). *Switch: How to change things when change is hard*. Toronto: Random House Canada.

Hénard, F. (2009). *Learning our lesson: Review of quality teaching in higher education. Draft report*. Paris: OECD. Retrieved from http://www.oecd.org/edu/imhe/44146968.pdf

Hénard, F. & Roseveare, D. (2012). *Fostering quality teaching in higher education: Policies and practices*. An Institutional Management in Higher Education (IMHE) Guide for Higher Education Institutions. Retrieved from http://www.oecd.org/edu/imhe/QT%20policies%20and %20practices.pdf

Higher Education Quality Council of Ontario [HEQCO]. *Quality: Shifting the focus. A report from the expert panel to asses the strategic mandate agreement submissions*. Retrieved from http://www.heqco.ca/SiteCollectionDocuments/FINAL%20SMA%20Report.pdf

Iqbal, I. (2013). Academics' resistance to summative peer review of teaching: questionable rewards and the importance of student evaluations. *Teaching in Higher Education*, *18*(5), 557-569.

Joughin, G. & Winer, L. (2014). *Response to Teaching Evaluation Toolkit preliminary reporting*. Panel discussion, Toronto, ON.

Keig, L. & Waggoner, M.D. (1994). Collaborative peer review: The role of faculty in improving college teaching. *ASHE-ERIC Higher Education Report, 2*. Washington, DC: George Washington University School of Education and Human Development.

Kelly, M. (2012). *Student evaluations of teaching effectiveness: Considerations for Ontario universities*. Toronto: Council of Ontario Universities.

Kiceniuk, D. (2012). Student ratings of instruction (SRI) in higher education-exploring the issues. *FOCUS on University Teaching and Learning*. Centre for Learning and Teaching, Dalhousie University.

Knight, P.T. & Trowler, P.R. (2000). Department level cultures and the improvement of learning and teaching. *Studies in Higher Education, 25*(1), 69-83.

Knol, M., In't Veld, R., Vorst, H., Driel, J., & Mellenbergh, G. (2013). Experimental effects of student evaluations coupled with collaborative consultation on college professors' instructional skills. *Research in Higher Education, 54*(8), 825-850.

Lang, J.W.B. & Kersting, M. (2007). Regular feedback from student ratings of instruction: Do college teachers improve their ratings in the long run? *Instructional Science, 35*(3), 187-205.

Langbein, L. (2008). Management by results: Student evaluation of faculty teaching and the mis-measurement of performance. *Economics of Education Review, 27*(4), 417-428.

Lee, A. (2007). *From teaching to learning. Leading change at a large* research-*intensive university: A personal reflection.* Retrieved from http://www.guidelinesonlearning.com/sites/default /files/From%20Teaching%20to%20 Learning.pdf

Ludlow, L. H., (2005). A longitudinal approach to understanding course evaluations. *Practical Assessment, Research & Evaluation, 10*(1). Retrieved from http://pareonline.net/pdf/v10n1.pdf

Ludlow, L. H. (1996). Instructor evaluation ratings: A longitudinal analysis. *Journal of Personnel Evaluation in Education*, *10*, 83-92.

Marsh, H.W. (2007). Students' evaluations of university teaching: A multidimensional perspective. In R.P. Perry & J.C. Smart (Eds.), *The scholarship of teaching and learning in higher education: An evidence-based perspective* (pp. 319-384). New York: Springer.

Marsh, H.W. & Roche, L. (1993). The use of students' evaluations and an individually structured intervention to enhance university teaching effectiveness. *American Educational Research Journal, 30*(1), 217–251.

McLean, M., Cilliers, F., & Van Wyk, J.M. (2008). Faculty development: Yesterday, today and tomorrow. *Medical Teacher, 30*(6), 555-584.

Meadows, D.H. & Wright, D. (2008). *Thinking in systems: A primer*. White River Junction, VT: Chelsea Green Publishers.

Murray, H.G. (2005). *Student evaluations of teaching: Has it made a difference?* Paper presented at Annual Meeting of the Society for Teaching and Learning in Higher Education, Charlottetown, PEI.

Murray, H.G. (1987). Acquiring student feedback that improves instruction. In M.G. Weimer (Ed.), *New directions for teaching and learning* (No. 32). San Francisco: Jossey-Bass.

Organization for Economic Co-operation and Development [OECD]. (2007). *Assessing higher education learning outcomes: Summary of a first meeting of experts*. Retrieved from http://www.oecd.org

Ontario Confederation of University Faculty Associations [OCUFA]. (2013). *Update on freedom of information and protection of privacy issues.* Retrieved from http://ocufa.on.ca/?s=Update +on+Freedom+of+Information+and+Protection+of+Privacy+Issues+t

Ontario Undergraduate Student Alliance [OUSA]. (2008). *Measuring up: Quality metrics and accountability in Ontario's post-secondary system*. Retrieved from http://www.ousa.ca/ wordpress/wp-content/uploads/2010/12/Quality-Metrics-and-Accountability.pdf

Ontario Undergraduate Student Alliance [OUSA]. (2010). *The differentiation debate: Submission to the Higher Education Quality Council of Ontario*. Retrieved from http://www.ousa.ca/dev/wp-content/uploads/2011/01/ Differentiation-for-print.pdf

Ontario Undergraduate Student Alliance [OUSA]. (2006). *Teaching quality: A balance of scholarly priorities and the need for innovation*. Retrieved from http://www.ousa.ca/uploaded_files/ pdf_files/Policy%20Papers%20and%20 Statements/TeachingQualityfinal.pdf

Ouimet, J.A. & Smallwood R.A. (2005). CLASSE: The class-level survey of student engagement. *Assessment Update*, 17(6), 13-15.

Pallett, W. (2006). Uses and abuses of student ratings. In P. Seldin (Ed.), *Evaluating faculty performance: A practical guide to assessing teaching, research, and service* (pp. 50-65). San Francisco, CA: Jossey-Bass.

Pallof, R.M. & Pratt, K. (2008). *Effective course, faculty and program evaluation*. Proceedings from the 24th Annual Conference on Distance Teaching and Learning, 1-5.

Patrick, C.L. (2011). Student evaluations of teaching: Effects of the big five personality traits, grades and the validity hypothesis. *Assessment & Evaluation in Higher Education, 36*(2), 239-249.

Penny, A.R. & Coe, R. (2004). Effectiveness of consultation on student ratings feedback: A meta-analysis. *Review of Educational Research, 74*(2), 215–253.

Performance Assessment for California Teachers. (n.d.). Retrieved from http://www.pacttpa.org/ _main/hub.php?pageName=Home

Piccinin, S.J. (2003). *Green Guide 4: Feedback key to learning*. Society for Teaching and Learning in Higher Education. London, ON.

Pratt, D.D. (1997). Reconceptualizing the evaluation of teaching in higher education. *Higher Education, 34*(1), 23-44.

Pratt, D.D. & Associates. (1998). *Five perspectives on teaching in adult and higher education*. Malabar, FL: Krieger Publishing.

Prebble, T., Hargraves, H., Leach, L., Naidoo, K., Suddaby, G., & Zepke, N. (2004). *Impact of student support services and academic development programmes on student outcomes in undergraduate tertiary study: A synthesis of the research*. Report to the Ministry of Education, Massey University College of Education.

Quality Assurance Agency for Higher Education [QAA]. (2012). *UK quality code for higher education*. Retrived from http://www.qaa.ac.uk/Publications/InformationAndGuidance/Documents/B3.pdf

Ramsden, P. (2003). *Learning to teach in higher education*. New York: RoutledgeFalmer.

Ramsden, P., Margetson, D., Martin, E., & Clark, S. (1995) *Recognising and rewarding good teaching in australian higher education*. A project commissioned by the Committee for the Advancement of University Teaching, Canberra: Government Printer.

Ravelli, B. (2000, June). Anonymous online teaching assessments: Preliminary findings. Paper presented at the Annual National Conference of the American Association for Higher Education, Charlotte, North Carolina.

Roche, L.A. & Marsh, H.W. (2002). Teaching self-concept in higher education. In N. Hativa & P. Goodyear (Eds.), *Teacher thinking, beliefs and knowledge in higher education* (pp. 179-218). Netherlands: Springer.

Sachs, J. (2012). *Teaching standards framework online development and pilot study project*. Macquarie University, Sydney, Australia.

Sachs, J. (2011). *Teaching standards framework online development and pilot study proposal*. Macquarie University, Sydney, Australia.

Shore, B., Foster, S., Knapper, C., Nadeau, G., Neill, N., & Sim, V. (1991). *The CAUT guide to the teaching dossier: Its preparation and us*e. Ottawa: Canadian Association of University Teachers.

Seldin, P. (2006). Essential operating principles and key guidelines. In P. Seldin (Ed.), *Evaluating faculty performance: A practical guide to assessing teaching, research, and service* (pp. 20-34). San Francisco, CA: Jossey-Bass.

Seldin, P. (1991). The teaching portfolio. Boston: Anker.

Senge, P.M. (1990). *The fifth discipline: The art and practice of the learning organization*. New York: Doubleday/Currency.

Silver, H. (2003). Does a university have a culture? *Studies in Higher Education, 28*(2), 157-169.

Slade, S. & Prinsloo, P. (2013). Learning analytics: ethical issues and dilemmas. *American Behavioral Scientist, 57*(10), 1509–1528.

Slocombe, T., Miller, D., & Hite, N. (2011). A survey of student perspectives toward faculty evaluations. *American Journal of Business Education, 4*(7), 51-57.

Smith, C., Herbert, D., Robinson, W., & Watt, K. (2001). Quality assurance through a continuous curriculum review (CCR) strategy: Reflections on a pilot project. *Assessment and Evaluation in Higher Education, 26*(5), 489-502.

Snavely, L. & Dewald, N. (2011). Developing and implementing peer review of academic librarians' teaching: An overview and case report. *The Journal of Academic Librarianship*, 37(4), 343–351.

Stakeholder Consultation Meeting. (2013, December). Toronto, ON.

Steinert, Y., Mann, K., Centeno, A., Dolmans, D., Spencer, J., Gelula, M., & Prideaux, D. (2006). A systematic review of faculty development initiatives designed to improve teaching effectiveness in medical education: BEME Guide No. 8. *Medical Teacher, 28*(6), 497-526.

Sterman, J.D. (2006). Learning from evidence in a complex world. *American Journal of Public Health, 96*(3), 505-514. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1470513/

Struyven, K., Dochy, F., & Janssens, S. (2010). Explaining students' appraisal of lectures and student-activating teaching: Perceived context and student characteristics. *Interactive Learning Environments*, *20*(5), 391–422.

Sullivan, S.L., Polnick, B., Nickson, L., Maninger, R., & Butler, J.Y. (2013). Student evaluation of teaching: The inequity of faculty scores in online versus face-to-face courses. *School Leadership*, *8*(2), 52-61.

Taylor, P. & Maor, D. (2000, February). *Assessing the efficacy of online teaching with the constructivist online learning environment survey.* Presentation at the ninth Annual Teaching Learning Forum, Perth, WA. Retrieved from http://researchrepository.murdoch.edu.au/8750/

Theall, M. (2002). *Student ratings: Myths vs. research evidence*. Retrieved from https://studentratings. byu.edu/info/faculty/myths.asp

Theall, M. & Franklin, J. (2001). Looking for bias in all the wrong places: A search for truth or a witch hunt in student ratings of instruction? In P. Theall, L. Abrami, & L. Mets (Eds.) *The student ratings debate: Are they valid? How can we best use them? New Directions in Educational Research, 109.* San Francisco: Jossey-Bass.

Theall, M. & Franklin, J. (2000). Creating responsive student ratings systems to improve evaluation practice. In K.E. Ryan (Ed.), *Evaluating teaching in higher education: A vision for the future: New Directions for Teaching and Learning, 83*. San Francisco: Jossey-Bass.

Thorpe, S.W. (2002). *Online student evaluation of instruction: An investigation of non-response bias*. Paper presented at the 42nd Annual Forum for the Association for Institutional Research, Toronto, ON.

Tobin, T.J. (2004). Best practices for administrative evaluation of online faculty. *Online Journal of Distance Learning Administration, 7*(2). Retrieved from http://www.westga.edu/ %7Edistance/ojdla/summer72/summer72.html

Trujillo, J., DiVall, M., Barr, J., Gonyeau, M., Van Amburgh, J. A., Matthews, J., & Qualters, D. (2008). Development of a peer teaching-assessment program and a peer observation and evaluation tool. *American Journal of Pharmaceutical Education,72*(6), 1-9.

Universities UK, (2008). *Quality and standards in UK universities: A guide to how the system works.* Retreived from http://www.eua.be/fileadmin/user_upload/files/Newsletter_new/44806UniUK_ guide_lores_1_.pdf

Üstünlüoğlu, E. & Güngör-Culha, D. (2012). Investigating student evaluation of teachers by using latent class analysis: A case study at a tertiary level. *International Journal of Education, 4(3),*147-159.

Vajoczki, S. (2008). *Effective teaching and teaching evaluation practices: Canadian university geography departments*. (Unpublished doctoral dissertation). Wilfrid Laurier University, Waterloo, ON.

Weimer, M. (2002). *Learner-centred teaching: Five key changes to practice*. San Francisco: Jossey-Bass.

Wellein, M., Ragucci K.R., & Lapointe, M. (2009). A peer review process for classroom teaching. *American Journal of Pharmaceutical Education*, *73*(5), 1-7.

Weschke, B. & Canipe, S. (2010) The faculty evaluation process: The first step in fostering professional development in an online university. *Journal of College Teaching & Learning, 7*(1), 45-58.

Whitehead, L. (2013). An influence pathway: Leadership to culture to practice to outcomes. In K. Norrie, & M.C. Lennon, *Measuring the value of a postsecondary education* (pp. 159-168). Kingston, ON: McGill-Queen's University Press.

Winchester, T.M. & Winchester, M. (2011). Exploring the impact of faculty reflection on weekly student evaluations of teaching. *International Journal for Academic Development, 16*(2), 119-131.

Winer, L. (2013, September). *Getting from there to here: The example of online course evaluations*. Presentation to SSMU Policy and Advocacy Resource Committee, McGill University, Montreal, QC.

Wininger, S. & Birkholz, P. (2013). Sources of instructional feedback, job satisfaction, and basic psychological needs. *Innovative Higher Education*, *38*(2), 159-170.

Wolf, R. (1987). The nature of educational evaluation. *International Journal of Educational Research, 11*(1), 7-19.

Wright, W.A., Mighty, J., Dyens, O., & Rogers, P. (2013, February). *Building capacity? Educational development pioneers in senior administration*. Panel Discussion presented at the Educational Developers' Caucus Annual Conference, Wilfrid Laurier University, ON.

Wright, W.A. & O'Neill, M.C. (1995). Teaching improvement practices: International perspectives. In W.A. Wright (Ed.), *Successful faculty development strategies*. Boston, MA: Anker Publishing Co.

Zakrajsek, T. (2006). Using evaluation data to improve teaching effectiveness. In P. Seldin (Ed.), *Evaluating faculty performance. A practical guide to assessing teaching, research and service* (pp. 166-179). Bolton, MA: Anker.

Appendix A

Team Mandates

Aggregate Data Team

Effective Practices Team

Environmental Scan Team

Policies and Procedures Team

Technical Team

Project Mandate: Aggregate Data Team

The Aggregate Data Tool Team is researching and establishing a recommended approach to aggregate SRI data modeling for the proposed Teaching Evaluation (TE) Toolkit Suite.

Purpose in Overall Project Section:

- What statistically sound approaches to aggregating SRI data-based should inform the proposed TE Suite?
- What is the potential, and potential challenges, of various uses of aggregate SRI data?
- Propose preliminary design for visual representation of aggregate SET data that includes options for data manipulation meeting the needs of instructors and administrators related to documenting and improving teaching practice at the individual, departmental, and institutional levels.
- What is the potential of such a tool for improving strategic planning?

Analysis:

- How can data be most effectively represented for use by instructors and administrators?
- What degree of data manipulation, and what kinds, would be desirable?
- What limits to access should be put in place, if any (i.e., who should have access to what?)?
- What are suggested design recommendations for an aggregate TE data tool?

Deliverable: A report on aggregate data addressing the questions above.

Project Mandate: Effective Practices Team

The Effective Practices Team will research and report on instances of exemplary practice in teaching evaluation, internationally.

Purpose in Overall Project:

- To identify core and desirable practices and elements that should be integrated into proposed Toolkit design and other planned contextualizing documents
- To identify leading practitioners for further consultation and involvement in later stages of project
- To establish persuasive evidence of impactful practices and identify approaches associated with success (and approaches that met with less success) for use in the feasibility study
- To identify the implications of this research for the Ontario context (note: will need to coordinate with Environmental scan group for this)
- To inform the work of the coordinating committee and the technical team in developing the proposed design

Project Elements:

- Annotated Bibliography
- Literature review
- Data analysis and synthesis
- Reporting

Information Gathering:

- What does research indicate are "effective practices" in teaching evaluation with an emphasis on facilitating teaching improvement?
- Based on discussion with leaders of TE initiatives, what are the conditions that favour successful implementation and faculty engagement, and the factors that make those "effective practices" operable?
- Effective practices:
    - in evaluation (design, implementation, analysis, reporting)
      in change/roll out practices related to teaching evaluation
    - in monitoring/assessing quality of teaching evaluation practice

Analysis:

- Which elements of these effective practices must (should?) be adopted in the Ontario context? Which elements would be desirable but not critical?
- What conditions and mechanisms are most commonly associated with greatest adoption and success in implementing teaching evaluation frameworks, both at the institutional and broader levels?
- What pitfalls must be avoided?
  What are the implications of these findings for the Ontario context?

Deliverable: **A report highlighting examples of national and international effective practices** in this area, including an assessment of what might be feasible and adaptable to the Ontario context.

Project Mandate: Environmental Scan Team

The Environmental Scan Team will research current practices and policy related to the evaluation of teaching in Ontario Universities (coordinating with the policies and procedures researcher).

Purpose in Overall Project:

- to establish a baseline description of what is currently happening with a view to understanding the challenges and barriers that a provincial framework would have to address
- to identify commonalities among universities to inform the framework and tool suite modeling
- to identify resources practices and procedures potentially worth adopting for a provincial framework

Project Elements:

- collection of materials being used to evaluate teaching at Ontario universities, as well as guides, contextualizing documents
- an online survey of TE practices at Ontario universities
- discussions with key individuals identified through survey (CTL Directors, Institutional Analysis Directors, in their official roles related to this work at their universities) at Ontario universities

Information Gathering:

- What is the state of current TE practice in Ontario Universities?
- How do they do it?
- What information do they gather?
- What is involved other than SRI? What is available, and what is recommended, and what is required?
- With regard to SRI, what is asked? What methods of validation/assessment of reliability went into the process, if any?
- What rules govern use of TE data at various institutions? *(Policies and Procedures to assist)*
- What is TE data generally used for, officially?
- Is there any important history around changing TE practices that it is important to know?

Analysis:

- What is the perceived role and significance of TE at Ontario universities?
- What are the strengths and weaknesses of current TE practices in Ontario Universities?
- What do individuals working with TE and teaching improvement at Ontario Universities identify as their needs in moving TE and TI forward?
- Comparison of SRI forms including identification of common and less common questions, inclusion of qualitative data, use of online format, use of third-party SRI.
- What are potential barriers to adoption of a possible TE Suite at Ontario institutions?
- What “risk factors” connected to the evaluation of teaching must be addressed given the political and economic climate?
- Can we get enough of a sense of current practice to estimate a provincial cost of TE at Ontario institutions?

Deliverable: A report on Environmental Scan addressing questions identified above.

Project Mandate: Policies and Procedures Team

The Policies and Procedures Team will research and report on common and outlier governance and labour agreement regulations related to teaching evaluation in the province of Ontario, using a representative sample of Ontario universities as a guide.

Purpose in Overall Project:

- To inform the conceptual and technical design of the Teaching Evaluation (TE) Toolkit and teaching evaluation framework, so that the eventual product will be adaptable to a broad range of institutional contexts
- To inform the environmental scan of teaching evaluation practices in Ontario, in particular with regard to explicit regulation of design, implementation, and use of teaching evaluation tools in the province, and any potential barriers to the adoption of a common but adaptable province-wide framework
- To demonstrate to the client and stakeholders that we have taken into account current regulatory frameworks and to raise awareness of how these kinds of challenges impede collaborative inter-institutional practice

Project Elements:

- Review of institutional documents
- Review of collective agreements
- Consultation with institutional representatives as necessary
- Data analysis and synthesis
- Reporting

Information Gathering:

- What policies, bylaws, laws, and labour agreements govern the design, implementation, and use of TE tools at Ontario universities? Are they generally approved by Senate or part of collective agreement? Are there review and renewal requirements?
- How are TE data intended to be used, according to policies, bylaws, laws, and labour agreements at Ontario institutions?

Analysis:

- What are potential legal, regulatory, and contractual barriers to adoption of a possible TE Suite at Ontario institutions?
- What "risk factors" connected to the evaluation of teaching need to be addressed given the political and economic climate?
- What pitfalls must be avoided?
- To what degree are the pitfalls/barriers changeable?
- To what degree does there appear to be variation in Ontario university practice based on legislation, which different institutions are interpreting differently?

Deliverable: This research **informs both the environmental scan and the conceptual modelling** of the tool. The deliverable for this team is a report with accompanying documentation.

Project Mandate: Technical Team

Based on the findings of the other teams and the coordinating committee's ongoing conceptual modeling, the Technical Team will prepare a preliminary technical model, a development methodology, and initial user interface designs for the Teaching Evaluation Toolkit. These deliverables will inform development and testing across member institutions in the second phase of the project.[1]

Purpose in Overall Project:

- to translate the conceptual design into preliminary technical designs, and identify and suggest preliminary solutions to identified technical design problems
- to discover and document high-level design requirements for the proposed Teaching Evaluation Toolkit design, in support of the Phase Two proposal
- to inform the feasibility study's selection of a development methodology suited to the adapting, research-driven design of the Teaching Evaluation Toolkit
- to identify potential technical barriers to the province-wide delivery of the Toolkit
- to prepare a preliminary visual design for the Toolkit interface
- to collaborate with the Aggregate Data team on the design and technical aspects of the SRI visualization and analysis tool

Information Gathering:

- What are the potential benefits and limitations of third party SRI management systems, as compared with a custom-developed system? (in collaboration with a "third-party SRI" researcher)
- How have other online Teaching Evaluation tools been developed? (in collaboration with the Effective Practices and Environmental Scan teams)

Analysis:

- What are the core technical challenges of the initially proposed suite of tools?
- As the requirements of the projects continue to evolve, what are the technical and project-management challenges of accommodating those revisions?
- Which aspects, if any, of the proposed Teaching Evaluation Toolkit system are technically impractical, risky, or prohibitively expensive to implement? What simpler alternatives could be proposed?
- What security and privacy issues need to be addressed?
- What should the development plan be for the proposed suite of tools?

[1] Of all the team mandates this one was the most fluid. In a traditional design process, the assessment of the client's needs would be completed before modeling and preliminary design began. This was impossible owing to the very significant time constraints faced by the project team; therefore, as the project evolved, so did the mandate. Given the "agile design" methodology identified and advocated by both the overall project team and the technical team, the provision of a "locked-in" plan was not a recommended approach. The technical team has worked with the project team through a number of iterations on the overall design for the Teaching Evaluation Toolkit, advising on the feasibility and challenges of various approaches, as the project team narrowed its focus to identify areas where intervention is likely to produce the most "traction" for improvement. While time-consuming, this iterative approach has proven invaluable to the establishment of an efficient, technically feasible, and culturally appropriate solution. As a result, the plan for a fully integrated Teaching Evaluation Toolkit has been broken into a number of discrete tools. For this reason the team has not provided a comprehensive, unifying design for a Toolkit interface; but they have provided preliminary visualizations of the individual tools. The detailed designs of the tools will evolve through consultation with multi-institutional user and design teams in Phase Two.

Appendix B

Core Practices Involved in the Evaluation of Teaching

Core Practices Involved in the Evaluation of Teaching

Student Ratings of Instruction

Student Ratings of Instruction (SRI) are also called Student Evaluations of Teaching (SET) in the literature. Some authors prefer the term ratings of instruction because they feel that students are not in a position to offer a complete evaluation of the many facets of instruction (Berk, 2013; Pallett, 2006; Hativa, 2013a), and because student surveys do not form an entire evaluation but are instead one piece of information that is considered during the evaluation process. We agree with this position and will use the term SRI throughout this review.

SRI is the most popular form of faculty appraisal in North America. Although some authors claim that SRI data lacks validity (Langbein, 2008), the majority of studies find that students are capable of making valid ratings of instruction (Beran & Rokosh, 2009; Hativa, 2013; Marsh, 2007; Theall, 2002). Berk (2009) states that 97% of department chairs in North America are using SRI to inform summative decisions. In Europe and Asia, SRI is less common but on the increase (Hallinger, 2010; Marsh, 2007; Surgenor, 2013). While the vast body of research surrounding SRI provides some contradictory findings, the five following practices have been found to be effective.

1. Students should only be asked to rate faculty on criteria where they can make appropriate judgments (Buller, 2012; Pallett, 2006; Theall, 2002). These criteria include questions about demeanour in the classroom, student experience, and student perceptions. Excluded questions, or those which students are not adequately equipped to make sound judgments, relate to faculty knowledge, curriculum design, assessment design, grading standards, and research ability (Hativa, 2013a)
2. Validate the instrument (Arreola, 2007; Marsh, 2007). Homegrown or instructor developed SRI surveys have unknown psychometric properties and questionable validity (Berk, 2013). Carefully constructed surveys with known psychometric properties that have been verified and refined through testing and research are recommended.
3. Design implementation in ways that reflect serious commitment to using student input. For example, implementing evaluation earlier in the semester allows for faculty to act on the feedback to benefit the students who gave the feedback, and signals to the student body that their ratings are valued and acted upon (Benton & Cashin, 2012; Knol, Veld, Vorst, Driel & Mellenbergh, 2013; Winchester & Winchester, 2011). Perceptions of the credibility and value placed on SRI are critical to engagement with the process and the subsequent data.
4. Interpret results with care. A variety of factors can impact student ratings of instruction: class size, instructor likeability, course difficulty, course level, faculty, delivery method, and age of instructor (Clayson, 2009; Galbraith, Merrill & Kline, 2012; Hativa, 2013b; Langbein, 2008; Patrick, 2011; Slocombe, Miller, & Hite, 2011; Sullivan, Polnick, Nickson, Maninger, & Butler, 2013). Any interpretation of SRI results must be performed thoughtfully and carefully (Buller 2012; Hativa 2013b; Struyven, Dochy & Janssens 2010; Theall 2002).
5. Have faculty discuss and reflect upon the ratings (Penny & Coe, 2004; Winchester & Winchester, 2011). According to Penny and Coe (2004), "Consultation on student ratings, or consultative feedback, is widely recognized as an effective support mechanism to help teachers use student ratings feedback to improve their teaching" (p. 215). In earlier research by Marsh and Roche (1993), they state that, "the most robust finding from SET feedback research is that consultation augments the effects of written summaries of SETs" (p. 223).

Evidence linking SRI results to student learning is weak. While "medium-size relationships between instructor ratings and student learning" exist, there are significant problems in measuring student

learning and its causal relationship with instruction (Berk, 2014; Hativa, 2013a). Clayson's (2009) meta-analysis of the research on SRI and student learning found that of the 42 studies included, 10 had negative correlations between SRI ratings and student learning and 32 had positive correlations. The net result was a statistically insignificant positive correlation. It is difficult to know how to interpret these findings given the significant variations and variable quality of SRI instruments, and the challenge of accurate measurement of student learning (Berk, 2014).

There is little evidence that an individual's SRI ratings change over time or are used effectively by faculty to improve their teaching (Beran & Rokosh, 2009; Beran, Violato, Kline, & Frideres, 2005; Lang & Kersting, 2007; Wininger & Birkholz, 2013). Success stories exist (Hallinger, 2010) and effective implementation of SRI is critical. According to Wininger and Birkholz (2013), faculty find that student feedback they solicit themselves (as opposed to mandated feedback) is the most useful information they have to improve their teaching. Knol et al. (2013) and Marsh and Roche (1993) provide evidence that prompt access to data in a mediated context during the course can create significant differences in teaching improvement.

Peer Evaluation of Instruction

Although not as common as student ratings, peer evaluation is gaining popularity. There are wide variations in practice and implementation yet these have yielded common themes.

1. Numerous studies indicate that peer review can provide more accurate input to teaching evaluation in some areas where students are less likely have the expertise to provide accurate feedback, such as course content, goals, and design; pedagogical content knowledge, evaluation practices, or ethical standards of practice (DeZure, 1990; Keig & Waggoner, 1994; Cohen & McKeachie, 1980).

2. Faculty need to be involved in all stages of the peer evaluation process (Snavely & Dewald, 2011; Trujillo et al., 2008; Wellein, Ragucci & Lapointe 2009). This includes establishing the purpose, design, implementation, and evaluation of the process. Because faculty are both evaluators and subjects, involvement in all aspects is essential.

3. To be effective, peer evaluation must take place in an academic culture that supports it (Blauvelt, Davenport & Spath 2012; Chism 2007; Snavely & Dewald 2011). Faculty must feel that the process is valued by administrators and that the peer evaluations are either followed up or have an impact on decisions (Iqbal 2013).

4. Some faculty feel that the peer evaluation exercise is a pointless "ticking the box exercise" or that their peers are not competent evaluators of teaching (Chamberlain, D'Artrey & Rowe 2011; Iqbal 2013). Limited engagement can be the result of faculty perceptions that only friends are chosen as evaluators, that the results are not as valued as SRI, or that single observations do not provide a valid picture of teaching quality. Chamberlain et al. (2011) noted that in about half of the cases studied the results of peer observation of teaching were discussed with faculty at their subsequent annual performance review.

5. Peer observations need to be followed by time for discussion and reflection (Boerboom et al., 2011; Hansen et al., 2007; Snavely & Dewald, 2007; Wellein et al., 2009). Ideally there should be a pre-observation meeting and a post-observation meeting. The pre-observation meeting helps to set a context for the observation and the post-observation meeting can be used as a debrief and also as an exchange of ideas. Formalized reflection processes seem to increase the quality of critical reflection (Boerboom et al. 2011).

6. As with all forms of evaluation, structures and training impact the degree of reliability of the activity (Centra, 1993; Goodwin & Stevens, 1996), but inter-rater reliability of peer review can be improved through effective process and training (Millis & Kaplan, 1997; DeZure, 1999). Arreola (2007) claims that it is preferable to have faculty peer evaluate the teaching materials, syllabi and other course documentation but not to directly observe or evaluate teaching.

Peer observation of teaching clearly has the potential to be an effective means of sharing ideas, improving collegiality and enacting meaningful improvements but these benefits only seem to occur if faculty are involved in the entire process and the process is demonstrably valued by administrators.

Teaching Dossiers

According to the Canadian Association of University Teachers (CAUT), a "teaching dossier is a summary of an academic's major teaching accomplishments and strengths. It is to an academic's teaching what lists of publications, grants, and academic honours are to research," (p. 2, 2007). They provide a formal structure for the summary of major teaching accomplishments and the documentation of practice, bringing together data from multiple sources with annotative text written by the instructor. Teaching dossiers include a range of quantitative and qualitative data, often including a record of teaching responsibilities, student ratings of instruction data, written feedback and comments from students, supervision responsibilities, a teaching philosophy, descriptions of pedagogical approaches employed across a broad range of contexts, peer observation reports, records of innovative practices and their impact, evidence of involvement with curriculum renewal or design, and student work samples (Wright and O'Neill, 1995).

As Seldin (1991) points out, teaching dossiers serve a dual purpose, firstly allowing for the collection and representation of hard evidence of teaching effectiveness for decision making and evaluative purposes, but secondly, providing an effective framework to facilitate reflection about areas of teaching that need improvement. Another benefit of using teaching dossiers is their ability to take into account variability in instructors' goals, beliefs about teaching, and practices, and disciplinary differences in teaching: the teaching dossier very explicitly demands that instructors demonstrate the connections between their approach to teaching, their practices, and student outcomes (Pratt, 2005).

Teaching portfolios or teaching dossiers are a well-accepted method for the evaluation and assessment of teaching (Burnham, Hooper, & Wright, 2010; Devanas, 2006; McColgan & Blackwood, 2009, O'Farrell, 2007, Schonwetter, Friesen, & Taylor, 2002).

1. Dossiers are important individual representations of the metacognitive, reflective, and cyclical nature of teaching effectiveness and improvement (Devanas, 2006; McColgan & Blackwood, 2009).

2. Dossiers provide a qualitative opportunity for linking and exhibiting research and teaching practice, and faculty perceptions indicate a strong belief in the existence of, and need for, a symbiotic relationship (Burnham, Hooper & Wright, 2010; Robertson & Bond 2001).

3. Teaching philosophies, which are often included in dossiers as reflections of disciplinary cultures, institutional structures, stakeholder expectations, and personal beliefs about teaching, can be evaluated using a conceptual model and rubric (Schonwetter, Friesen, & Taylor, 2002).

4. Faculty buy-in is enhanced when teaching effectiveness practices are directly relatable to parallel professional concerns (Burnham, Hooper & Wright, 2010; Way, 2002). The teaching dossier is directly relatable to the tripartite requirements of the promotion and tenure (P & T) process: teaching, research, and service. Institutional emphasis on each requirement varies; therefore, a flexible model that is comprehensive yet adaptable would have the most utility.

Self-evaluations

Numerous universities mount self-evaluation tools on their websites as part of their teaching and learning support offerings. The practice of self-evaluation is suggested as a key piece in raising instructor awareness of individual teaching styles, enhancing reflective capacity in instructors, and in assisting them to set goals for improved practice. Some institutions use self-evaluation as part of teaching assessment and personnel decisions. Acceptance of the value of self-evaluation as a component both of improvement-oriented evaluation and of evaluation for personnel purposes has grown over the last two decades (Centra, 2000).

1. Whether the impetus is the improvement of practice or personnel decisions, engaging in the process of self-evaluation can influence faculty performance. Teacher self-ratings are a powerful mediator of teacher performance outcomes, motivation to teach, satisfaction with teaching, and likelihood they will engage in teaching development activities (Amundsen et al., 2005; Marsh & Roche, 1993; Roche & Marsh, 2002).
2. Instruments include TEEQ and ATI (Prosser & Trigwell, 2006) Teaching Perspectives Inventory (Collins & Pratt, 2011), CLASSE (Smallwood & Ouimet, 2009), and numerous homegrown offerings.
3. Self-evaluations should precede student ratings (Zakrajsek, 2006) and be used to contextualize growth and improved practice in teaching dossiers (Devanas, 2006).
4. Self-evaluations should be constructive, designed to improve teaching, and relevant to more formalized evaluation efforts (Fang, 2007; Weschke & Canipe, 2010).
5. Faculty self-evaluations of teaching and student evaluations of the same instruction produce similar results (Barnett, Matthews & Jackson, 2001).

Curriculum evaluations

A component of effective teaching and learning is the design and implementation of curriculum, course content, and student learning contexts (Hénard & Roseveare, 2012). Evaluating and improving the curriculum is part of quality teaching. Effective practices include:

1. Consider the range of student experiences, which are part of an educational course (Palloff & Pratt, 2008; Hénard & Roseveare, 2012).
2. Use cyclical feedback mechanisms that measure established competencies against achievement outcomes (Dunn, Morgan, O'Reilly & Parry, 2004; Palloff & Pratt, 2008; Smith, Herbert, Robinson, & Watt, 2001).
3. Differentiate the unique aspects of online environments (Sullivan, Polnick, Nickson, Maninger, & Butler, 2013; Tobin, 2004; Weschke & Canipe, 2010).
4. Use collaborative and responsive tools such as the Constructivist On-Line Learning Environment Survey, or COLLES (Taylor & Maor, 2000).

Online TE Tools

Several universities in Canada and internationally have implemented online SRIs accessed through individual university web portals. Although perceptions and evidence about online vs. paper response rates are mixed (Ravelli, 2000; Thorpe, 2002), Dalhousie University cites references

indicating that students prefer online evaluations (Kuhtman, 2004 in Donovan et al., 2006) and provide more thoughtful comments when using user-friendly online tools (Ravelli, 2000).

1. Online surveys should be anonymous and confidential (Benton & Cashin, 2012; Berk, 2013).
2. Frequent email reminders, posters (Tucker, 2008), early grade release (e.g. Stanford, UTDallas) prizes (e.g., UTDallas) and opportunities for in-class completion (Kiceniuk, 2012) are some of the incentives for completing course evaluations. In some instances students may only complete teacher evaluations once they have submitted other types of course or unit feedback (Tucker, 2008).
3. There is an indication from the literature (Gravestock & Gregor-Greenleaf, 2008; Kiceniuk, 2012; Tucker, 2008) and numerous university websites that information clarifying the purpose of evaluations may support student engagement in the process.
4. There is some evidence that mean scores from online teaching evaluation are lower than pencil-and-paper scores, and comparisons between these two versions of evaluation may be questionable. It is not clear that this finding is generalizable at this point, and so transition to online ratings of instruction must incorporate a stage of statistical analysis of outcomes at specific institutions.
5. Some studies suggest that students write more, and more substantively, when asked for written feedback in the online format.
6. Cost savings from online teaching evaluation have been shown to be as much as 50% (Winer, 2013).
7. Effective transitions have generally involved sustained and well-managed investment of resources and high degrees of stakeholder engagement (Joughin & Winer, 2014).

Appendix H provides further information regarding the use of third-party SRI provision.

References

Amundsen, C., Abrami, P., McAlpine, L., Weston, C., Krbavac, M., Mundy, A., & Wilson, M. (2005, April). *The what and why of faculty development in higher education: A synthesis of the literature*. Paper presented at the American Educational Research Association (AERA), Montreal, QC.

Arreola, R. (2007). *Developing a comprehensive faculty evaluation system: A guide to designing, building and operating large-scale faculty evaluation systems*. San Francisco: Anker Publishing.

Barnett, C. W., Matthews, H. W., & Jackson, R. A. (2003). A comparison between student ratings and faculty self-ratings of instructional effectiveness. *American Journal of Pharmaceutical Education, 67*(4), 1-6.

Benton, S. L., Cashin, W. E., & Kansas, E. (2012). *IDEA PAPER# 50: Student ratings of teaching: A summary of research and literature*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.388.8561

Beran, T. & Rokosh, J. (2009). Instructors' perspectives on the utility of student ratings of instruction. *Instructional Science, 37*(2), 171-184.

Beran, T., Violato, C., Kline, D., & Frideres, J. (2005). The utility of student ratings of instruction for students, faculty, and administrators: A "consequential validity" study. *The Canadian Journal of Higher Education, 35*(2), 49-70.

Berk, R.A. (2014). Should student outcomes be used to evaluate teaching? *Journal of Faculty Development, 28*(2), 87-96.

Berk, R.A. (2013). *Top 10 flashpoints in student ratings and the evaluation of teaching*. Stylus Publishing, Sterling VA.

Berk, R. A. (2009). Using the 360° multisource feedback model to evaluate teaching and professionalism. *Medical Teacher, 31*(12), 1073-1080.

Blauvelt, M., Erickson, C., Davenport, N., & Spath, M. (2012). Say yes to peer review: A collaborative approach to faculty development. *Nurse Educator, 37*(3), 126–130.

Boerboom, T., Jaarsma, D., Dolmans, D., Scherpier, A., Mastenbroek, N., & Van Beukelen, P. (2011). Peer group reflection helps clinical teachers to critically reflect on their teaching. *Medical Teacher, 33*, 615-623.

Buller, J. (2012). *Best practices in faculty evaluation: A practical guide for academic leaders*. Jossey-Bass: San Francisco.

Burnham, V., Hooper, L., & Wright, V. (2010). *Tools for dossier success*. New York: Routledge.

Canadian Association of University Teachers [CAUT]. (2007). *CAUT teaching dossier*. Retrieved from

http://www.caut.ca/docs/default-source/professional-advice/teaching-dossier.pdf

Centra, J.A. (2000). Evaluating the teaching portfolio: a role for colleagues. *New Directions for Teaching and Learning, 83*, 87-93.

Centra, J.A. (1993). *Reflective faculty evaluation*. San Francisco: Jossey-Bass.

Chamberlain J., D'Artrey M. & Rowe, D. (2011). Peer observation of teaching: A decoupled process. *Active Learning in Higher Education, 12*(3) 189–201.

Chism, N. V. N. (2007). *Peer review of teaching: A sourcebook* (2nd ed.). Bolton, MA: Anker.

Clayson, D. (2009). Student evaluations of teaching: Are they related to what students learn? A meta-analysis and review of the literature. *Journal of Marketing Education, 31*(1), 16-30.

Cohen, P.A. & McKeachie, W.J. (1980). The role of colleagues in the evaluation of teaching. *Improving College and University Teaching, 28*(4), 147-154.

Collins, J. B., & Pratt, D. D. (2011). The teaching perspectives inventory at 10 years and 100,000 respondents: Reliability and validity of a teacher self-report inventory. *Adult Education Quarterly, 61*(4), 358-375.

Devanas, M. A. (2006). Teaching portfolios. In P. Seldin (Ed.), *Evaluating faculty performance. A practical guide to assessing teaching, research and service* (pp. 111-130). Bolton, MA: Anker Publishing.

DeZure, B. (1999). Evaluating teaching through peer observation. In P. Seldin (Ed.), *Changing practices in evaluating teaching: A practical guide to improved faculty performance and promotion/tenure decisions* (pp. 71-96) Bolton, MA: Anker Publishing.

Donovan, J., Mader, C. E., & Shinsky, J. (2006). Constructive student feedback: Online vs. traditional course evaluations. *Journal of Interactive Online Learning, 5*(3), 283-296.

Dunn, L., Morgan, C., O'Reilly, M., & Parry, S. (2004). *The student assessment handbook: New directions in traditional and online assessment*. London: Routledge Falmer.

Fang, Berlin (2007, May-June). A performance-based developmental model for online faculty. *Performance Improvement, 46*, 17-24.

Goodwin, L.D. & Stevens, E.A. (1996). The influence of gender on university faculty members' perceptions of 'good' teaching. In D.E. Finnegan, D. Webster, & F. Gamson (Eds.), *Faculty and faculty issues in colleges and universities* (pp. 27-286). Needham Heights, MA: Simon & Schuster.

Gravestock, P. & Gregor-Greenleaf, E. (2008). *Student course evaluations: Research, models and trends*. Toronto: Higher Education Quality Council of Ontario. Retrieved from http://www.heqco.ca/SiteCollectionDocuments/Student%20Course%20Evaluations_Research,%20Models%20and%20Trends.pdf

Hallinger, P. (2010). Using faculty evaluation to improve teaching quality: A longitudinal case study

of higher education in Southeast Asia. *Educational Assessment Evaluation and Accountability, 22*, 253–274.

Hansen L., McCollum M., Paulsen S., Cyr T., Jarvis C., Tate G., & Altiere, R. (2007). Evaluation of an evidence-based peer teaching assessment program. *American Journal of Pharmaceutical Education, 71* (3), 1-45.

Hativa, N. (2013a). *Student Ratings of Instruction: A practical approach to designing, operating, and reporting*. Charleston, SC: CreateSpace Independent Publishing Platform.

Hativa, N. (2013b). *Student ratings of instruction: Recognizing effective teaching*. USA: Oron Publications.

Hénard, F. & Roseveare, D. (2012, Sept). *Fostering quality teaching in higher education: Policies and practices. An IMHE Guide for Higher Education Institutions*. (Institutional Management in Higher Education) http://www.oecd.org/edu/imhe/QT%20policies %20and%20practices.pdf

Iqbal, I. (2013). Academics' resistance to summative peer review of teaching: questionable rewards and the importance of student evaluations. *Teaching in Higher Education, 18*(5), 557-569.

Joughin, G., & Winer, L. (2014). *Response to Teaching Evaluation Toolkit preliminary reporting*. Panel discussion presented March 5, 2014 in Toronto, Ontario.

Keig, Larry, & Waggoner, Michael D. (1994). *Collaborative peer review: The role of faculty in improving college teaching*. Washington, D.C. ASHE-ERIC Higher Education Report No. 2.

Kiceniuk, D. (2012). *Student ratings of instruction (SRI) in higher education-exploring the issues. FOCUS on university teaching and learning*. Centre for Learning and Teaching, Dalhousie University.

Knol, M., in't Veld, R., Vorst, H., Driel, J., & Mellenbergh, G. (2013). Experimental effects of student evaluations coupled with collaborative consultation on college professors' instructional skills. *Research in Higher Education, 54*(8), 825-850.

Lang, J. W. B., & Kersting, M. (2007). Regular feedback from student ratings of instruction: Do college teachers improve their ratings in the long run? *Instructional Science, 35*(3), 187-205.

Langbein, L. (2008). Management by results: Student evaluation of faculty teaching and the mis-measurement of performance. *Economics of Education Review, 27*(4), 417-428.

Marsh, H. W. (2007). Students' evaluations of university teaching: A multidimensional perspective. In R.P. Perry & J. C. Smart (Eds.), *The scholarship of teaching and learning in higher education: An evidence-based perspective* (pp. 319-384). New York: Springer.

Marsh, H. W., & Roche, L. (1993). The use of students' evaluations and an individually structured intervention to enhance university teaching effectiveness. *American Educational Research Journal, 30*(1), 217–251.

McColgan, K., & Blackwood, B. (2009). Systematic review protocol on the use of teaching portfolios

for educators in further and higher education. *Journal of Advanced Nursing, 65*(12), 2500-2507.

Millis, B., & Kaplan, B. (1997). Enhancing teaching through peer classroom observations. In P. Seldin & Associates (Eds.), *Improving College Teaching*, (pp. 137-151). Bolton, MA: Anker Publishing Company, Inc.

O'Farrell, C., (ed) 2007. *Teaching Portfolio Practice in Ireland: A Handbook.* Trinity College Dublin and All Ireland Society for Higher Education (AISHE), Dublin. Retrieved from http://www.aishe.org/readings/2007-2/

Pallett, W. (2006). Uses and abuses of student ratings. In P. Seldin (Ed.), *Evaluating faculty performance: A practical guide to assessing teaching, research, and service*, (pp. 50-65). San Francisco, CA: Jossey-Bass.

Pallof, R. M. & Pratt, K. (2008). *Effective course, faculty and program evaluation*. Proceedings of the 24th Annual Conference on Distance Teaching and Learning, 1-5.

Patrick, C. L. (2011). Student evaluations of teaching: effects of the big five personality traits, grades and the validity hypothesis. *Assessment & Evaluation in Higher Education, 36*(2),239–249

Penny, A. R., & Coe, R. (2004). Effectiveness of consultation on student ratings feedback: A meta-analysis. *Review of Educational Research, 74*(2), pp. 215–253

Pratt, D.D. (2005). Teaching. In L. M. English (Ed.) *International Encyclopedia of Adult Education* (pp. 610-615). New York: Palgrave Macmillan, 610-615.

Prosser, M., & Trigwell, K. (2006). Confirmatory factor analysis of the Approaches to Teaching Inventory. *British Journal of Educational Psychology 76*(2), 405-419.

Ravelli, B. (2000). *Anonymous Online Teaching Assessments: Preliminary Findings*. Paper presented at the Annual National Conference of the American Association for Higher Education, June 14 -18, 2000. Charlotte, North Carolina.

Robertson, J. & Bond, C. (2001) Experiences of the relation between teaching and research: what do academics value? *Higher Education Research and Development, 20*(1), pp.5-19.

Roche, L. A., & Marsh, H. W. (2002). Teaching self-concept in higher education. In N. Hativa & P. Goodyear (Eds.), *Teacher thinking, beliefs and knowledge in higher education* (pp. 179-218). Netherlands: Springer.

Schonwetter, D.J. Sokal, L., Friesen, M. and Taylor, K.L. (2002). Teaching philosophies reconsidered: A conceptual model for the development and evaluation of teaching philosophy statements. *The International Journal for Academic Development, 7*(1), 83-97.

Seldin, P. (1991). *The teaching portfolio*. Boston: Anker.

Slocombe, T., Miller, D., & Hite, N. (2011). A survey of student perspectives toward faculty evaluations. *American Journal of Business Education, 4*(7), 51-57.

Smallwood, R.A. & Ouimet, J. A. (2009) CLASSE: Measuring student engagement at the classroom level. In T. W. Banta, E. Jones, & K. Black (Eds.), *Designing effective assessment: Principles and profiles of good practice*. San Francisco, CA: Jossey-Bass

Smith, C., Herbert, D., Robinson, W., & Watt, K. (2001). Quality assurance through a continuous curriculum review (CCR) strategy: Reflections on a pilot project. *Assessment and Evaluation in Higher Education, 26*(5), 489-502.

Snavely, L. & Dewald, N. (2011). Developing and implementing peer review of academic librarians' teaching: An overview and case report. *The Journal of Academic Librarianship, 37*(4), 343–351.

Struyven, K., Dochy, F., & Janssens, S. (2010). Explaining students' appraisal of lectures and student-activating teaching: perceived context and student characteristics. *Interactive Learning Environments, 20*(5), 391–422.

Sullivan, S. L., Polnick, B., Nickson, L., Maninger, R., & Butler, J. Y. (2013). Student evaluation of teaching: The inequity of faculty scores in online versus face-to-face courses. *School Leadership, 8*(2), 52-61. Retrieved from http://www.tcpea.org/cms/lib07/TX01923126/Centricity/Domain/191/summer13.pdf.

Surgenor, P. (2013). Obstacles and opportunities: Addressing the growing pains of summative student evaluation of teaching. *Assessment & Evaluation in Higher Education, 38*(3), 363–376.

Taylor, P., & Maor, D. (2000). *Assessing the efficacy of online teaching with the constructivist online learning environment survey*. 9th Annual Teaching Learning Forum, 2-4 February 2000, Perth, W. A. Retrieved from http://researchrepository.murdoch.edu.au/8750/.

Theall, M. (2002). *Student ratings: Myths vs. research evidence*. Retrieved from https://studentratings.byu.edu/info/faculty/myths.asp.

Thorpe, S. W. (2002). *Online student evaluation of instruction: An investigation of non-response bias*. Paper presented at the 42nd Annual Forum for the Association for Institutional Research, 2002; Toronto, Ontario

Tobin, T.J. (2004). Best practices for administrative evaluation of online faculty. *Online Journal of Distance Learning Administration, 7*(2). Retrieved April 19, 2005, from http://www.westga.edu/%7Edistance/ojdla/summer72/summer72.html

Tucker, B. M. (2013). Student evaluation to improve the student learning experience: an Australian university case study. *Educational Research and Evaluation, 19*(7), 615-627.

Way, D. (2002). *Evaluating teaching portfolios: Moving from a competitive to a collaborative culture*. Paper presented at the 4th World Conference of the International Consortium for Educational Development in Higher Education, Perth, Australia http://www.csd.uwa.edu.au/iced2002/publication/index.html.

Wellein, M., Ragucci K. R., & Lapointe, M. (2009). A peer review process for classroom teaching.

American Journal of Pharmaceutical Education, 73(5), 1-7.

Weschke, B., & Canipe, S. (2010) The Faculty Evaluation Process: The First Step in Fostering Professional Development in an Online University. *Journal of College Teaching & Learning, 7*(1), 45-58.

Winchester, T. M., & Winchester, M. (2011). Exploring the impact of faculty reflection on weekly student evaluations of teaching. *International Journal for Academic Development, 16*(2), 119-131.

Wininger, S., & Birkholz, P. (2013). Sources of instructional feedback, job satisfaction, and basic psychological needs. *Innovative Higher Education, 38*(2), 159-170.

Wright, W.A., & O'Neill, M.C. (1995). Teaching improvement practices: International perspectives. In W.A. Wright (Ed.), *Successful Faculty Development Strategies.* Boston, MA: Anker Publishing Co.

Zakrajsek, T. (2006). Using evaluation data to improve teaching effectiveness. In P. Seldin (Ed.), *Evaluating faculty performance. A practical guide to assessing teaching, research and service* (pp. 166-179). Bolton, MA: Anker.

Appendix C

Teaching Frameworks

Teaching Standards Framework

The UK Professional Standards Framework

UK Quality Code for Higher Education

TEACHING STANDARDS FRAMEWORK **TSF 2013 : Blank Print Version**

This is a blank version of the TSF suitable for printing and reviewing.

Theme -Teaching

Teaching - Management Responsibilities

Criteria

TMR1 **Senior authority with institutional oversight of learning and teaching**
eg Deputy Vice-Chancellor (Academic) or equivalent with responsibility for learning and teaching in job description.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

TMR2 **Institutional academic structure facilitates quality learning and teaching**
eg Faculties, departments and disciplines are organised in a clear and systematic way with appropriate leadership structures with clearly defined roles.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

TMR3 **Inter-faculty quality assurance and quality enhancement process for learning and teaching**
eg Committee reporting to academic board (or equivalent) with responsibility for learning and teaching, moderation or standards processes in terms of reference.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

TMR4 **Faculty quality assurance and quality enhancement process for learning and teaching**
eg Reporting line within faculty for learning and teaching (eg dedicated learning and teaching committee with appropriate terms of reference and/or an associate dean with an appropriate job description).

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

TMR5 **Departmental quality assurance and quality enhancement process for learning and teaching**
eg Reporting line within department for learning and teaching (eg through a dedicated learning and teaching committee with appropriate terms of reference and/or a department learning and teaching director with an appropriate job description).

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

Teaching-Management Responsibilities- FOCUS AREA CONTEXTUAL STATEMENTS

Teaching - Planning

Criteria

TPL1 **Clearly defined values for learning and teaching incorporated into institutional-level planning**
eg Institutional-level plans incorporate a set of values that underpin learning and teaching (eg student centred perspective, diversity, quality assurance and quality enhancement).

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

TPL2 **Institutional-level teaching plans outline: a mechanism/framework for teaching quality assurance and quality enhancement**
eg Institutional-level plans include clear protocols, responsibilities, roles and timeframes for assuring and enhancing teaching quality.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

TPL3 **Institutional-level teaching plans outline: how to assure and enhance the qualifications of academic staff**
eg Institutional-level plans include strategies for employing quality academic staff (eg staff with teaching qualifications, teaching awards).

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

TPL4 **Institutional-level teaching plans outline: how to support and enhance quality teaching**
eg Institutional-level plans include strategies for ongoing professional development, grants and fellowships, workload allocation.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

TPL5 **Institutional-level teaching plans outline: how to recognise and reward quality teaching**
eg Institutional-level plans include strategies for academic promotions, awards, prizes and grants.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

TPL6 **Institutional-level teaching plans outline: how to benchmark teaching performance**
eg Institutional-level plans include strategies for benchmarking activities with other institutions, industry providers or internal departments.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

TPL7 **Institutional-level plans incorporate feedback from stakeholders for future planning**
eg Stakeholders provide feedback into, and on, teaching plans.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

TPL8 **Plans for quality teaching assurance and enhancement at a faculty level**
eg Faculty plans, based on higher-level institutional plans, have clear protocols, responsibilities, roles and timeframes for quality assurance and quality enhancement.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

TPL9 **Plans for quality teaching assurance and enhancement at a department level**
eg Departmental plans, based on higher-level faculty and institutional plans have clear protocols, responsibilities, roles and timeframes for quality assurance and quality enhancement.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

Teaching-Planning- FOCUS AREA CONTEXTUAL STATEMENTS

Teaching - Resources

Criteria

TRE1 **Allocation of adequate financial resources to support quality teaching include: provision of adequate human resources**
eg The institution evaluates systematically and transparently the levels of staffing required to conduct and support teaching adequately, sets appropriate targets (eg for staff : student ratios) and allocates resources accordingly.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

TRE2 **Allocation of adequate financial resources to support quality teaching include: recruiting and retaining quality academic staff**
eg Institutional financial resources to recruit and retain quality academic staff (eg staff with teaching qualifications, teaching awards).

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

TRE3 **Allocation of adequate financial resources to support quality teaching include: supporting and enhancing quality teaching**
eg Institutional financial resources for the ongoing professional development of staff (eg an office focused on research into learning and teaching, professional development, and enhancement) and allows staff time to pursue professional development.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

TRE4 **Allocation of adequate financial resources to support quality teaching include: recognition and reward for quality teaching**
eg Institution has financial resources to support academic promotions, awards, prizes and grants.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

Teaching-Resources- FOCUS AREA CONTEXTUAL STATEMENTS

Teaching - Policies and Procedures

Criteria

TPP1 **Established, systematic and accessible policy framework**
eg Institutional policy repository accessible to staff and students. The institution has clear and transparent procedures for the review and renewal of policy. Institution sets targets for policy quality (eg proportion of policy documents publicly accessible, proportion reviewed within review period and for reducing the number of out-of-date documents).

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

TPP2 **Policies and procedures reflect the institution's learning and teaching values**
eg Defined learning and teaching values can be tracked across a range of institutional policies and procedures.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

TPP3 **The institution has policies and procedures for: assuring and enhancing the quality of teaching**
eg Institutional policies include strategies and procedures for attracting and retaining quality academic staff (eg appointment policy, promotions policy).

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

TPP4 **The institution has policies and procedures for: supporting and enhancing teaching**
eg Institutional policies include strategies and procedures for the ongoing professional development of staff (eg training courses, grants and fellowships, staff health and wellbeing, workload allocation).

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

TPP5 **The institution has policies and procedures for: recognising and rewarding quality teaching**
eg Institutional policies have strategies and procedures for recognising and rewarding quality teaching (eg promotions, awards, prizes and grants).

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

TPP6 **The institution has policies and procedures for: benchmarking teaching quality**
eg Institutional policies include protocols and procedures for benchmarking with other institutions, including for establishing targets and conducting assessments.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

Teaching-Policies and Procedures- FOCUS AREA CONTEXTUAL STATEMENTS

Teaching - Practices

Criteria

TPR1 **Quality academic staff are recruited and retained**
eg Goals and strategies are in operation for recruiting and retaining quality academic staff (eg staff with teaching qualifications, teaching awards).

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

TPR2 **Ongoing professional development is encouraged and available to all academic staff**
eg Academic staff undertake professional development. Professional development is included in the

review and planning processes for all staff.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

TPR3 **Institution undertakes research into learning and teaching**
eg Teachers undertake pedagogical research leading to quality publications, grants, fellowships and exchanges.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

TPR4 **Institutional teaching responds to current research in learning and teaching**
eg Institution communicates innovations in research in learning and teaching to staff. Academic staff incorporate pedagogical research into their teaching practice as reported in departmental, faculty and institutional teaching reviews and staff professional development planning.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

TPR5 **Institutional learning and teaching plans are enacted**
eg Academic staff are aware of the content of institutional (including faculty/department) learning and teaching plans, learning and teaching goals and priorities, and address these in their practice. Subject outlines reflect learning and teaching plans. Student and other feedback indicate success in meeting plans.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

TPR6 **Quality teaching is recognised and rewarded**
eg The institution rewards and celebrates quality teaching (eg through promotions, awards, prizes and grants). The institution sets and reaches targets for applications for promotions, and for learning and teaching awards.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

TPR7 **Quality teaching practice is shared**
eg Academic staff undertake peer observation of teaching. Staff implement quality teaching practices. Staff collaboratively plan to implement effective learning experiences.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

Teaching-Practices- FOCUS AREA CONTEXTUAL STATEMENTS

Teaching - Outcomes

Criteria

TOU1 **Quality academic staff are recruited and retained**
eg The institution fulfils plans and meets established targets for the number of quality academic staff (eg based on information supplied in applications for appointment and promotion in relation to academic staff with formal teaching qualifications; numbers of teaching staff with teaching awards, grants or fellowships; and numbers of staff conducting pedagogical research).

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

TOU2 **Student performance meets defined goals**
eg The institution fulfils plans and meets established targets for student performance (eg measured by student result evaluation processes, and measurable attainment of learning outcomes).

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

TOU3 **Graduates are appropriately qualified**
eg Graduates attain graduate capabilities, workplace skills and discipline-based knowledge as defined in the Australian Qualifications Framework or relevant industry-based policy documents.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

TOU4 **Students report positive learning experience**
eg The institution meets established goals for student satisfaction in national, sector-wide and internal student feedback.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

TOU5 **Graduates find appropriate employment**
eg The institution meets established goals for graduate employment.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

TOU6 **Employers report satisfaction with graduate employees**
eg Surveys of employers indicate employer satisfaction with institutional teaching.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

Teaching-Outcomes- FOCUS AREA CONTEXTUAL STATEMENTS

Teaching - Monitoring and Evaluation

Criteria

TME1 **Regular reviews of: institutional management responsibilities for learning and teaching**
eg Drawing on feedback from staff, students and stakeholders, the institution conducts regular reviews of its: management responsibilities for learning and teaching according to clear protocols, responsibilities, roles and timeframes which specify how such reviewing is reported and acted upon.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

TME2 **Regular reviews of: learning and teaching plans**
eg Drawing on feedback from staff, students and stakeholders, the institution conducts regular reviews of its: learning and teaching plans according to clear protocols, responsibilities, roles and timeframes which specify how such reviewing is reported and acted upon.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

TME3 **Regular reviews of: resources for learning and teaching**
eg Drawing on feedback from staff, students and stakeholders, the institution conducts regular reviews of its: learning and teaching resources according to clear protocols, responsibilities, roles and timeframes which specify how such reviewing is reported and acted upon.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

TME4 **Regular reviews of: learning and teaching policy and procedures**
eg Drawing on feedback from staff, students and stakeholders, the institution conducts regular reviews of its: learning and teaching policy and procedures according to clear protocols, responsibilities, roles and timeframes which specify how such reviewing is reported and acted upon.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

TME5 **Regular reviews of: learning and teaching practices**
eg Drawing on feedback from staff, students and stakeholders, the institution conducts regular reviews of its: learning and teaching practices according to clear protocols, responsibilities, roles and timeframes which specify how such reviewing is reported and acted upon.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

TME6 **Regular reviews of: learning and teaching outcomes**
eg Drawing on feedback from staff, students and stakeholders, the institution conducts regular reviews of its: learning and teaching outcomes according to clear protocols, responsibilities, roles and timeframes which specify how such reviewing is reported and acted upon.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

TME7 **Regular reviews include national and international benchmarking**
eg The institution has an established benchmarking program at department, faculty and institutional level including clear protocols, responsibilities, roles and timeframes for agreeing on standards, and how the review will be reported and implemented.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

Teaching-Monitoring and Evaluation- FOCUS AREA CONTEXTUAL STATEMENTS

Theme -Learning Environment

Learning Environment - Management Responsibilities

Criteria

LMR1 **Senior authority with institutional oversight of learning and teaching resources and support**
eg Deputy Vice-Chancellor (Academic) or equivalent with responsibility for learning and teaching resources and support in job description.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

LMR2 **The institution has a senior officer/committee for: assuring and enhancing student welfare**
eg Senior officer responsible for student affairs (including pastoral care, academic advice and referral to student services).

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

LMR3 **The institution has a senior officer/committee for: quality assurance and quality enhancement in the provision of learning spaces**
eg Delegated authority entrusted with the administration of learning and teaching spaces, with effective lines of communication with faculties and departments.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

LMR4 **The institution has a senior officer/committee for: quality assurance and quality enhancement in the provision of information resources**
eg Delegated authority entrusted with the administration of information resources, with effective lines of communication with faculties and departments.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

LMR5 **The institution has a senior officer/committee for: quality assurance and quality enhancement in the provision of learning and teaching technologies**
eg Delegated authority entrusted with the administration of learning and teaching technologies, with effective lines of communication with faculties and departments.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

Learning Environment-Management Responsibilities- FOCUS AREA CONTEXTUAL STATEMENTS

Learning Environment - Planning

Criteria

LPL1 **Institutional-level plans include: arrangements for academic support for students**
eg Institutional-level plans ensure the appropriate provision of study skills training, transition support, academic advice, mentoring and student orientations.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

LPL2 **Institutional-level plans include: health and wellbeing support for students**
eg Institutional-level plans ensure the appropriate provision of accommodation, careers advice, medical facilities, counselling, financial support, scholarships.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

LPL3 **Institutional-level plans include: clearly defined expectations about student rights and responsibilities**
eg Institutional-level plans clearly communicate expectations about student rights and responsibilities (eg through a student charter or code of conduct).

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

LPL4 **Institutional-level plans include: appropriate provision of learning spaces**
eg Institutional-level plans define goals and strategies for the assessment, allocation and management of learning spaces.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

LPL5 **Institutional-level plans include: appropriate provision of information resources**
eg Institutional-level plans define goals and strategies for the assessment, allocation and management of information resources (eg library).

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

LPL6 **Institutional-level plans include: provision of quality technology to support learning**
eg Institutional-level plans define goals and strategies for the assessment, allocation and management of technology (eg network, hardware).

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

LPL7 **The faculty has plans for quality learning support**
eg Faculty plans have clear protocols, responsibilities, roles and timeframes for quality learning support, and both enable and fulfil institutional plans.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

LPL8 **The department has plans for quality learning support**
eg Department plans have clear protocols, responsibilities, roles and timeframes for quality learning support, and both enable and fulfil faculty and institutional plans.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

LPL9 **The institution has plans for systematic data collection**
eg Plans include strategies, clear protocols, responsibilities, roles and timeframes for systematic data collection (eg for statutory reporting).

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

Learning Environment-Planning- FOCUS AREA CONTEXTUAL STATEMENTS

Learning Environment - Resources

Criteria

LRE1 **Allocation of financial and human resources provide: <u>academic support for students</u>**
eg The institution assesses the need for resources for student academic support taking into account its diverse student population (eg study skills, transition support, mentoring and orientation), sets appropriate targets (eg for cost to students and waiting times) and allocates resources accordingly.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

LRE2 **Allocation of financial and human resources provide: <u>support for student health and wellbeing</u>**

eg The institution assesses the need for resources for student health and wellbeing taking into account its diverse student population (eg accommodation services, careers advice, medical facilities, counselling, financial support, scholarships etc) , sets appropriate targets (eg for cost to students and waiting times) and allocates resources accordingly.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

LRE3 **Allocation of financial and human resources provide: <u>adequate infrastructure</u>**

eg The institution assesses the need for resources to provide adequate materials and infrastructure taking into account its diverse student population (eg learning spaces and spaces for support services), sets appropriate targets (eg utilisation rates of learning spaces, effective class sizes, staff : student ratios, quantity of on-campus accommodation) and allocates resources accordingly.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

LRE4 **Allocation of financial and human resources provide: <u>appropriate information resources</u>**

eg The institution assesses the need for the provision of adequate information resources taking into account its diverse student population (eg library resources for both on-campus and distance students), sets appropriate targets (eg number of resources on request, proportion of students accessing resources, waiting times for borrowing) and allocates resources accordingly.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

LRE5 **Allocation of financial and human resources provide: <u>appropriate technology to support learning</u>**

eg The institution assesses the need for resources for technology taking into account its diverse student population, sets appropriate targets (eg for cost to students, renewal of equipment, range of software, and waiting times) and allocates resources accordingly.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

Learning Environment-Resources- FOCUS AREA CONTEXTUAL STATEMENTS

Learning Environment - Policies and Procedures

Criteria

LPP1 **The institution has policies and procedures to provide access to: academic support**
eg Institutional policies and procedures outline clear protocols, responsibilities, roles and timeframes for the provision of academic support for students (eg in the areas of study skills, transition support, mentoring and orientation).

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

LPP2 **The institution has policies and procedures to provide access to: health and wellbeing support**
eg Institutional policies and procedures outline clear protocols, responsibilities, roles and timeframes for the provision of services to support the health and wellbeing of students (eg in the areas of accommodation services, careers advice, medical facilities, counselling, financial support, scholarships etc).

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

LPP3 **The institution has policies and procedures for: the use of learning spaces**
eg Institutional policies and procedures outline clear protocols, responsibilities, roles and timeframes for assessing a need for space allocation and for monitoring of the use of learning spaces in the light of established goals and targets.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

LPP4 **The institution has policies and procedures for: <u>the provision of information resources</u>**
eg Institutional policies and procedures outline clear protocols, responsibilities, roles and timeframes for assessing a need for information resources and for monitoring of the use of information resources in the light of established goals and targets.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

LPP5 **The institution has policies and procedures for: <u>the provision of technology to support learning and teaching</u>**
eg Institutional policies and procedures outline clear protocols, responsibilities, roles and timeframes for assessing a need for learning and teaching technologies, and for monitoring of the use of learning and teaching technologies in the light of established goals and targets.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

Learning Environment-Policies and Procedures- FOCUS AREA CONTEXTUAL STATEMENTS

Learning Environment - Practices

Criteria

LPR1 **Academic support services are available**
eg The institution establishes appropriate facilities for student academic support in response to identified student needs.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

LPR2 **Students access academic support services**
eg Assessment of student needs identifies targets for the provision and use of student academic support services and the targets are met.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

LPR3 **Non-academic support services are available**
eg The institution establishes facilities for non-academic student support in response to measured student need and available institutional resources.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

LPR4 **Students access non-academic support services**
eg Assessment of student need identifies targets for the provision and use of non-academic student support services and the targets are met.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

LPR5 **The institution systematically collects and uses data about the provision of support services**
eg The institution analyses and uses data collected (eg for statutory reporting purposes)

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

Learning Environment-Practices- FOCUS AREA CONTEXTUAL STATEMENTS

Learning Environment - Outcomes

Criteria

LOU1 **The provision of academic support services is adequate**
eg Student feedback and other indicators (eg improved student performance, higher retention rates, a reduction in the number of disputes and complaints) indicate that the institution is meeting established goals for the provision of academic support services.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

LOU2 **The provision of non-academic support services is adequate**
eg Student feedback and other indicators (eg improved student performance, higher retention rates, a reduction in the number of disputes and complaints) indicate that established goals for the provision of non-academic support services are being met.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

LOU3 **Students report a quality learning experience**
eg Student involvement in cultural, social, sporting and volunteering activities improve the quality of student learning experience.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

Learning Environment-Outcomes- FOCUS AREA CONTEXTUAL STATEMENTS

Learning Environment - Monitoring and Evaluation

Criteria

LME1 **The institution regularly reviews the provision of: academic support services**
eg Drawing on feedback from staff, students and stakeholders, the institution conducts regular reviews of its provision of: academic support services according to clear protocols, responsibilities, roles and timeframes which specify how the review will be reported and implemented.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

LME2 **The institution regularly reviews the provision of: non-academic support services**
eg Drawing on feedback from staff, students and stakeholders, the institution conducts regular reviews of its provision of: non-academic support services according to clear protocols, responsibilities, roles and timeframes which specify how the review will be reported and implemented.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

LME3 **The institution regularly reviews the provision of: infrastructure for learning and teaching**
eg Drawing on feedback from staff, students and stakeholders, the institution conducts regular reviews of its provision of: adequate infrastructure for learning and teaching, according to clear protocols, responsibilities, roles and timeframes which specify how the review will be reported and implemented.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

LME4 **The institution regularly reviews the provision of: information resources**
eg Drawing on feedback from staff, students and stakeholders, the institution conducts regular reviews of its provision of: information resources, according to clear protocols, responsibilities, roles and timeframes which specify how the review will be reported and implemented.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

LME5 **The institution regularly reviews the provision of: technological support for learning and teaching**
eg Drawing on feedback from staff, students and stakeholders, the institution conducts regular reviews of its provision of: adequate technological resources for learning and teaching, according to clear protocols, responsibilities, roles and timeframes which specify how the review will be reported and implemented.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

Learning Environment-Monitoring and Evaluation- FOCUS AREA CONTEXTUAL STATEMENTS

Theme -Curriculum

Curriculum - Management Responsibilities

Criteria

CMR1 **Senior authority with institutional oversight of curriculum quality**
eg Deputy Vice-Chancellor (Academic) or equivalent with responsibility for curriculum quality.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

CMR2 **Institutional curriculum quality assurance and quality enhancement process**
eg Committee reporting to an academic board (or equivalent) with curriculum quality responsibilities in its terms of reference.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

CMR3 **Faculty curriculum quality assurance and quality enhancement process**
eg Reporting line within faculty for curriculum quality (eg dedicated curriculum quality committee with appropriate terms of reference and/or an associate dean with appropriate job description).

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

CMR4 **Department curriculum quality assurance and quality enhancement process**
eg Reporting line within department for curriculum quality (eg dedicated curriculum quality committee with appropriate terms of reference and/or a department learning and teaching director with appropriate job description).

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

Curriculum-Management Responsibilities- FOCUS AREA CONTEXTUAL STATEMENTS

Curriculum - Planning

Criteria

CPL1 **Institutional-level plans set standards and protocols for curriculum quality assurance and quality enhancement, including: curriculum development**
eg Institutional-level plans identify responsibilities, procedures and timeframes for the development of high quality curriculum, and, where relevant, its conformity to professional accreditation requirements.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

CPL2 **Institutional-level plans set standards and protocols for curriculum quality assurance and quality enhancement, including: curriculum review**
eg Institutional-level plans identify responsibilities, procedures and timeframes for curriculum review.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

CPL3 **Institutional-level plans set standards and protocols for curriculum quality assurance and quality enhancement, including: curriculum renewal**
eg Institutional-level plans identify responsibilities, procedures and timeframes for the periodic renewal of curriculum.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

CPL4 **Institutional-level plans set standards and protocols for curriculum quality assurance and quality enhancement, including: curriculum benchmarking**
eg Institutional-level plans identify responsibilities, procedures and timeframes for the benchmarking of curriculum between appropriate departments and institutions.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

CPL5 **Faculty plans set standards and protocols for curriculum quality assurance and quality enhancement**
eg Faculty plans align with institutional curriculum quality assurance and quality enhancement plans.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

CPL6 **Department plans set standards and protocols for curriculum quality assurance and quality enhancement**
eg Department plans align with institutional and faculty curriculum quality assurance and quality enhancement plans.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

Curriculum-Planning- FOCUS AREA CONTEXTUAL STATEMENTS

Curriculum - Resources

Criteria

CRE1 **Allocation of financial resources to support curriculum quality assurance and quality enhancement**
eg The institution supplies sufficient financial resources for the cost of curriculum development, review, renewal and benchmarking (eg costs of equipment, materials, technology, communication, travel and dissemination).

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

CRE2 **Allocation of human resources to support curriculum quality assurance and quality enhancement**
eg The institution allocates sufficient qualified academic and professional staff and staff-time for curriculum development, review, renewal and benchmarking.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

CRE3 **The institution provides academic and technical expertise to support curriculum quality assurance and quality enhancement**

eg Academic and professional staff have access to expert advice in curriculum development, review, renewal and benchmarking.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

Curriculum-Resources- FOCUS AREA CONTEXTUAL STATEMENTS

Curriculum - Policies and Procedures

Criteria

CPP1 **Institutional policies and procedures enable quality assurance and quality enhancement in the areas of: curriculum development**
eg Institutional policies and procedures specify: responsibilities for initiating and approving curriculum development, as well as required evidence and relevant timeframes.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

CPP2 **Institutional policies and procedures enable quality assurance and quality enhancement in the areas of: curriculum design**
eg Institutional policies and procedures specify: standards of subject design (including quality of subject materials) in terms of teaching and assessment practice and learning outcomes.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

CPP3 **Institutional policies and procedures enable quality assurance and quality enhancement in the areas of: curriculum and subject reviews**
eg Institutional policies and procedures specify: responsibilities, timeframes and documentation for subject and curriculum review, including protocols for gathering and assessing feedback from students, staff and other stakeholders.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

CPP4 **Institutional policies and procedures enable quality assurance and quality enhancement in the areas of: curriculum and subject renewal**
eg Institutional policies and procedures specify: responsibilities, timeframes and documentation for subject and curriculum renewal, including protocols for gathering and assessing feedback from students, staff and other stakeholders.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

CPP5 **Institutional policies and procedures enable quality assurance and quality enhancement in the areas of: benchmarking**
eg Institutional policies and procedures specify: responsibilities, timeframes and documentation for subject and curriculum benchmarking to ensure (i) parity of standards within the institution and (ii) comparability of standards between universities.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

CPP6 **Faculty policies and procedures enable quality assurance and quality enhancement in the areas of curriculum development, design, review, renewal and benchmarking**
eg Faculty policies and procedures specify responsibilities, timeframes and documentation for curriculum development, design, review, renewal and benchmarking in co-ordination with institutional policies and procedures.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

CPP7 **Department policies and procedures enable quality assurance and quality enhancement in the areas of curriculum development, design, review, renewal and benchmarking**
eg Department policies and procedures specify responsibilities, timeframes and documentation for curriculum development, design, review, renewal and benchmarking in co-ordination with

institutional and faculty policies and procedures.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

Curriculum-Policies and Procedures- FOCUS AREA CONTEXTUAL STATEMENTS

Curriculum - Practices

Criteria

CPR1 **The curriculum is informed by current research in the discipline**
eg Review and benchmarking processes confirm the curriculum is informed by current research in the discipline.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

CPR2 **The curriculum conforms to professional accreditation standards (where appropriate)**
eg Courses are awarded professional accreditation (where appropriate).

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

CPR3 **The curriculum is designed to ensure specified student learning outcomes**
eg Staff, student and stakeholder feedback, peer review, and review and benchmarking processes confirm link between curriculum design and student learning outcomes.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

CPR4 **The curriculum is presented using the most appropriate technologies and modes of delivery**
eg Staff, student and stakeholder feedback, peer review, and review and benchmarking processes confirm effectiveness of teaching technologies and modes of delivery.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

CPR5 **Teaching materials provide students with all necessary information about content, requirements and resources**
eg Student feedback confirms quality of teaching materials.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

CPR6 **Assessment practice is aligned to stated learning outcomes**
eg Student feedback, review and benchmarking processes confirm link between assessment and stated student learning outcomes.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

Curriculum-Practices- FOCUS AREA CONTEXTUAL STATEMENTS

Curriculum - Outcomes

Criteria

COU1 **The institution meets appropriate targets for student achievement**

eg Department, faculty and institution confirm students have met specified learning outcomes.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

COU2 **The institution develops graduate capabilities, workplace skills and discipline-based knowledge**
eg Feedback from staff, students and stakeholders (eg employers, other academic institutions, professional bodies and other relevant groups) confirm graduates have attained graduate capabilities, workplace skills and discipline-based knowledge.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

COU3 **The institution meets appropriate targets for student satisfaction with the curriculum**
eg Student feedback, application and retention rates confirm student satisfaction with the curriculum.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

COU4 **The institution meets appropriate targets for employer satisfaction with the curriculum**
eg Employer feedback and application rates confirm employer satisfaction with and reputation of the curriculum.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

Curriculum-Outcomes- FOCUS AREA CONTEXTUAL STATEMENTS

Curriculum - Monitoring and Evaluation

Criteria

CME1 **The institution regularly reviews: institutional strategies for curriculum development**
eg Drawing on feedback from staff, students and stakeholders: institution, faculties and departments undertake regular reviews of curriculum strategies and report on outcomes.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

CME2 **The institution regularly reviews: curriculum content**
eg Drawing on feedback from staff, students and stakeholders: departments and faculties provide written reviews of curriculum content according to a specified schedule.

O **Yes** O **Yes, but** O **No** O **N/A**

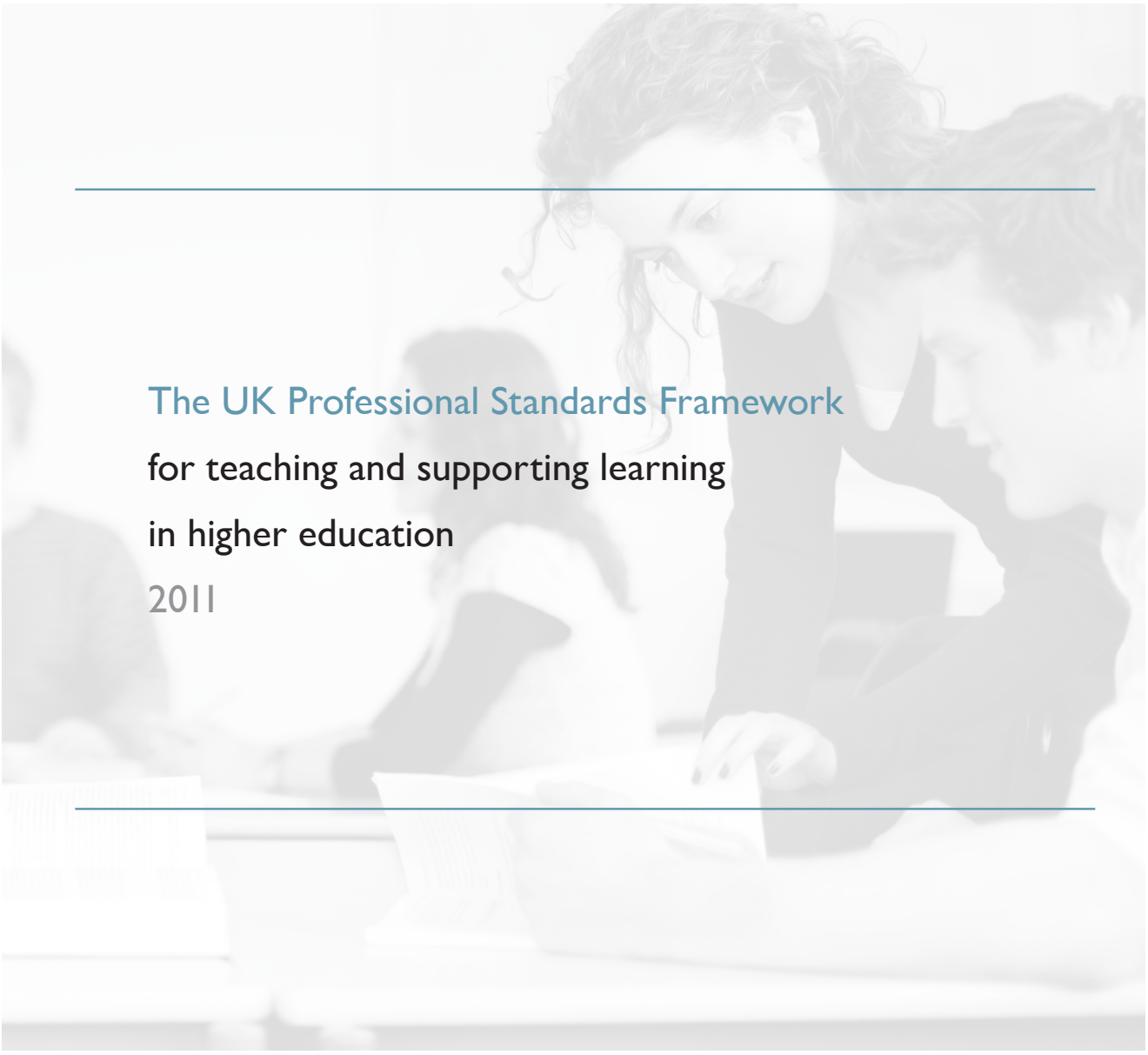
CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

CME3 **The institution regularly reviews: curriculum development, design and delivery**
eg Drawing on feedback from staff, students and stakeholders: departments, faculties and relevant institutional offices provide written feedback on the effectiveness of curriculum development, design and delivery.

O **Yes** O **Yes, but** O **No** O **N/A**

CRITERIA CONTEXTUAL STATEMENTS (OPTIONAL)

Curriculum-Monitoring and Evaluation- FOCUS AREA CONTEXTUAL STATEMENTS



The UK Professional Standards Framework

for teaching and supporting learning

in higher education

2011

Aims of the Framework

The UK Professional Standards Framework:

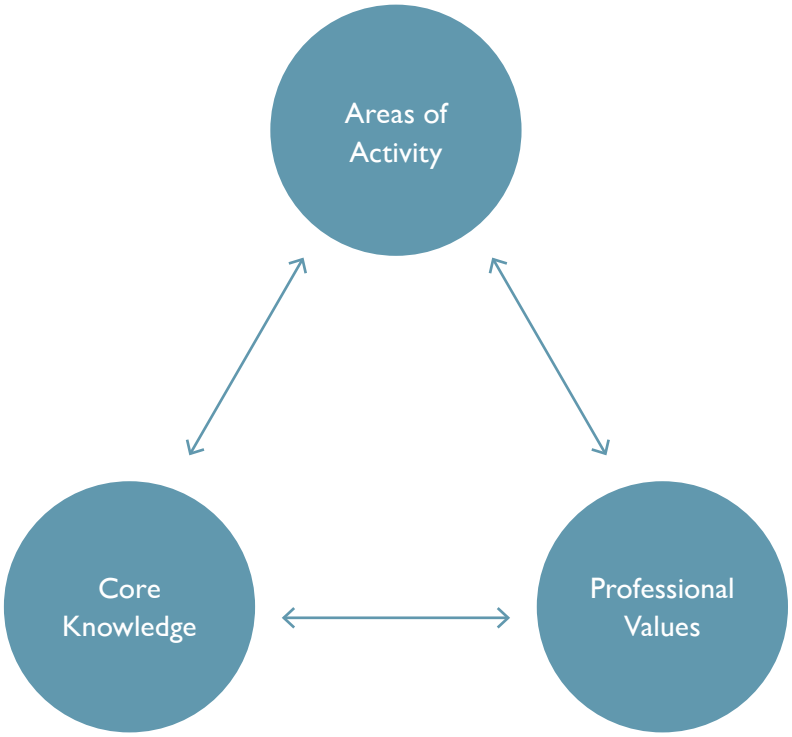
1. Supports the initial and continuing professional development of staff engaged in teaching and supporting learning

2. Fosters dynamic approaches to teaching and learning through creativity, innovation and continuous development in diverse academic and/or professional settings

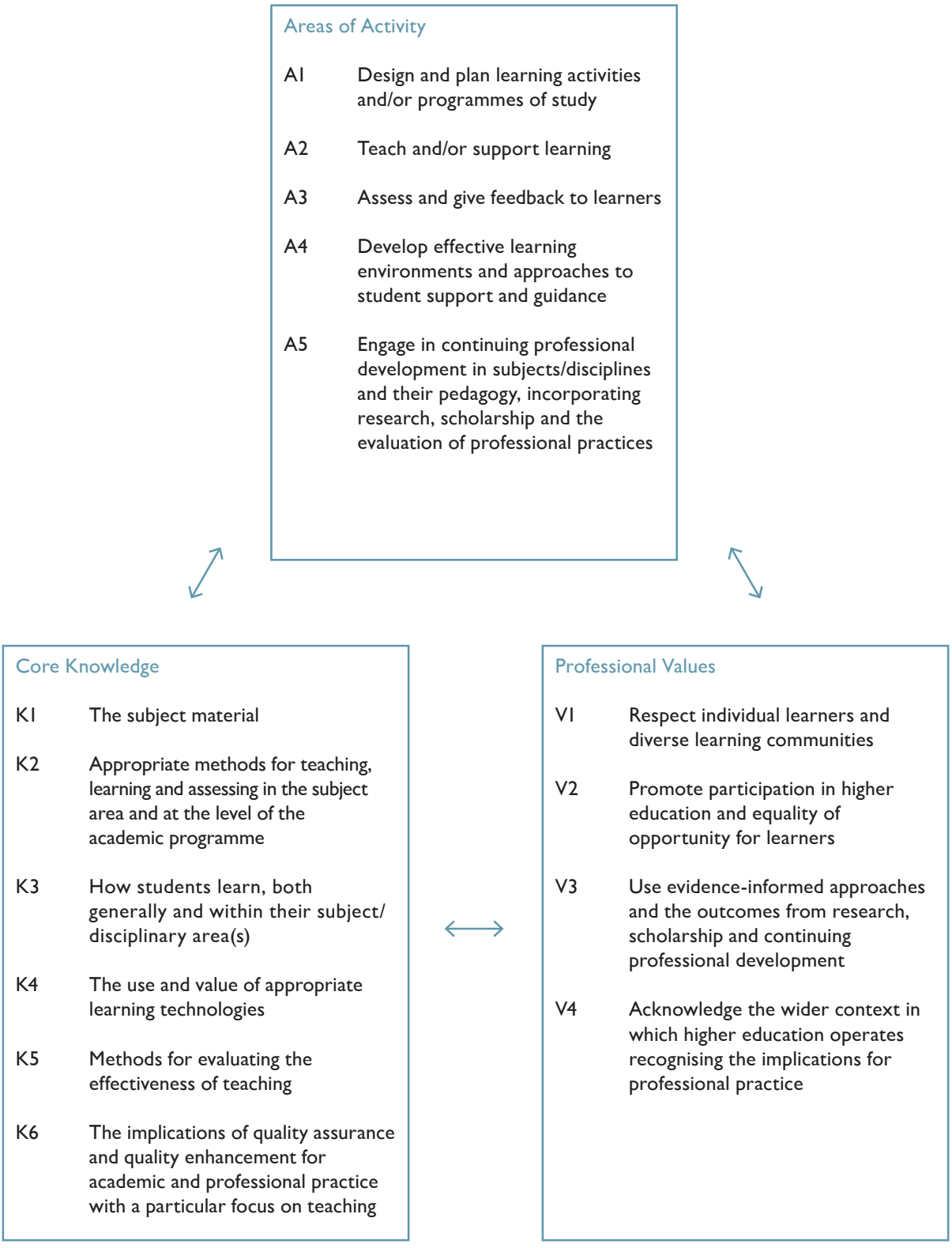
3. Demonstrates to students and other stakeholders the professionalism that staff and institutions bring to teaching and support for student learning

4. Acknowledges the variety and quality of teaching, learning and assessment practices that support and underpin student learning

5. Facilitates individuals and institutions in gaining formal recognition for quality-enhanced approaches to teaching and supporting learning, often as part of wider responsibilities that may include research and/or management activities



Dimensions of the Framework



Descriptor 1	Typical individual role/career stage	Related HEA recognition
Demonstrates an understanding of specific aspects of effective teaching, learning support methods and student learning. Individuals should be able to provide evidence of: I. Successful engagement with at least two of the five Areas of Activity II. Successful engagement in appropriate teaching and practices related to these Areas of Activity III. Appropriate Core Knowledge and understanding of at least K1 and K2 IV. A commitment to appropriate Professional Values in facilitating others' learning V. Relevant professional practices, subject and pedagogic research and/or scholarship within the above activities VI. Successful engagement, where appropriate, in professional development activity related to teaching, learning and assessment responsibilities	Individuals able to provide evidence of effectiveness in relation to their professional role(s), which, typically, will include at least some teaching and/or learning support responsibilities. This teaching and learning role may sometimes be undertaken with the assistance of more experienced teachers or mentors. Typically, those likely to be at Descriptor 1 (D1) include: a. Early career researchers with some teaching responsibilities (e.g. PhD students, GTAs, contract researchers/post doctoral researchers etc.) b. Staff new to teaching (including those with part-time academic responsibilities) c. Staff who support academic provision (e.g. learning technologists, learning developers and learning resource/library staff) d. Staff who undertake demonstrator/ technician roles that incorporate some teaching-related responsibilities e. Experienced staff in relevant professional areas who may be new to teaching and/or supporting learning, or who have a limited teaching portfolio	Associate Fellow

Descriptor 2	Typical individual role/career stage	Related HEA recognition
Demonstrates a broad understanding of effective approaches to teaching and learning support as key contributions to high quality student learning. Individuals should be able to provide evidence of: I. Successful engagement across all five Areas of Activity II. Appropriate knowledge and understanding across all aspects of Core Knowledge III. A commitment to all the Professional Values IV. Successful engagement in appropriate teaching practices related to the Areas of Activity V. Successful incorporation of subject and pedagogic research and/or scholarship within the above activities, as part of an integrated approach to academic practice VI. Successful engagement in continuing professional development in relation to teaching, learning, assessment and, where appropriate, related professional practices	Individuals able to provide evidence of broadly based effectiveness in more substantive teaching and supporting learning role(s). Such individuals are likely to be established members of one or more academic and/or academic-related teams. Typically, those likely to be at Descriptor 2 (D2) include: a. Early career academics b. Academic-related and/or support staff holding substantive teaching and learning responsibilities c. Experienced academics relatively new to UK higher education d. Staff with (sometimes significant) teaching-only responsibilities including, for example, within work-based settings	Fellow

Descriptor 3	Typical individual role/career stage	Related HEA recognition
Demonstrates a thorough understanding of effective approaches to teaching and learning support as a key contribution to high quality student learning. Individuals should be able to provide evidence of: I. Successful engagement across all five Areas of Activity II. Appropriate knowledge and understanding across all aspects of Core Knowledge III. A commitment to all the Professional Values IV. Successful engagement in appropriate teaching practices related to the Areas of Activity V. Successful incorporation of subject and pedagogic research and/or scholarship within the above activities, as part of an integrated approach to academic practice VI. Successful engagement in continuing professional development in relation to teaching, learning, assessment, scholarship and, as appropriate, related academic or professional practices VII. Successful co-ordination, support, supervision, management and/or mentoring of others (whether individuals and/or teams) in relation to teaching and learning	Individuals able to provide evidence of a sustained record of effectiveness in relation to teaching and learning, incorporating for example, the organisation, leadership and/or management of specific aspects of teaching and learning provision. Such individuals are likely to lead or be members of established academic teams. Typically, those likely to be at Descriptor 3 (D3) include: a. Experienced staff able to demonstrate, impact and influence through, for example, responsibility for leading, managing or organising programmes, subjects and/or disciplinary areas b. Experienced subject mentors and staff who support those new to teaching c. Experienced staff with departmental and/or wider teaching and learning support advisory responsibilities within an institution	Senior Fellow

Descriptor 4	Typical individual role/career stage	Related HEA recognition
Demonstrates a sustained record of effective strategic leadership in academic practice and academic development as a key contribution to high quality student learning. Individuals should be able to provide evidence of: I. Active commitment to and championing of all Dimensions of the Framework, through work with students and staff, and in institutional developments II. Successful, strategic leadership to enhance student learning, with a particular, but not necessarily exclusive, focus on enhancing teaching quality in institutional, and/or (inter)national settings III. Establishing effective organisational policies and/or strategies for supporting and promoting others (e.g. through mentoring, coaching) in delivering high quality teaching and support for learning IV. Championing, within institutional and/or wider settings, an integrated approach to academic practice (incorporating, for example, teaching, learning, research, scholarship, administration etc.) V. A sustained and successful commitment to, and engagement in, continuing professional development related to academic, institutional and/or other professional practices	Individuals, as highly experienced academics, able to provide evidence of a sustained and effective record of impact at a strategic level in relation to teaching and learning, as part of a wider commitment to academic practice. This may be within their institution or wider (inter)national settings. Typically, those likely to be at Descriptor 4 (D4) include: a. Highly experienced and/or senior staff with wide-ranging academic or academic-related strategic leadership responsibilities in connection with key aspects of teaching and supporting learning b. Staff responsible for institutional strategic leadership and policy-making in the area of teaching and learning c. Staff who have strategic impact and influence in relation to teaching and learning that extends beyond their own institution	Principal Fellow

Framework Guidance Notes (FGN)

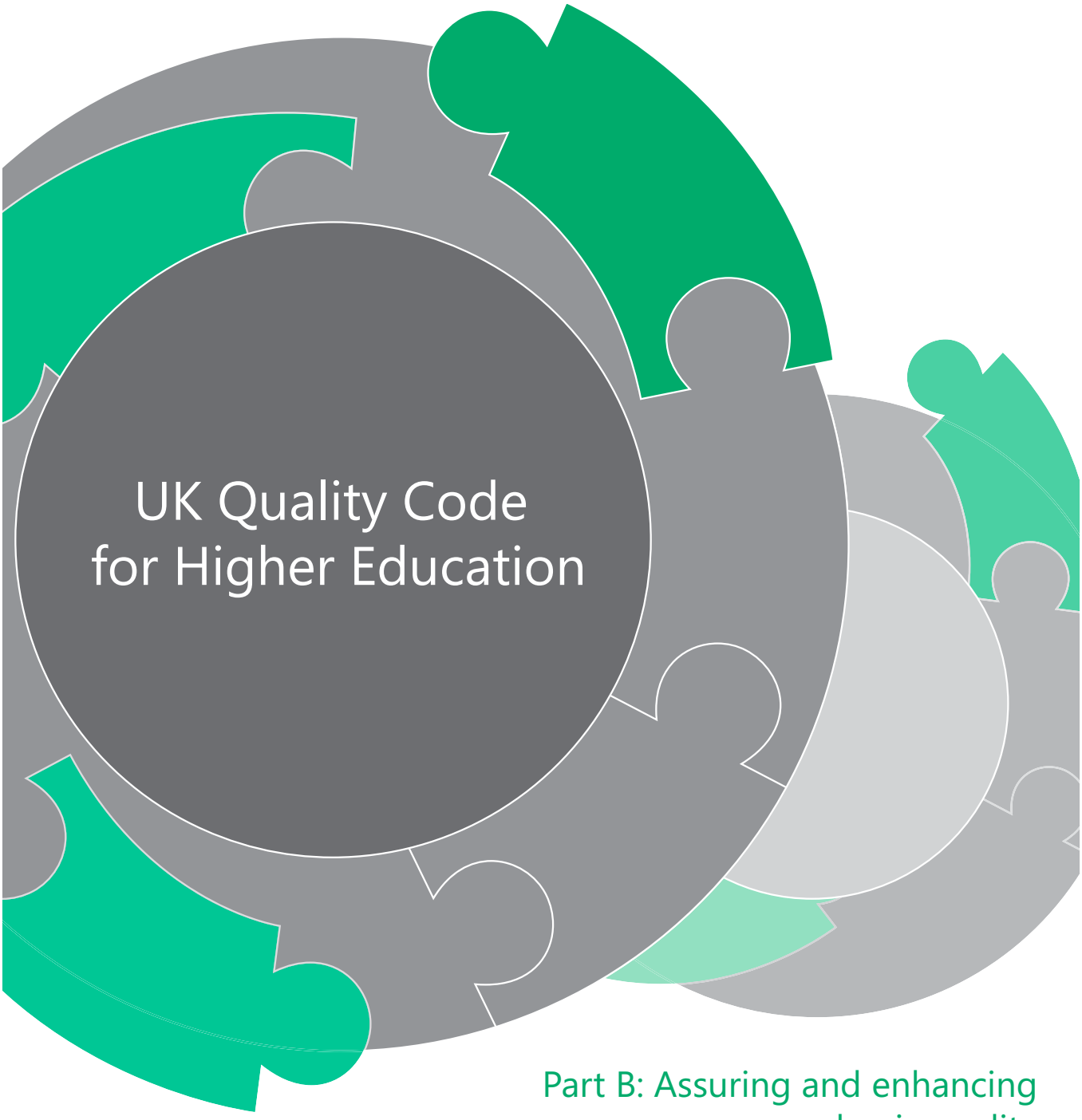
The UK Professional Standards Framework is supplemented and supported by a series of Framework Guidance Notes (FGN). These are designed to highlight and disseminate good practice in a given area as well as outline issues that institutions and individuals may want to consider in using the UK Professional Standards Framework.

Relationship to the Higher Education Academy National Accreditation Scheme

The Higher Education Academy recognises the importance and value of the UK Professional Standards Framework (UKPSF) and aligns its Fellowship categories with the Descriptors. Its accreditation scheme provides a national professional benchmarking of provision that reflects the best practices in the sector. Guidance and support is offered through the work of the Higher Education Academy to higher education institutions wishing to be accredited for their application and use of the UKPSF.

Please see http://www.heacademy.ac.uk for further details.

QAA



Part B: Assuring and enhancing academic quality

Chapter B3: Learning and teaching

Contents

About the Quality Code

The UK Quality Code for Higher Education (the Quality Code) is the definitive reference point for all UK higher education providers.[1] It makes clear what higher education providers are required to do, what they can expect of each other, and what the general public can expect of them. The Quality Code covers all four nations of the UK and all providers of UK higher education operating overseas. It protects the interests of all students, regardless of where they are studying or whether they are full-time, part-time, undergraduate or postgraduate students.

Each Chapter contains a single Expectation, which expresses the key principle that the higher education community has identified as essential for the assurance of academic standards and quality within the area covered by the Chapter. Higher education providers reviewed by the Quality Assurance Agency for Higher Education (QAA) are required to meet all the Expectations. The manner in which they do so is their own responsibility. QAA carries out reviews to check whether higher education providers are meeting the Expectations.[2]

Each Chapter has been developed by QAA through an extensive process of consultation with higher education providers; their representative bodies; the National Union of Students (NUS); professional, statutory and regulatory bodies; and other interested parties.

Higher education providers are also responsible for meeting the requirements of legislation and any other regulatory requirements placed upon them, for example by funding bodies. The Quality Code does not interpret legislation nor does it incorporate statutory or regulatory requirements. Sources of information about other requirements and examples of guidance and good practice are signposted within the Chapter where appropriate. Higher education providers are responsible for how they use these resources.

The Expectation in each Chapter is accompanied by a series of Indicators that reflect sound practice, and through which providers can demonstrate they are meeting the relevant Expectation. Indicators are not designed to be used as a checklist; they are intended to help providers reflect on and develop their regulations, procedures and practices to demonstrate that the Expectations in the Quality Code are being met. Each Indicator is numbered and printed in bold and is supported by an explanatory note that gives more information about it, together with examples of how the Indicator may be interpreted in practice.

The *UK Quality Code for Higher Education: General introduction*[3] should be considered in conjunction with this document. It provides a technical introduction for users, including guidance concerning the terminology used and a quick-reference glossary.

About this Chapter

This publication is a new Chapter of the Quality Code. It incorporates and supersedes the *Code of practice for the assurance of academic quality and standards in higher education (Code of practice), Section 2: Collaborative provision and flexible and distributed*

[1] www.qaa.ac.uk/qualitycode

[2] www.qaa.ac.uk/InstitutionReports/types-of-review

[3] www.qaa.ac.uk/publications/informationandguidance/pages/quality-code-introduction.aspx

learning (*including e-learning*) (2010), Part B: Aspects specific to flexible and distributed learning, and the *Code of practice, Section 9: Work-based and placement learning* (2007). It also incorporates and supersedes those parts of the *Code of practice, Section 3: Disabled students* (2010) relating to learning and teaching. The evaluation of the Academic Infrastructure and consultation on subsequent changes which resulted in the development of the Quality Code identified the need for the introduction of a Chapter on learning and teaching covering all modes of study.[4]

This draft was subject to public consultation between May and July 2012. The final version of the Chapter was published in September 2012 and will be used as a reference point in reviews coordinated by QAA from **August 2013**.

Learning and teaching

This Chapter focuses on the learning opportunities that higher education providers make available to students and on the staff who teach and who support learning, including those staff who are not employees of the higher education provider and/or are not based at the provider. It applies to any learning opportunities that lead to a UK higher education award or award of credit, whether through short courses involving single modules or multi-year programmes of study. It covers students studying at all academic levels in *Chapter A1: The national level*[5] (that is, all undergraduate, master's and doctoral students) and irrespective of their:

- location - for example, campus-based, on placement or otherwise in a workplace, distance learning, or with a collaborative partner within the UK or internationally
- mode of study - for example face to face, e-learning, blended learning, or work-based learning, whether full-time or part-time
- academic subject
- age, disability, gender reassignment, pregnancy or maternity, race, ethnic origin or national identity, religion or belief, sex, or sexual orientation
- previous educational background, breaks since the last experience of academic study, or time spent in the workplace
- nationality or fee status - for example home, European Union or international.

UK higher education embraces a wide variety of modes of both learning and teaching, which can be used in different combinations. These include flexible and distributed learning, work-based or placement learning, and technology-enhanced learning. At one end of the continuum, programme delivery, learner support and assessment are all provided by staff of the degree-awarding body on its campus(es). The other end of the continuum can be represented by a distance-learner who:

- has no direct contact with the degree-awarding body, its staff or other students
- has a programme of study delivered by an organisation that is not the degree-awarding body, and

[4] *Changes to the Academic Infrastructure: final report* (June 2011): www.qaa.ac.uk/publications/informationandguidance/pages/changes-to-academic-infrastructure.aspx.

[5] www.qaa.ac.uk/Publications/InformationAndGuidance/Pages/quality-code-A1.aspx

- is provided with support for learning from an organisation that is neither part of the delivery organisation nor the degree-awarding body.

Between these two ends of the continuum, there are many possible combinations involving learning and teaching, support, and assessment delivered by the degree-awarding body and/or delivering partner, and/or support provider. Some or all of the learning opportunities may be provided, for example, by an employer or organisation offering a work-based or placement learning opportunity, or an employer supporting its employee(s) on a higher education programme where the workplace is used as a learning environment. Work-based or placement learning opportunities can take place outside the UK, for example through exchange schemes.

The Chapter **does not** assume that:

- a student's place of study is physically located within the higher education provider whose academic award will be given on successful completion of the programme of study
- a student's programme of study is delivered directly by the degree-awarding body
- a student is directly supported by staff of the degree-awarding body
- a student is routinely working with other students
- the assessment of a student's achievement takes place at the location of the degree-awarding body.

Where learning and teaching is delivered through a collaborative arrangement, the responsibilities of each of the partners is defined in a written agreement. Ultimate responsibility for the academic standards of the award and the quality of learning opportunities made available rests with the degree-awarding body. Sound practice in the management of such an arrangement is articulated in Chapter *B10: Management of collaborative arrangements*.[6]

More detailed information about research degrees is provided in *Chapter B11: Research degrees.*[7]

What contributes to effective learning and teaching?

Identifying effective learning and teaching is complex, as it is influenced by many factors, which vary between individuals and different learning environments. Individuals learn when they acquire new (or modify existing) knowledge, behaviours, skills or values. There are many different ways to teach, all aimed at helping a student to learn. There are many ways to empower learners with the confidence to participate, critically and creatively, in the study of their subject area(s).

The central themes of this Chapter are:

- inclusive learning through promoting equality, diversity and equal opportunity
- higher education providers, staff (all those who facilitate learning), students and other stakeholders working in partnership
- learning facilitated by enthusiastic and capable staff through teaching and other types of support for learning, whether formal or informal.

[6] www.qaa.ac.uk/Publications/InformationAndGuidance/Pages/quality-code-B10.aspx

[7] www.qaa.ac.uk/Publications/InformationAndGuidance/Pages/quality-code-B11.aspx

3

Equality, diversity and equal opportunity

Diversity of staff and students brings with it different ideas, knowledge and experiences that contribute to an enriched learning environment. Promoting equality involves treating everyone with equal dignity and worth, irrespective of the group or groups to which they belong, while also raising aspirations and supporting achievement for people with diverse requirements, entitlements and backgrounds. An inclusive environment for learning anticipates the varied requirements of learners, for example because of a declared disability, specific cultural background, location, or age, and aims to ensure that all students have equal access to educational opportunities. Higher education providers, staff and students all have a role in and responsibility for promoting equality.

Equality of opportunity involves enabling access for people who have differing individual requirements as well as eliminating arbitrary and unnecessary barriers to learning. The nature of students' particular learning experiences may vary according to location of study, mode of study, or academic subject, as well as whether they have any protected characteristics, but every student experiences parity in the quality of learning opportunities. In addition, disabled students and non-disabled students are offered learning opportunities that are equally accessible to them, by means of inclusive design wherever possible and by means of reasonable individual adjustments wherever necessary.

Offering an equal opportunity to learn is distinguished from offering an equal chance of success.

Working in partnership

In general, effective learning in higher education occurs when there is a partnership between the higher education provider, their staff and students, and any external stakeholder that results in:

- teaching that aims to be inspirational
- the co-production and enhancement of creative and transformational learning, and
- a schedule of assessment that is effective both in supporting learning and safeguarding academic standards.

Transformational learning involves a process of becoming critically aware of one's own tacit assumptions and expectations and those of others, and assessing their relevance before making an interpretation.

Partnership in this Chapter uses the definition set out in *Chapter B5: Student engagement*,[8] emphasising that partnership is about joint working and not a formal legal relationship based on equal responsibility and liability. Joint working is based on a mature relationship and mutual respect. In the case of learning and teaching, a partnership between staff and students can empower students to develop further as active and independent learners who recognise and take responsibility for their own learning. Such a partnership, whether through formal or informal arrangements, acknowledges that while providers are responsible for providing inclusive learning opportunities and support for learning, the effectiveness with which the learning opportunities are used is a matter for students themselves.

This Chapter address the concept of students engaging with learning opportunities

[8] www.qaa.ac.uk/Publications/InformationAndGuidance/Pages/quality-code-B5.aspx

4

and teaching activities. *Chapter B5: Student engagement* considers how higher education providers engage with students through their quality systems.

Teaching and support for learning

This Chapter is underpinned by recognition that learning and teaching take place in a variety of forms, involving a mix of formal and informal arrangements, and that what is commonly described as 'support for learning' plays a key role. For that reason the term 'teacher' is used to refer to any member of staff involved in facilitating student learning. Teaching is anything a teacher does to facilitate learning.

The term 'staff' refers to anyone involved in teaching or supporting student learning. It includes, but is not limited to, academic staff, graduate teaching assistants, specialist learning support staff, library staff and technicians employed by the higher education provider. It also includes staff not employed by the higher education provider but who interact with students studying for one of their awards; for example, through a collaborative arrangement or through supporting placement learning.

Effective learning and teaching activities and practices are enabled through, and depend on, staff who are appropriately qualified for their role and who engage throughout their career in continuing professional development, in the evaluation of their practice, and in developing their understanding of their subject and the learning process as it relates to their subject.

Student support and the provision of learning resources underpin effective learning and teaching. These topics are covered in *Chapter B4: Student support, learning resources and careers education, information advice and guidance.*[9]

The role of assessment in learning and teaching

This Chapter addresses the role of assessment in promoting learning, especially as the basis for reflection and dialogue between staff and students. This activity is often referred to as assessment **for** learning and equated with formative assessment, being developmental. It is contrasted with assessment **of** learning - often equated with summative assessment - the principal purpose of which is to determine student attainment against predetermined criteria. An effective approach to summative assessment, however, reflects the capacity of any form of assessment to enable ongoing learning. Assessment of learning is addressed in *Chapter B6: Assessment of students and accreditation of prior learning.*[10]

To derive maximum learning benefit from assessment, students need to receive timely feedback in a manner that is supportive, yet includes some personal challenge. Effective feedback is the result of: agreeing and communicating clear criteria before students complete the task; assessment that uses those criteria; and feedback based on the criteria and timed so that students can use it constructively in their next stage of learning. Effective feedback for learning also takes into account the need to build confidence as well as to communicate where and how improvements can be made. Effective feedback to support learning involves an ongoing dialogue between the student and staff, the student and their peers, and the student and an employer, for example, where the student is in a work-based or placement learning environment.

[9] www.qaa.ac.uk/Publications/InformationAndGuidance/Pages/quality-code-B4.aspx

[10] www.qaa.ac.uk/Publications/InformationAndGuidance/Pages/quality-code-B6.aspx

5

Expectation

The Quality Code sets out the following Expectation about learning and teaching, which higher education providers are required to meet.

Higher education providers, working with their staff, students and other stakeholders, articulate and systematically review and enhance the provision of learning opportunities and teaching practices, so that every student is enabled to develop as an independent learner, study their chosen subject(s) in depth and enhance their capacity for analytical, critical and creative thinking.

6

Indicators of sound practice

The Indicators below are mutually dependent, so they should not be considered in isolation. For example, strategy is not divorced from learning design, and learning design is informed by the careful evaluation of the effectiveness of actual practices through the collection and analysis of robust evidence. Furthermore, it may become evident from the evaluation of learning and teaching that the strategy itself requires modification.

The basis for effective learning and teaching

Indicator 1

Higher education providers articulate and implement a strategic approach to learning and teaching and promote a shared understanding of this approach among their staff, students and other stakeholders.

One strength of UK higher education is the diversity of its providers in terms of mission and range of provision. Each higher education provider sets out, in a form it deems appropriate, and implements its strategic approach to learning and teaching, indicating how it promotes student learning across all of its provision, across all academic levels and all organisational levels (for example, provider, faculty and department).

This strategic approach addresses the learning opportunities for every student, irrespective of: location of study (for example, whether on a campus or in a work-based or placement environment, in the UK or outside the UK) or mode of study (for example, part-time or full-time, or distance learning supported by technology). It considers how every student can experience parity in the quality of learning opportunities (see *Chapter B10: Management of collaborative arrangements*).[11]

In agreeing their strategic approach, higher education providers recognise that effective learning occurs when students are enabled to:

- engage actively in learning, and participate fully in the learning opportunities that are presented to them
- understand their learning environment, its culture and resources
- embrace the aims and expectations of their chosen programme of study
- demonstrate understanding of, and an ability to reflect upon, the ways in which their skills and knowledge are developing
- recognise and value their existing knowledge and skills, and build upon them
- make effective and responsible use of advice, guidance and feedback from formative and summative assessment that is provided during their programme of study
- avail themselves of the opportunities to acquire and develop learning skills
- understand sound academic practice and behave with integrity
- contextualise and apply their developing knowledge and skills to their wider experience and plans for the future

[11] www.qaa.ac.uk/Publications/InformationAndGuidance/Pages/quality-code-B10.aspx

7

- work together in an informal environment as well as in formal learning sessions
- act responsibly in planning their use of the resources available to support learning, including the effective use of their own time in independent study
- give constructive feedback on their perception of the quality of their learning experience
- engage in monitoring and influencing the curriculum
- show commitment to attaining the academic standards that have been defined for their programme of study.

Students are uniquely positioned to comment on how the strategic approach to learning and teaching adopted by the higher education provider enables and supports student learning, and how enhancement of the opportunities provided can benefit their learning. Consequently, higher education providers involve students in developing, implementing and monitoring the strategic approach (see *Chapter B5: Student engagement*).[12]

Themes that cross subject boundaries

In addition to subject-specific content, higher education providers consider the way their strategic approach reflects themes that cross subject boundaries. These themes reflect topics which may be considered to have a broad relevance to the purposes of higher education and its wider context in society. Where the themes are embedded within the curriculum and form an integral part of a programme of study, learning and teaching activities are designed to take them into account.

These themes may include:

- academic and digital literacies appropriate to the academic level of the student
- education for sustainability
- citizenship
- enterprise and entrepreneurship
- internationalisation
- ethical behaviour.

The design and approval of programmes is addressed in *Chapter B1: Programme design and approval.*[13]

[12] www.qaa.ac.uk/Publications/InformationAndGuidance/Pages/quality-code-B5.aspx

[13] www.qaa.ac.uk/Publications/InformationAndGuidance/Pages/quality-code-B1.aspx

8

Higher education providers are responsible for ascertaining which laws and regulations apply to them. To meet the Expectation of this Chapter of the Quality Code, higher education providers may wish to consider the indicative list of reference points, guidance and examples of good practice below.

NUS (2012) *Student Experience Research Part 1: Teaching and Learning*
www.qaa.ac.uk/Publications/InformationAndGuidance/Pages/Student-Experience-Research-12-Part-1.aspx

NUS (2012) *Student Experience Research Part 3: Subject Differences*
www.qaa.ac.uk/Publications/InformationAndGuidance/Pages/Student-Experience-Research-12-Part-3.aspx

QAA (2011) *Outcomes from Institutional Audit: 2007-09: Managing learning opportunities*
www.qaa.ac.uk/Publications/InformationAndGuidance/Pages/outcomes-audit-learning-opps.aspx

QAA Scotland Enhancement Themes: Graduates for the 21st Century
www.enhancementthemes.ac.uk/enhancement-themes/completed-enhancement-themes/graduates-for-the-21st-century

Higher Education Academy: Education for sustainable development
www.heacademy.ac.uk/education-for-sustainable-development

QAA (2012) *Enterprise and entrepreneurship education: Guidance for UK higher education providers*
www.qaa.ac.uk/Publications/InformationAndGuidance/Pages/enterprise-entrepreneurship-guidance.aspx

QAA (2008) *Outcomes from Institutional Audit, Second series: Work-based and placement learning, and employability*
www.qaa.ac.uk/Publications/InformationAndGuidance/Pages/Outcomes-institutional-audit-Second-series-Work-based-placement-learning-employability.aspx

Indicator 2

Learning and teaching activities and associated resources provide every student with an equal and effective opportunity to achieve the intended learning outcomes.

Learning and teaching activities, support for learning, and resources together enable every student who engages appropriately to achieve the learning outcomes. The learning opportunities support students to make the transition into higher education and to progress academically and intellectually through their programme of study, whatever its level.

Learning outcomes

Opportunities for learning in higher education may be formal or informal. The term 'learning outcomes' recognises that learning will generate a range of outcomes, including the intended learning defined in the programme of study and learning which is generated by or as experience or which is the by-product of other learning activities.

9

The learning experience may therefore involve informal learning over and above the structured formal learning focused on the defined learning outcomes.

The concept of learning outcomes is applicable at all academic levels (*Chapter A1: The national level*)[14] and to every subject of study (*Chapter A2: The subject and qualification level*).[15] For postgraduate research students it links to Indicators 13 and 14 of *Chapter B11: Research degrees*.[16]

Once determined, the learning outcomes for the programme of study map directly to the summative assessment, with the assessment methods being appropriate to offer every student an equal opportunity to demonstrate their achievement of the intended learning outcomes irrespective of how and where the student has studied (see *Chapter B6: Assessment of students* of the Quality Code).[17]

Providing an effective opportunity to achieve learning outcomes

The planning and design of learning and teaching activities and associated resources:

- provides breadth, depth, pace and challenge appropriate for the learning outcomes, subject and level of study
- takes an inclusive approach
- develops appropriate knowledge, skills and understanding.

Achievement of learning outcomes may depend on fieldwork, placement or work-based learning and may need to satisfy the requirements of a professional, statutory and regulatory body or other external framework. In these situations the learning outcomes encapsulate the academic and professional competence standards essential to the programme of study and, at the same time, are achievable by every student undertaking that programme of study. Where optional or compulsory fieldwork, practical work, or work placement is part of a programme, prior consideration of reasonable adjustments enables the participation of disabled students. Higher education providers engage students and external partners providing fieldwork, placement or work-based learning opportunities in their internal quality processes to assure themselves that the learning outcomes are relevant and achievable (see *Chapter B5: Student engagement*[18] and *Chapter B8: Programme monitoring and review*[19]).

Providing an equal opportunity to achieve learning outcomes

The nature of students' particular learning experiences may vary according to location of study, mode of study, or academic subject, as well as whether they have any protected characteristics, but every student experiences parity in the quality of learning opportunities.

[14] www.qaa.ac.uk/Publications/InformationAndGuidance/Pages/quality-code-A1.aspx

[15] www.qaa.ac.uk/Publications/InformationAndGuidance/Pages/quality-code-A2.aspx

[16] www.qaa.ac.uk/Publications/InformationAndGuidance/Pages/quality-code-B11.aspx

[17] www.qaa.ac.uk/Publications/InformationAndGuidance/Pages/quality-code-B6.aspx

[18] www.qaa.ac.uk/Publications/InformationAndGuidance/Pages/quality-code-B5.aspx

[19] www.qaa.ac.uk/Publications/InformationAndGuidance/Pages/quality-code-B8.aspx

All methods of learning and teaching including group work, seminars, lectures, placements and practical classes create a variety of challenges for students. Providing each student with an equal opportunity involves enabling access for people who have differing individual requirements as well as eliminating arbitrary and unnecessary barriers to learning. A flexible and inclusive approach to learning and teaching enables and empowers every student to fulfil their individual potential and minimises the need for individual alterations or adjustments. Through adopting inclusive learning and teaching practices, higher education providers enhance the learning opportunities for every student.

Diversity of staff and students brings with it different ideas, knowledge and experiences that contribute to an enriched learning environment. Higher education providers recognise that students have differing learning styles and come from diverse educational, linguistic and cultural backgrounds, and they consider whether examples and resources used in learning and teaching are drawn from a sufficiently broad range of sources, cultures and viewpoints. Higher education providers, their staff, and staff of external partners involved in facilitating student learning create a positive environment by aiming to eliminate harassment or discriminatory behaviour from among the peer group of students.

Higher education providers have a specific responsibility under UK equality legislation to ensure that disabled people are not put at a disadvantage. Providers offer disabled students learning opportunities which are accessible to them, by means of inclusive design wherever possible and by means of reasonable individual adjustments wherever necessary. Staff work in partnership with individual students to understand the implication of any specific requirements or reasonable adjustments for their teaching practices.

Where a programme includes learning that takes place in a virtual learning environment, providers ensure that it can be accessed fully by every student, or that reasonable adjustments can be made so that every student has an equal learning opportunity.

Higher education providers are responsible for ascertaining which laws and regulations apply to them. To meet the Expectation of this Chapter of the Quality Code, higher education providers may wish to consider the indicative list of reference points, guidance and examples of good practice below.

Equality Challenge Unit (2010) *Disability legislation: practical guidance for academic staff (revised)* www.ecu.ac.uk/publications/disability-legislation-practical-guidance-for-academic-staff-revised

Equality and Human Rights Commission: Guidance for providers of further and higher education
www.equalityhumanrights.com/advice-and-guidance/further-and-higher-education-providers-guidance

Higher Education Academy (2010) *Inclusive Learning and Teaching in Higher Education*
www.heacademy.ac.uk/resources/detail/inclusion/LTsummit_final_report

Higher Education Academy and Equality Challenge Unit: Ethnicity, Gender and Degree Attainment
www.heacademy.ac.uk/resources/detail/inclusion/Ethnicity/ethnicity

11

Higher Education Academy and UK Council for International Student Affairs: Teaching International Students Project
www.heacademy.ac.uk/teaching-international-students

JISC TechDis: Teaching Inclusively Using Technology
www.jisctechdis.ac.uk/pages/detail/online_resources/Teaching_Inclusively_Using_Technology

ASET (2009) *A Good Practice Guide for Placement and Other Work-Based Learning Opportunities in Higher Education* (2nd edition)
www.asetonline.org/documents/ASETCodeofPractice-Version2.1.pdf

Indicator 3

Learning and teaching practices are informed by reflection, evaluation of professional practice, and subject-specific and educational scholarship.

Effective teaching and support for learning occurs when staff display a sound understanding and up-to-date knowledge of their subject and/or professional practice and they bring this to a variety of appropriately designed learning and teaching activities and assessment methods. They communicate enthusiasm, and draw on scholarship, research and professional activity to facilitate student learning. Staff create opportunities for learning which are effective by recognising the value of both individual and collaborative learning activities, the value of learning how to learn, and that learning is about interpretation, analysis and synthesis underpinned by reflection, not just the repetition of facts.

Reflective practice

In order that teaching and support for learning remains effective, staff reflect on their practice and consider how it might be changed and improved. They draw on a variety of sources to inform this, including feedback from: students gathered at different organisational levels from the individual teaching session and module upwards; examination boards and external examiners; and recent alumni and employers.

Effective reflective practice retains a practical focus. It may involve different organisational levels, from the individual teacher or programme teaching team, to a whole department, faculty or higher education provider. Successful reflective practice depends on self-awareness, critical analysis, synthesis and evaluation leading to new perspectives.

Scholarship and professional practice

Scholarship and research lie at the heart of higher education, but their nature will depend on the academic level of the programme, the subject area and the provider or providers of the programme. Scholarship may include conventional research (discovery of new knowledge), innovative application or integration of existing knowledge, for example in professional practice, or the study of learning and teaching processes and practices.

In subject areas such as clinical medicine, music or performing arts among others, or where learning is taking place in the workplace, evaluation of professional practice directly informs student learning. The requirements of professional, regulatory or statutory bodies are also taken into account where applicable.

Higher education providers are responsible for ascertaining which laws and regulations apply to them. To meet the Expectation of this Chapter of the Quality Code, higher education providers may wish to consider the indicative list of reference points, guidance and examples of good practice below.

Higher Education Academy: Resource centre
www.heacademy.ac.uk/resources

QAA Scotland Enhancement Themes: Research-Teaching Linkages
www.enhancementthemes.ac.uk/enhancement-themes/completed-enhancement-themes/research-teaching-linkages

Higher Education Academy (2006) *Scholarly Activity in the context of HE in FE*
www.heacademy.ac.uk/ourwork/universitiesandcolleges/heinfe/alldisplay?type=resources&newid=resource_database/web0462_scholarly_activity_in_the_context_of_he_in_fe_June_2006&site=york

Higher Education Academy (2007) *Linking teaching and research in disciplines and departments*
www.heacademy.ac.uk/assets/documents/teachingandresearch/LinkingTeachingAndResearch_April07.pdf

Higher Education Academy: Becoming a reflective practitioner
www.heacademy.ac.uk/resources/detail/subjects/medev/Focus-_Becoming_a_reflective_practitioner

Indicator 4

Higher education providers assure themselves that everyone involved in teaching or supporting student learning is appropriately qualified, supported and developed.

Effective student learning is facilitated by interaction with appropriately qualified, supported and developed teaching and support staff. Staff may be employed by the degree-awarding body, by a collaborative partner, a work-based learning or placement provider, or may be a member of visiting staff; where responsibility sits for staff appointment, support and development is defined by the terms of the relevant collaborative arrangement (see *Chapter B10: Management of collaborative arrangements*).[20]

[20] www.qaa.ac.uk/Publications/InformationAndGuidance/Pages/quality-code-B10.aspx

13

Appointment, support, and continuing development of staff

Higher education providers determine what is necessary to demonstrate that a member of staff is qualified to fulfil their role in teaching or supporting learning; whether this means the individual holds a relevant formal qualification will depend on the circumstances. Staff recruitment and appointment procedures include a means of ensuring new staff have an appropriate level of competence for the teaching and/or learning support role to which they are being appointed. In particular the provider considers the extent to which members of staff have:

- appropriate and current practitioner knowledge and an understanding of the subject they teach (which may be demonstrated by a qualification) and an understanding of the disciplinary scholarship appropriate to the academic level of the students they are teaching
- the necessary skills and experience to facilitate learning in the students they are interacting with, and to use approaches grounded in sound learning and teaching scholarship and practice.

Members of staff new to their teaching or supporting student learning role are encouraged to engage in appropriate induction and mentoring opportunities made available by the higher education provider.

Once appointed, and throughout their career, staff engage with opportunities to develop and extend their teaching capabilities and to reflect upon their teaching practice. Staff are encouraged to value their own and others' skills, to recognise that they have a responsibility to identify their own development needs, and to engage in initial and continuing professional development activities. Higher education providers make opportunities available for all those involved in teaching and supporting student learning to inform each other's practice and professional development.

Continuing professional development activities made available by the higher education provider are planned strategically, including the allocation of sufficient resources to cover the needs of both research and learning and teaching development. Protected staff time to engage in continuing professional development is identified and factored into workload considerations.

Higher education providers assure themselves of the effectiveness of their approach to staff development and support. Aspects considered may include any or all of the following: working with staff development teams; having online continuing professional development resources and modules for staff; and ensuring the availability of sufficient administrative support.

Higher education providers also have agreed procedures to identify staff in need of additional support to ensure their effectiveness, and provide them with opportunities (which the provider expects them to take up), support and mentoring to enable improvement of their skills and competency to an agreed level.

Individual staff members are able to access appropriate and timely support to develop inclusive forms of learning, teaching and assessment which are supported by technology. They are provided with guidance and support to understand the impact of equal opportunity and equality legislation on their roles, and have free

access to information on creating an inclusive learning culture for which they have a shared responsibility with the higher education provider. Higher education providers also recognise the importance of digital literacy for staff and make available suitable development opportunities.

Higher education providers have in place transparent staff recognition and reward processes, and promotion opportunities for all career paths.

The UK Professional Standards Framework for teaching and supporting learning in higher education

The UK higher education sector has endorsed the UK Professional Standards Framework for teaching and supporting learning in higher education (UKPSF), which is published by the Higher Education Academy (HEA) on behalf of the four UK funding bodies as well as Universities UK, GuildHE, and NUS. The purpose of the framework is to help individuals and higher education providers enhance the learning experience of their students, by improving the quality of their teaching and learning support. It is written from the perspective of the practitioner and outlines the national framework for recognising and benchmarking teaching and learning support roles within higher education. The UKPSF has two components. The descriptors are a set of statements outlining the key characteristics of someone performing four broad categories of typical teaching and learning support roles within higher education. The dimensions of practice are a set of statements outlining: the **areas of activity** undertaken by teachers and supporters of learning within higher education; the **core knowledge** needed to carry out those activities at the appropriate level; and the **professional values** that someone performing these activities embraces and exemplifies.

The UKPSF provides a UK-wide benchmark by which higher education providers can demonstrate how they support staff and assure themselves that they are qualified to teach and support learning. It also enables higher education providers to demonstrate that their professional development programmes and activities meet expected national professional standards.

Higher education providers are responsible for ascertaining which laws and regulations apply to them. To meet the Expectation of this Chapter of the Quality Code, higher education providers may wish to consider the indicative list of reference points, guidance and examples of good practice below.

UK Professional Standards Framework
www.heacademy.ac.uk/ukpsf

Staff and Educational Development Association (SEDA)
www.seda.ac.uk

Association for Learning Development in Higher Education (ALDinHE)
www.aldinhe.ac.uk

JISC: Learning Literacies in a Digital Age
www.jisc.ac.uk/whatwedo/projects/elearningllida.aspx

JISC TechDis
www.jisctechdis.ac.uk

15

ASET (The Placement and Employability Professionals' Body)
www.asetonline.org

ENQA (2009) *Standards and Guidelines for Quality Assurance within the European Higher Education Area* (3rd edition)
www.enqa.eu/pubs_esg.lasso

Indicator 5

Higher education providers collect and analyse appropriate information to ensure the continued effectiveness of their strategic approach to, and the enhancement of, learning opportunities and teaching practices.

Higher education providers use a range of internal and external information and feedback from diverse sources along with examples of sound practice and innovation to enable them to keep their strategic approach to learning and teaching under review, to modify it as appropriate and to facilitate the continuous improvement of the learning opportunities they provide.

Data sources on which they draw may include:

- feedback from students on their learning experience collected through internal mechanisms
- feedback from students through external instruments like the National Student Survey (NSS), the Postgraduate Taught Experience Survey (PTES) and the Postgraduate Research Experience Survey (PRES)
- routine evaluations of modules and programmes incorporating feedback from staff and external examiners
- feedback from alumni and employers and placement providers
- retention statistics
- mark profiles for students, modules and programmes
- availability and quality of teaching and learning spaces for formal and informal learning
- uptake and utilisation of any virtual learning environment and assistive technology
- student academic appeals and complaints
- feedback from external reviews and accreditations, such as those of professional, regulatory and statutory bodies.

Evaluation takes place at different levels from the module, by the individual teacher or module team, through to senior management level and is appropriate to the mode and level of the provision.

The use of information for assurance and enhancement at programme and award level is addressed in *Chapter B8: Programme monitoring and review*.[21]

The involvement of students in processes to assure and enhance providers' approaches is covered in *Chapter B5: Student engagement*.[22]

Higher education providers are responsible for ascertaining which laws and regulations apply to them. To meet the Expectation of this Chapter of the Quality Code, higher education providers may wish to consider the indicative list of reference points, guidance and examples of good practice below.

EFQM Excellence Model Higher Education Version (2003)
www.osti.gov/eprints/topicpages/documents/record/884/1265593.html

HEFCE (2010) *Enhancing and Developing the National Student Survey*
www.hefce.ac.uk/media/hefce/content/pubs/2010/rd1210/rd12_10a.pdf

Higher Education Academy (2012) *Using PRES to enhance the experience of postgraduate researchers*
www.heacademy.ac.uk/resources/detail/postgraduate/using_pres_to_enhance

[21] www.qaa.ac.uk/Publications/InformationAndGuidance/Pages/quality-code-B8.aspx

[22] www.qaa.ac.uk/Publications/InformationAndGuidance/Pages/quality-code-B5.aspx

17

The learning environment

Indicator 6

Higher education providers maintain physical, virtual and social learning environments that are safe, accessible and reliable for every student, promoting dignity, courtesy and respect in their use.

The learning environment

The learning and teaching activities made available by a higher education provider take place within a broad learning environment which comprises both physical and virtual facilities and the culture and ethos of learning promoted by the provider and its staff working with students and other stakeholders. This learning environment also encompasses learning opportunities provided, for example, through fieldwork or workplace or placement settings.

Physical learning environments

Higher education providers make available physical environments that are applicable and appropriate to each mode and location of learning, and which are safe and accessible to students. Learning spaces are suitable for the nature of the learning activities being provided, for example in seating arrangements, lighting and acoustics, and availability of technology for planned activities. Where appropriate, learning spaces enable informal as well as formal learning.

Higher education providers decide how to access specialist advice on accessible and inclusive learning space design that best meets their requirements and suits their context. Providers involve students in the process for developing a fully accessible environment, especially students with specific requirements, such as disabled or part-time students. Higher education providers have a systematic approach, for example to the refurbishment of learning spaces, rather than relying on making adjustments on an ad hoc basis.

Virtual learning environments

The use of technology to enhance learning can be an important means to enable students to engage fully in their programme of study. Higher education providers develop technological facilities and services (including virtual learning environments and library systems) that are accessible, inclusive and cater for a wide range of potential student requirements. Systematic consultation with students and staff about the accessibility of technological facilities enhances standards of usability.

Assistive technology can make methods of learning and teaching more accessible to a wide range of students. Wherever possible, assistive technologies are made available to all students through integrated organisation-wide systems, rather than through distributed facilities or those targeted at a subset of students.

Social learning environments

All interactions among students and staff (including staff who contribute to learning through fieldwork, placements and work-based learning), whether in person or through electronic means, reflect the following characteristics:

- dialogue based on mutual dignity and respect
- a safe environment for exploring new ideas and for providing feedback even when that is negative.

Higher education providers are responsible for ascertaining which laws and regulations apply to them. To meet the Expectation of this Chapter of the Quality Code, higher education providers may wish to consider the indicative list of reference points, guidance and examples of good practice below.

Equality Challenge Unit (2010) *Disability legislation: practical guidance for academic staff* (revised)
www.ecu.ac.uk/publications/disability-legislation-practical-guidance-for-academic-staff-revised/

Equality and Human Rights Commission: Guidance for providers of further and higher education
www.equalityhumanrights.com/advice-and-guidance/further-and-higher-education-providers-guidance

QAA Scotland Enhancement Themes: Flexible Learning
www.enhancementthemes.ac.uk/enhancement-themes/completed-enhancement-themes/flexible-delivery

Higher Education Academy: Flexible learning
www.heacademy.ac.uk/flexible-learning

JISC (Joint Information Systems Committee)
www.jisc.ac.uk

JISC TechDis
www.jisctechdis.ac.uk

Universities and Colleges Employers Association (2009) *Health and safety guidance for the placement of HE students*
www.ucea.ac.uk/en/publications/index.cfm/HSplace

Association of Learning Technologists (ALT): What research has to say for practice
wiki.alt.ac.uk/index.php/What_research_has_to_say_for_practice

Student engagement in learning

Indicator 7

Every student is provided with clear and current information that specifies the learning opportunities and support available to them.

Higher education providers make clear to students the learning opportunities and learning and teaching support available to them and how they can access these.

19

Care is taken to ensure coherence in information that relates to different organisational levels of the provider or elements of study (for example department, programme or module) and which is given at different times.

The information provided to every student reflects the specific nature of the learning opportunities available and learning support provided, including any study undertaken at a location that is not at the campus of the provider (for example in a fieldwork, work-based or placement environment) or through a virtual learning environment. It also indicates whether the learning opportunities and support are made available by the provider or by an external stakeholder such as a collaborative partner or employer.

Higher education providers actively engage students to monitor, review and evaluate this information, to ensure that it meets the requirements of different groups of students. The involvement of students in quality assurance and enhancement is addressed in *Chapter B5: Student engagement*.[23]

The provision of information for prospective and current students is addressed in Part C: Information about higher education provision.[24]

Higher education providers are responsible for ascertaining which laws and regulations apply to them. To meet the Expectation of this Chapter of the Quality Code, higher education providers may wish to consider the indicative list of reference points, guidance and examples of good practice below.

Teachability project: Creating accessible information about courses or programmes of study for disabled students
www.teachability.strath.ac.uk/chapter_1/tableofcontents1.html

Universities UK (2002) *Student Services: Effective approaches to retaining students in higher education*
www.universitiesuk.ac.uk/Publications/Documents/services.pdf

Higher Education Academy (2012) *Building student engagement and belonging in Higher Education at a time of change: final report from the What Works? Student Retention and Success programme*
www.heacademy.ac.uk/assets/documents/retention/What_works_final_report.pdf

NUS (2012) *Student Experience Research Part 2: Independent Learning and Contact Hours*
www.qaa.ac.uk/Publications/InformationAndGuidance/Pages/Student-Experience-Research-12-Part-2.aspx

QAA (2011) *Contact hours: a guide for students*
www.qaa.ac.uk/Publications/InformationAndGuidance/Pages/contact-hours-student.aspx

[23] www.qaa.ac.uk/Publications/InformationAndGuidance/Pages/Quality-Code-Part-B5.aspx

[24] www.qaa.ac.uk/Publications/InformationAndGuidance/Pages/Quality-Code-Part-C.aspx

Indicator 8

Higher education providers take deliberate steps to assist every student to understand their responsibility to engage with the learning opportunities provided and to shape their learning experience.

A key characteristic of UK higher education is the emphasis placed on students to engage in independent learning, working in partnership with staff and displaying academic behaviour and integrity appropriate to the level of study. Consequently, providers explain what they expect of students and how they will assist them to make the transition into and through their studies.

Students undertaking fieldwork, work-based or placement learning may have additional responsibilities to the learning provider and to others such as customers, clients, service users, other employees and the general public they may encounter. This includes the responsibility to meet the norms and expectations for professional conduct in the particular field of work or study that they are undertaking. For students using their existing workplace for their work-based learning, such norms, expectations and responsibilities are often covered in an employment contract and may be more obvious than for students joining a workplace to undertake a placement.

As active members of a learning community, students depend on interaction with staff and with their peers to support their learning. Achieving independence in learning means that there are always some opportunities for students to shape their learning experience. For some students this may not extend beyond selecting optional modules, undertaking additional reading or practice of relevant skills. For others it may extend to the negotiation of assessment titles or engagement in self-selected research for a dissertation or equivalent practice-based module.

Higher education providers are responsible for ascertaining which laws and regulations apply to them. To meet the Expectation of this Chapter of the Quality Code, higher education providers may wish to consider the indicative list of reference points, guidance and examples of good practice below.

NUS (2012) Student Experience Research Part 4: First Year Student Experience
www.qaa.ac.uk/Publications/InformationAndGuidance/Pages/Student-Experience-Research-12-Part-4.aspx

QAA Scotland Enhancement Themes: First Year: Engagement and Empowerment
www.enhancementthemes.ac.uk/enhancement-themes/completed-enhancement-themes/first-year

Higher Education Academy: Academic integrity service
www.heacademy.ac.uk/academic-integrity

Indicator 9

Every student is enabled to monitor their progress and further their academic development through the provision of regular opportunities to reflect on feedback and engage in dialogue with staff.

Assessment for learning

The design and delivery of learning and teaching activities, including assessment, provide opportunities for feedback that promotes learning. The activities are informed by an understanding that, while the purpose of formative assessment is to facilitate learning, summative assessment also has a formative element which may relate to the development of transferable knowledge and skills beyond the specific subject matter of the assessment task itself. The engagement of students in learning is stimulated by their understanding of the value of feedback obtained through those activities provided by staff and peers, and the opportunity to reflect on that feedback to inform further development. Effective dialogue with staff builds on that feedback and reflection.

The assessment schedule of a programme, module or session considers the dual needs of assessment for learning and assessment of learning. It facilitates feedback, reflection and dialogue, taking into account the value of students having time to put their learning into practice, including in the next applicable summative assessment.

Reflection and personal development planning

Students are encouraged to reflect on the formal and informal feedback they receive and use it to engage in a dialogue with staff to help plan their future learning. Many higher education providers factor into the student learning opportunities offered a process based on personal development planning (PDP). PDP is a structured and supported process undertaken by a learner to reflect upon their own learning, performance and/or achievement and to plan for their personal, educational and career development. It is an inclusive process, open to all learners, in all higher education provision settings, and at all levels.

Higher education providers are responsible for ascertaining which laws and regulations apply to them. To meet the Expectation of this Chapter of the Quality Code, higher education providers may wish to consider the indicative list of reference points, guidance and examples of good practice below.

QAA (2012) *Understanding assessment: its role in safeguarding academic standards and quality in higher education (2nd edition)*
www.qaa.ac.uk/Publications/InformationAndGuidance/Documents/understanding-assessment-second-edition.pdf

QAA (2009) *Personal development planning: guidelines for institutional policy and practice in higher education*
www.qaa.ac.uk/Publications/InformationAndGuidance/Pages/Personal-development-planning-guidance-for-institutional-policy-and-practice-in-higher-education.aspx

Centre for Recording Achievement (CRA)
www.recordingachievement.org

QAA Scotland Enhancement Themes: Integrative Assessment
www.enhancementthemes.ac.uk/enhancement-themes/completed-enhancement-themes/integrative-assessment

QAA Scotland Enhancement Themes: Assessment
www.enhancementthemes.ac.uk/enhancement-themes/completed-enhancement-themes/assessment

QAA Scotland Personal development planning toolkit
www.qaa.ac.uk/scotland/DevelopmentAndEnhancement/Pages/Personal-development-planning-Scotland.aspx

Higher Education Academy: Assessment and feedback
www.heacademy.ac.uk/assessment

NUS: Ten Principles of good feedback
www.nus.org.uk/en/advice/course-reps/feedback/feedback-what-you-can-expect-/

23

Appendix 1: The Expectation and Indicators

The Expectation

The Quality Code sets out the following Expectation about learning and teaching, which higher education providers are required to meet:

> **Higher education providers, working with their staff, students and other stakeholders, articulate and systematically review and enhance the provision of learning opportunities and teaching practices, so that every student is enabled to develop as an independent learner, study their chosen subject(s) in depth and enhance their capacity for analytical, critical and creative thinking.**

The Indicators of sound practice

Indicator 1

Higher education providers articulate and implement a strategic approach to learning and teaching and promote a shared understanding of this approach among their staff, students and other stakeholders.

Indicator 2

Learning and teaching activities and associated resources provide every student with an equal and effective opportunity to achieve the intended learning outcomes.

Indicator 3

Learning and teaching practices are informed by reflection, evaluation of professional practice, and subject-specific and educational scholarship.

Indicator 4

Higher education providers assure themselves that everyone involved in teaching or supporting student learning is appropriately qualified, supported and developed.

Indicator 5

Higher education providers collect and analyse appropriate information to ensure the continued effectiveness of their strategic approach to, and the enhancement of, learning opportunities and teaching practices.

Indicator 6

Higher education providers maintain physical, virtual and social learning environments that are safe, accessible and reliable for every student, promoting dignity, courtesy and respect in their use.

Indicator 7

Every student is provided with clear and current information that specifies the learning opportunities and support available to them.

Indicator 8

Higher education providers take deliberate steps to assist every student to understand their responsibility to engage with the learning opportunities provided and shape their learning experience.

Indicator 9

Every student is enabled to monitor their progress and further their academic development through the provision of regular opportunities to reflect on feedback and engage in dialogue with staff.

Appendix 2: Membership of the Advisory Group for this Chapter

Name	Position	Affiliation
Dr Mark Atlay	Director of Teaching and Learning	University of Bedfordshire
Harriet Barnes	Development Officer	QAA
Dr Tim Burton	Assistant Director	QAA (Chair)
Professor Alan Davidson	Dean for the Enhancement of Learning, Teaching and Assessment	Robert Gordon University
Dr Ian Giles	Emeritus Fellow, formerly Director of Learning and Teaching Enhancement Unit	University of Southampton (specialist writer)
Julie Hall	Director of Learning and Teaching Enhancement Unit	Roehampton University
Katya Hosking	Inclusive Curriculum Officer	Cardiff University
Professor Sharon Huttly	Dean of Studies	London School of Hygiene and Tropical Medicine
Alan M Jones	Director of Education (Architecture)	Queen's University Belfast
Matthew Kitching	Representation and Development Manager	Buckinghamshire New University
Professor Rose Luckin	Professor of Learner Centred Design	Institute of Education
Sarah MacDonald	Director of Standards and Quality Assurance	Pearson Higher Education Awards
Dorothy McElwee	Head of Higher Education and Training Programmes	North West Regional College
Dr Christine Macpherson	Assistant Director	QAA Scotland
Dr Jenny Naish	Assistant Dean (Learning, Teaching and Quality), Teesside University Business School	Teesside University
Dr Kathleen M Quinlan	Head of Educational Development (Oxford Learning Institute)	University of Oxford

Dr Simon Rouse	Principal Lecturer, Faculty of Health and Life Sciences	York St John University
Dr Jon Scott	Academic Director, College of Medicine, Biological Sciences and Psychology	University of Leicester
Dr Jo Smedley	Director, Centre for Excellence in Learning and Teaching	University of Wales, Newport
Ellen Thinnesen	Dean of Higher Education	Grimsby Institute of Further and Higher Education
Helen Thomas	Independent Formerly Head of Teacher Excellence	Higher Education Academy
Professor Mary Thorpe	Professor of Educational Technology	Open University
Kate Wicklow	Head of Quality and Student Engagement	National Union of Students

27

The Quality Assurance Agency for Higher Education
Southgate House
Southgate Street
Gloucester
GL1 1UB

Tel 01452 557000
Fax 01452 557070

Email comms@qaa.ac.uk
Web www.qaa.ac.uk

ISBN 978 1 84979 708 5

All QAA's publications are available on our website www.qaa.ac.uk

Registered charity numbers 1062746 and SC037786

QAA 494 09/12

Appendix D

21 Key Elements to Consider in Fostering Quality Teaching

21 Key Elements to Consider in Fostering Quality Teaching

- The ultimate goal of quality teaching policies is to improve the quality of the learning experiences of students and – through this – the outcomes of learning. Policies and practices to foster quality teaching should therefore be guided by this ultimate goal.

- Teaching and learning are inherently intertwined and this necessitates a holistic approach to any development initiative.

- Sustained quality teaching policies require long-term, non-linear efforts and thus call for a permanent institutional commitment from the top-leadership of the institution.

- Definitions and conceptions of quality teaching are varied across contexts and evolve over time. They require adaptability and an empirical basis to remain useful for development. Instilling a culture of change will be key in ensuring relevance and sustainability.

- Quality teaching initiatives respond to specific objectives of an institution and could therefore be irrelevant when implemented in another institution, or in another department or school within the same institution. Ensuring the alignment of differing approaches in regard to teaching and learning and their contribution to the institutional strategy are key.

- Quality teaching policies should be designed consistently at institutional, programme and individual levels. The programme levels are the pivotal place where quality teaching is likely to flourish.

- Encouraging a quality teaching culture will consist in inter-linking the various types and levels of support so that collaboration and its likely impacts on the teaching and learning are enhanced among leaders, teachers, students, staff and other stakeholders.

- Strengthening horizontal linkages and creating synergies is a particularly effective way of supporting the development of quality teaching.

- Learning experiences can be gained in many different forms of learning environments, not to be limited to auditoriums and class-rooms. Learning happens also outside the institution and also from a distance.

- The temporal dimension counts in quality teaching: what can be done at a certain point of time cannot be done later and vice-versa. There are “opportunity windows” to catch.

- The environment, students’ profiles and demands, job markets requirements, reputation and history of the institution are the prominent factors amongst others that influence a strategy of teaching improvement.

- There are no predetermined thresholds to be attained in quality teaching. The lack of quantitative indicators should not be a barrier to assess the impacts. Interpreting results of the impact of quality teaching initiatives is key.

- Orchestrating the implementation, setting the right pace of change, leaving room for experiments enable a steady improvement in the quality of teaching.

- Few quantitative standards can be prescribed and measured. Each institution is primarily responsible for the quality of its teaching and should set the bar internally. Comparative analysis within and across institutions is however likely to provide new benchmarks, as long as the method used is reliable and transparent.

- Quality teaching is a part of a global quality approach and of the institutional strategy and should not be isolated from the institutional quality culture.

- Incentives are more impactful than regulations and coercive stands. Ministerial authorities, funding bodies and quality assurance agencies should contribute to foster a climate for change. Robust and trustful partnership between actors is key.

- The size of an institution is irrelevant with respect to quality teaching. Small specialised polytechnics or large multi-disciplinary universities can equally improve quality teaching provided:
  - A teaching and learning framework is set and understood by the community,
  - Resources, time and provisions are provided consistently,
  - Leadership is a driver for change and is clearly identified at all levels,
  - Synergy of policies is sought as it serves teaching and learning improvement.

- Although money matters, the quality of teaching can start improving without a significant investment.

- Sustaining quality improvement will require prioritisation, consistent with the educational model and goals set by the institution.

- Quality teaching happens first in the classroom. Not all teachers are innovators, and few innovations can be disseminated and sustained without an efficient organisational structure.

- Higher education institutions ought to cast themselves as learning organisations in order to embrace quality teaching.

Hénard & Roseveare, 2012, p. 9-11*

*Hénard, F. & Roseveare, D. (2012). Fostering quality teaching in higher education: Policies and practices. An Institutional Management in Higher Education (IMHE) Guide for Higher Education Institutions. Retrieved from http://www.oecd.org/edu/imhe/QT%20policies%20and %20practices.pdf

Appendix E

An Environmental Scan of Teaching Evaluation in Ontario

Teaching Evaluation Practices Survey

Policies and Administrative Responsibilities Survey

Survey 1
Teaching Evaluation Practices Survey

Which of the following are used to evaluate teaching at your institution?

	Used formatively	Used summatively
Student ratings of instruction (SRI)	29%	82%
Peer observation of teaching	67%	20%
In-class surveys or other in-course classroom assessment techniques	43%	7%
Self-evaluation instruments	43%	29%
Review of video-recordings	31%	8%
Teaching dossiers	31%	50%

Are there standard processes and instruments used across your institution for the following?

	Yes	No
Student ratings of instruction (SRI)	94%	6%
Peer observation of teaching	38%	62%
In-class surveys or other in-course classroom assessment techniques	33%	67%
Self-evaluation instruments	27%	73%
Review of video-recordings	13%	87%
Teaching dossiers	56%	44%

What challenges does your institution face with the implementation of SRIs?

1. Student completion of online surveys is very low. Lot of emotion around this topic because findings are seen as a judgement of the teacher. Most recent Collective Agreement not allowed to assess the teacher just the course
2. Ensuring that students complete them.
3. Faculty member concerns that the rankings are the only measure used by Promotion and Tenure committees in their evaluation (which is not true); Faculty concerns that the instrument does not accurately assess their teaching (they fail to see it as one measure only); Faculty concerns that only students "with a beef" bother to provided feedback. These concerns have been exacerbated recently as our institution has just gone on line this year (though we allowed people to "opt out" and continue to use paper if they wished (about 25% chose to opt out).
4. Response rate and interpretation of results.
5. Nous avons un projet pilote en cours de réalisation depuis l'automne 2014. La personne responsable est monsieur Jovan Groen, conseiller pédagogique au Centre de pédagogie universitaire.

6. Low response rates to on-line version. Tenured faculty can opt for paper-based. Probationary faculty must use paper-based.
7. No mechanism to include in formal performance reviews at this time. We have just moved to a new system and things are functioning very well. In the change management process, we need to carefully manage and help to repair faculty confidence in the system - a work in progress.
8. paper is clumsy; takes up in-class time; must be coordinated through student reps
9. Completion rates online, turn-around time to receive results, effective question design, effective structures and processes to help faculty interpret the results and use them
10. response rate
11. very low student completion rates across all faculties and departments
12. The issues right now are the length of time it takes to make changes, that the negotiating units may not have the same goals regarding making changes, and that transition to online will also be time consuming.
13. Most SRI's are currently paper based. We have a small percentage that use our LMS system (with the assistance of CPI) for online SRI's, however, we anticipate a LMS upgrade in the Spring which will enable all courses who use the LMS to have online course evaluation capacity. A major challenge is that we currently do not have a standard set of questions for the SRI's across the institution.
14. The lack of standardization and the response rate are challenges.
15. No standardization of instrument
    No comparators re dept avergages, university averages,
    Information not available to students
16. Response rates are poor.

What challenges has your institution faced with the implementation of Teaching Dossiers?

1. Teaching dossiers is being loosely interpreted her to mean the provision of a teaching philosophy and a reflection on teaching practice, as provided for Tenure and Promotion and for Merit applications. The challenge is that many faculty members do not provide these elements and only provide the subset of scores required as part of our collective agreement. As such, committees have a difficult time assessing teaching competence. It seems that for many faculty members, they are unsure how to present their reflections. (we have work to do in this regard).
2. Currently there are no formal process or supports in place for faculty members to develop their teaching dossiers. The Teaching and Learning Centre is able to provide resources and basic information in this area, but support is limited.
3. Ce n'est pas une obligation. Par contre, lors de l'obtention de la permanence ou d'une promotion, cela devient un incontournable.
4. Variation in level of detail required in different program areas.
5. inconsistently used and not officially required.
6. This is not a standardized procedure at all. It is recommended for tenure and promotion, but I'm not aware that we track the adoption rate.
7. Acceptance of value; time commitment; committee and heads understanding of how to evaluate dossiers
8. Helping faculty to understand what goes in a dossier
9. we are in our first year of rolling out e-portfolios so we are making these known to faculty and working with them to have them consider using this tool.
10. The challenges are a) we have great variation in the type of teaching dossiers that are constructed; b) we have no online tool for a standardized 'template' for dossier construction.

11. expectations of contents of teaching dossier are different based on different T and P committee (faculty based); Are they really read or do these committees simply look at CVs for publication record?
12. Limited use - not required but recommended.

What challenges or successes would you like us to be aware of your institution's use of your teaching evaluation tools?

1. These tools are used in assigning teaching assignments.
2. I have covered most of the challenges.
3. Le taux de réponses des étudiants n'est pas très élevé.
4. Several years ago, attempts were made to develop a formative evaluation process, but these were deemed too resource intensive and impractical for implementation.
5. Having an institutional level recommended Teaching Dossier and self-evaluation is particularly valuable.
6. Assessing a dossier is a challenge since those assessing do not really understand them
7. Have successfully changed questions to make them more relevant for formative and summative evaluation.
8. The challenge is that we do not have a standardized tool. We anticipate the next steps forward to entail: a) developing the online course evaluation capacity; and b) adding an additional small bank of key questions to each existing SRI so as to enable the collection of key performance indicators across the institution
9. Moving to an online system in the near future .....
10. Their validity is constantly being called into question (i.e., student evaluations of teaching); the peer instruments are focused on "how" we deliver instruction, in contrast to "what - content" is

Survey 2
Policies and Administrative Responsibilities Survey

Who takes part in summative evaluation of teaching?

Response	Chart	Percentage	Number
Students		86%	12
Peers		36%	5
Instructor		36%	5
Department Head		50%	7
Teaching and Learning Centre		7%	1
Varies by Department		7%	1
Other, please specify...		21%	3

Who takes part in summative evaluation of teaching? (Other, please specify...)

1. I am not sure what the question is asking
2. For tenure track instructors, an evaluation committee is formed
3. Dean

Please indicate the key themes addressed by the Student Ratings of Instruction (SRI) for both Instructor and Course Evaluations.

	Instructor Evaluation	Course Evaluation
Organization	64%	73%
Clarity	80%	70%
Enthusiasm/Stimulation of Interest	78%	44%
Rapport with students/Interaction	91%	27%
Responsiveness	78%	44%
Instructor accessibility	89%	33%
Assessment	62%	62%
Exam/Grading fairness	56%	56%
Materials	33%	89%
Activities	20%	100%
Course difficulty	40%	100%
Workload	33%	100%

	Instructor Evaluation	Course Evaluation
Recommend to others	50%	100%
Enthusiasm for the course	83%	33%
Student self-assessment of learning	43%	86%
Overall effectiveness	80%	60%
Other	100%	0

Can departments or individuals add additional questions to the SRIs?

	Yes	No
Departments	67%	33%
Individuals	73%	27%

How are teaching evaluations used at your institution?

Response	Chart	Percentage
Personal use		100%
Performance review		83%
Promotion		92%
Tenure		92%
Departmental, Faculty, or Institutional Planning		50%
Student course selection		8%
Other, please specify...		0

Who has access to teaching evaluation data?

Response	Chart	Percentage
Individual Faculty Member		100%
Department Head		58%
Faculty Office		58%
Students		8%
Centre for Teaching and Learning		0
Provost's Office		50%
Institutional analysis		8%

What teaching evaluation data do instructors receive?

Response	Chart	Percentage
Raw Scores		75%
Weighted Scores		42%
Ranked Scores		42%
Contextualized Scores*		58%
Aggregate results		42%
Data visualizations		42%
Comments		67%
Other, please specify...		8%

* Scores in the context of departmental means or other contextualizing data

Appendix F

Terminology Used in Tenure Policies to Describe Teaching Expectations

Terminology Used in Tenure Policies to Describe Teaching Expectations

Standards of Performance for Teaching Contributions	Institutions
Effectiveness/effective	British Columbia, Calgary, Carleton, Dalhousie, Laurentian, Manitoba, McMaster, Nipissing, Northern British Columbia, Ryerson, Saskatchewan, Simon Fraser, Toronto, Victoria
Demonstrated effectiveness	Ontario Institute of Technology
Documented effectiveness	Memorial
Established effectiveness	Lethbridge
High degree of effectiveness	Algoma
Quality and effectiveness	Prince Edward Island, St. Mary's, St. Thomas
Sustained satisfactory and effective	Brock
Record of performance	Guelph
Good performance	Acadia
Acceptable performance	Cape Breton
Reasonable performance	McGill
Satisfactory performance	Bishop's, Lakehead, Laurentian, Mount Saint Vincent, Prince Edward Island, Winnipeg
Strong performance	Waterloo
High standard of performance	British Columbia
Superior performance	McGill
Teaching excellence	St. Francis Xavier, Toronto, Windsor
Satisfactory quality	Lakehead
Good quality	St. Thomas
High quality	Mount Allison, Trent
Exceptional quality	Lakehead, Winnipeg
Demonstrated superiority	York
Satisfactory record	Northern British Columbia, Thompson Rivers, Wilfrid Laurier
Strong record of achievement	Alberta
Sufficiently strong record	Western Ontario
Competence	New Brunswick, Ryerson, Toronto
Quality of competence	Concordia
Demonstrated competence and responsibility	Northern British Columbia
Scholarly competence	Laurentian, Manitoba, Memorial, Nipissing, Ryerson, Wilfrid Laurier
Meets expectations	Ottawa
Good teacher committed to academic and pedagogical excellence	Queen's, Waterloo
Success	Brandon, Prince Edward Island, St. Mary's
Promise	Nipissing
Sustained commitment	Simon Fraser

Gravestock, 2011, p. 149*

*Gravestock, P. (2011). Does teaching matter? The role of teaching evaluation in tenure policies at selected Canadian universities. (Unpublished doctoral dissertation). University of Toronto, Toronto, Ontario.

Appendix G

Teaching Evaluation Toolkit:
Report of the Technical Team

Teaching Evaluation Toolkit: Report of the Technical Team

Bill Muirhead
UOIT

Mark Naylor
UOIT

Graham Fawcett
University of Windsor

June 2014

Contents

1 Executive Summary

The technical elements of the Teaching Evaluation Toolkit, initially envisioned as a large integrated system, will best be developed through two independent, focused prototyping projects:

1. an Electronic Teaching Dossier Development Tool
2. an SRI Visualization and Analysis Tool

We recommend that the management and delivery of SRI surveys is best addressed by licensing a third-party, best-of-breed SRI management solution, and not through development of new software.

Our selected methodology, *Disciplined Agile Delivery*, ensures a process that fosters continuous stakeholder engagement. Through incremental development and frequent milestone reviews, the methodology enables rapid development of always-usable software throughout the project lifecycle.

We describe the technical deliverables, known requirements, and initial development plans for the both Phase Two prototyping projects, and conclude with a set of technical notes intended to bootstrap the Phase Two development process.

2 Introduction

The initial concept for the Teaching Evaluation Toolkit (see Figure 1) was a sophisticated and deeply integrated system comprised of many parts: an SRI reporting framework; tools for aggregate data analysis; facilities for managing academic surveys, peer reviews, and other evaluation activities; and a teaching dossier development component, which incorporated evidence from the other tools into a unified dossier.

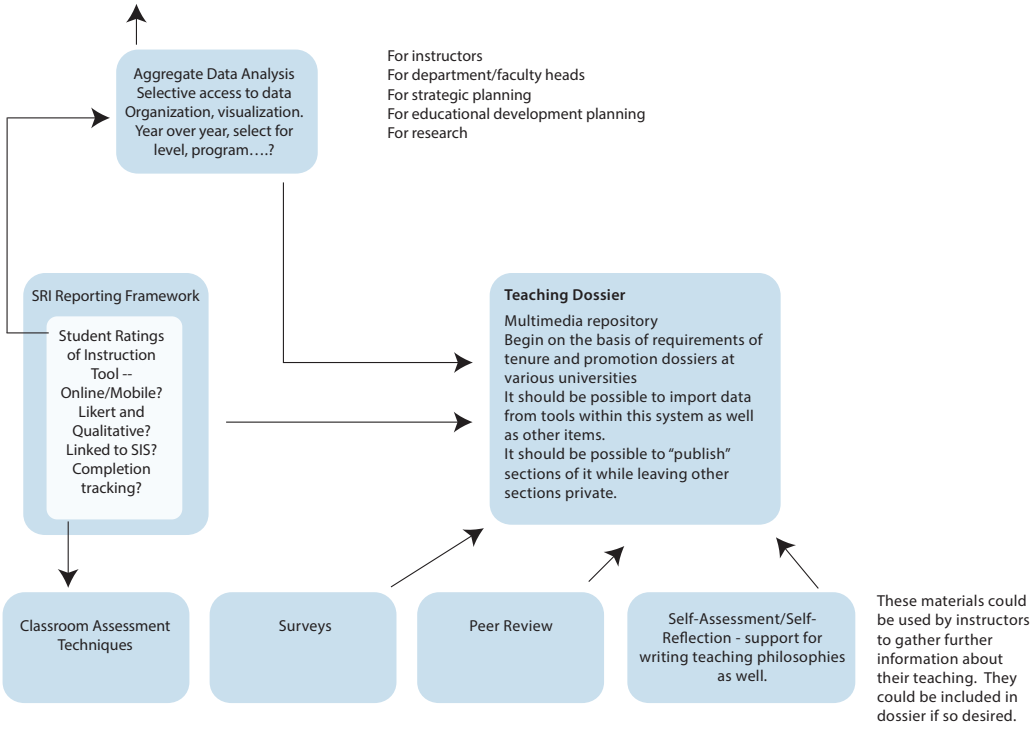


Figure 1: Initial conceptual diagram

However, during the research phase, the principal investigators concluded that universities in Ontario are neither culturally ready for, nor actively demanding, such a large-scale system. Based on an assessment of feasibility, demand in the milieu, best chance for making change, and evidence from the Effective Practices research group, the principal investigators determined that a number of smaller, highly-focused projects was a more effective approach to providing very similar functionality (see Figure 2).[1] So, although we had determined that the large, integrated system was technically feasible, we began to gather requirements and evaluate possible design approaches for two focused projects, which will be discussed in this report: the *Electronic Teaching Dossier Development Tool*, and the *SRI Visualization and Analysis Tool*.

[1] For further discusison of this decision, please refer to the recommendations of the Feasibility Study, p. 53.

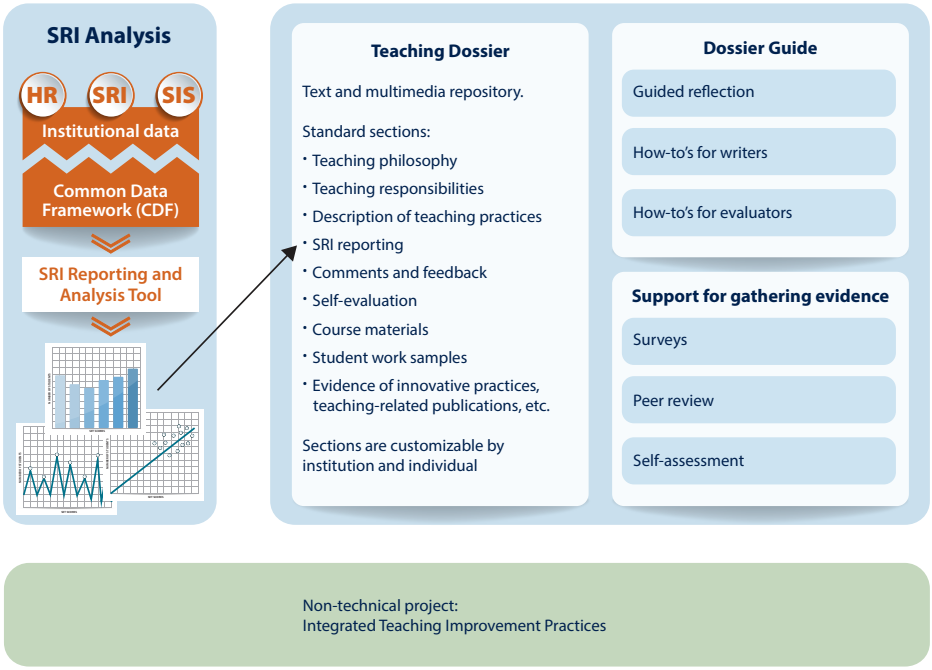


Figure 2: Revised conceptual diagram

3 The case for third-party SRI tools

The initial proposal for the Teaching Evaluation Toolkit project called for our team to produce an "internal SRI development proposal including proposed interface, analysis potential, and costs for comparison with third-party information." As discussed within the Feasibility Study, the custom development of an SRI management system — while technically feasible — was ruled out during our Phase One investigation based on two factors. First, our preliminary requirements gathering (based on input from the research team, as well as research on existing internal and third-party SRI systems) highlighted the tremendous potential complexity of such a system, particularly given the need to deeply customize the application to accommodate the varying needs of each participating university. Second, the environmental review conducted by Joughin and Boujos (2014)[2] establishes a solid business case for adopting a third-party SRI system, and identifies specific SRI systems worthy of consideration.

Joughin and Boujos identify two best-of-breed third-party SRI solutions. The IDEA system[3] is well regarded and has a solid research foundation; but they conclude that IDEA is an unlikely candidate for the Canadian sector, given its data-retention policies and the standardization of questions (which conflicts with current provincial culture). Joughin and Boujos' concluding endorsement is for Blue, a survey management system offered by the Montreal-based firm eXplorance.

Joughin and Boujos note that their review did not consider the possibility of consortial development of an SRI management system, nor of consortial licensing of a third-party solution. In our experience, consortial system development is especially challenging, particularly in the academic sector; and the consortial development of such a complex system as an SRI management system should be regarded as a high-risk and expensive venture. Consortial licensing of a third-party system is a viable alternative, with precedent in the sector, and is worthy of further consideration.

[2] Joughin and Boujos (2014) is included as Appendix "H" of the Feasibility Study.
[3] IDEA home page: http://ideaedu.org

If the Ministry is seeking a standardized, sector-wide SRI management system, our team concurs with Joughin and Boujos' recommendation to adopt a third-party system, rather than to build a custom system. However, our researchers have determined that Ontario universities are **not culturally prepared for sweeping changes to SRI practice**. This finding should be strongly evaluated before such a project is undertaken.

4 Development methodology

Selecting an appropriate development methodology is crucial to project success. For these two projects, we have selected a methodology that is well suited to exploratory, participatory development and which emphasizes engagement with stakeholders across the university sector. The methodology, known as *Disciplined Agile Delivery* (DAD; Ambler & Lines, 2011, 2012), derives best practices from lean development methods such as Scrum and the Agile Unified Process. It emphasizes stakeholder participation, purposeful milestone reviews, incrementally delivering consumable solutions at each iteration, and allowing stakeholders to revise requirements throughout the development process (Ambler & Lines, 2011). The DAD process covers the entire lifecycle of product development, from conception through ongoing maintenance and adaptation (see Figure 3). Since the two proposed projects are intended to foster incremental and lasting cultural change, it follows that the tools produced must grow and adapt at a common pace with the cultural change effort. The DAD process provides a comprehensive framework for guiding development in a manner that continuously reflects the changing needs of the stakeholders.

Roles The DAD process identifies five primary personnel roles which interact to ensure project success (Ambler & Lines, 2011). Each project team will include:

- **stakeholders** who are materially impacted by the outcome of the project. Our key stakeholders are the principal investigators of the research project, but also include: chairs of promotion/tenure committees and University Secretariats at participating universities; IT professionals who must integrate their systems with our tools; and the instructors and administrators who will use our prototypes and ensure that the needs of their constituencies are being met. Ensuring that technical development is driven and steered by stakeholder input is a key deliverable of the DAD process.
- the **product owner**, who speaks as the "single voice" of the customer. In our context, the product owner will be a key member of the research team, though not necessarily the principal investigator. This is the primary non-technical contact person for the development team. She prioritizes work, seeks answers to developer's questions about the project, and controls the scope of development. The product owner is also the "voice" of the developer team back to the stakeholders, giving status reports and coordinating demonstrations and presentations.
- **team members** who perform the technical work. The DAD process encourages teams comprised of "generalizing specialists" — team members with expertise in a few disciplines, but with a wide range of general skills. Such a team design improves collaboration and communication, and allows the project to adapt to changing priorities more efficiently. Given the generalist IT practices at many universities, building an effective DAD team of academic technologists is a feasible goal.
- the **team lead**, a member of the technical team responsible for scheduling, motivation, communication and resource management; and the **architecture owner**, a technical team member with a specific responsibility to oversee the evolution of the overall system design. Architectural design is largely about risk management, and ensuring that all technical effort is focused on delivering a coherent and correct solution. Given the small size of these projects, it is likely that the architecture owner and the team lead will be the same person.

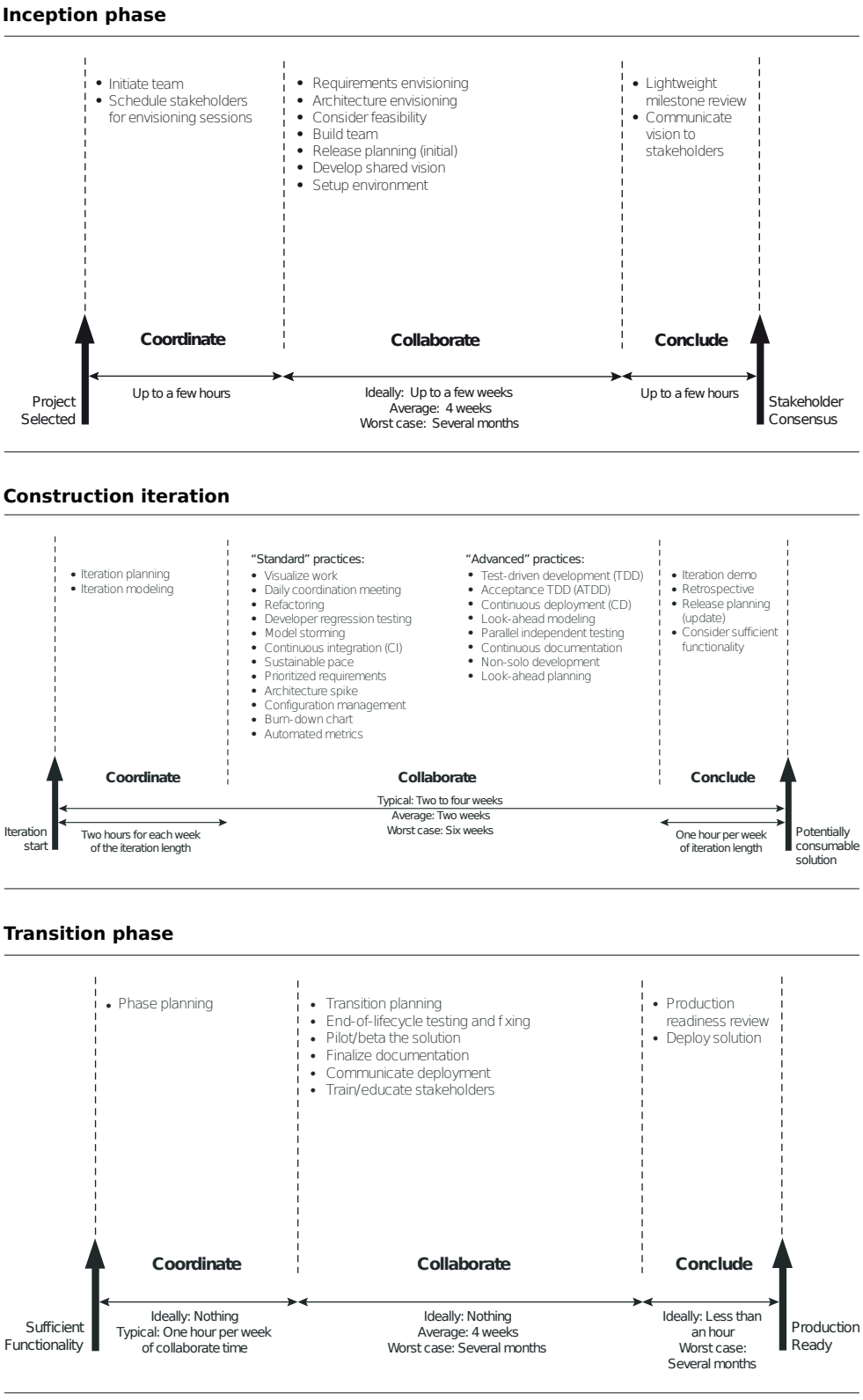


Figure 3: The Disciplined Agile Delivery lifecycle (Ambler & Lines, 2011).

Each of the two development projects will involve multiple institutions. Optimally, the research team and stakeholders will include people from all institutions; but it will be most efficient if the team members, team lead, and product owner are co-located at one university. (This approach is also recommended by the DAD process.) Institutional teams will be comprised of members from various stakeholder groups; they will use and refine each iterative version of the software, feeding back recommendations to the development group at each iteration. This approach will optimize design efficiency while also ensuring a final design most likely to meet needs across all Ontario institutions.

Iterative evolution The speed of the rapid-application design process is both its strength and a potential risk: **high-velocity teams can potentially make very rapid progress toward the wrong goal**. The DAD process provides several checks and balances to ensure that rapid development is disciplined and continuously realigned. The establishment of primary roles, and communication among them, is one such feature. A second is that each short development iteration is followed by a *retrospective*, in which progress is carefully reviewed and next-step decisions are made — not just on the next piece of work to complete, but on the overall direction of the project, and the refinement of its overarching goals. With its emphasis on "always-working" software, DAD calls for stakeholder demonstrations after each short iteration, so that course corrections are made based on the best available evidence.

After the prototyping phase Each of the two Phase Two research projects has a key milestone, at which the researchers will decide whether (a) to build the final, production system using the current development team, or (b) to contract a third-party vendor to build the final system. The complexity of the functional prototypes will be well established by this point, and estimates for completing the production system will have a high degree of accuracy. The determination of how to build the final system will be informed by the technical group, but will ultimately be determined by the research lead and product owner in consultation with the funding agency.

If the decision is to hire a third-party vendor, we strongly recommend that the vendor be required to adopt the Disciplined Agile Delivery methodology. The "product owner" from the prototyping phase can continue in that role, optimizing the transfer of knowledge and the continuity of process. Continuing with DAD will also guarantee that the same quality controls and development transparency which enabled effective prototyping are carried into the production phase. We further recommend that the first iteration of the production development project involves the prototype development team: the prototypes will be demonstrated, explained in detail, and where appropriate, the internal designs of the prototypes will be explored in detail. This will ensure continuity during the transitional phase.

5 Project #1: Electronic Teaching Dossier Development Tool

The electronic teaching dossier will be a customizable electronic record of an instructor's practice. It will include sections for each of the elements of teaching dossiers required or recommended by Ontario institutions (such as a teaching philosophy, description of teaching practices, course syllabi, student evaluation of teaching data, teaching-related publications, etc.). Documents will be importable in a range of formats, or can be produced directly within the dossier. The electronic dossier will be multi-media compatible: instructors will be able to include video and photographic evidence as well as written documentation. The electronic teaching dossier will also be exportable as a printable document (such as Microsoft Word or PDF). Instructors will be able to give other people access to their dossier if they wish to do so.

The electronic teaching dossier is framed by a variety of customizable guide materials that support the instructors through the process of developing their teaching dossiers: in particular, self-reflection, data gathering, writing, and documentation. The framing materials will also provide guidance for those reading and evaluating teaching dossiers.

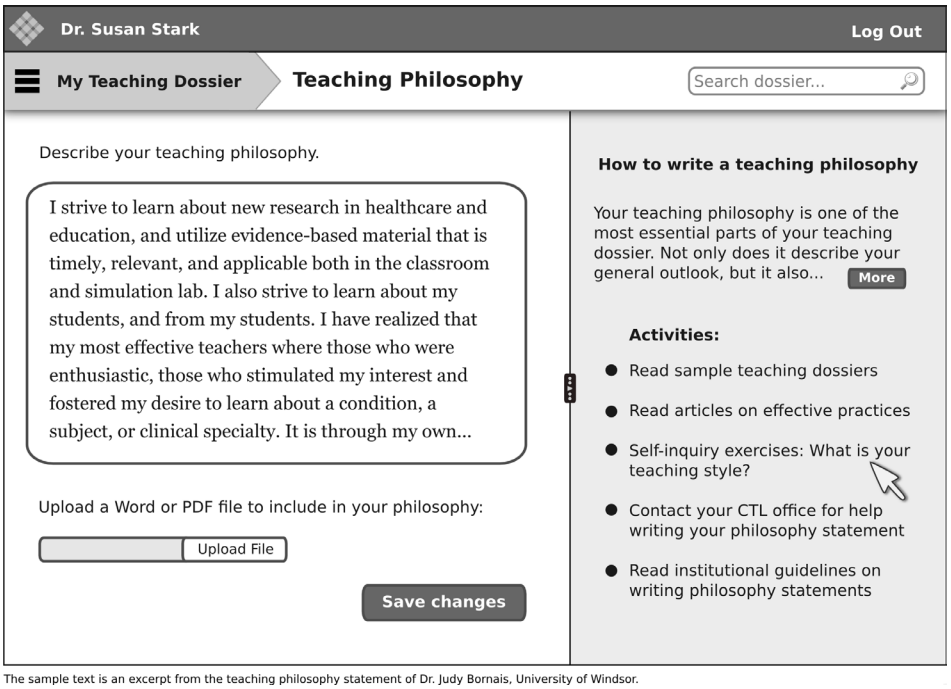


Figure 4: Adding content to a section of a teaching dossier. The integrated Dossier Guide, on the right, guides instructors through the dossier development process. (See Section 8 for more user interface diagrams.)

5.1 Intended outcomes of the project

Based on a multi-institutional consultative process, including the development of a comprehensive Dossier Development Guide:

- Create a prototype of the Electronic Teaching Dossier.
- Pilot the dossier tool with a trial group of users from the participating institutions.

- Determine how best to build a production-ready system from the prototype: (a) continuing development with the current technical team; or or (b) engaging a third-party vendor.

5.2 Description of the Electronic Teaching Dossier

The Electronic Teaching Dossier is a **Web-based** application that will enable instructors to research, organize, write, and submit a teaching dossier. Teaching dossiers are **structured documents**. They contain several sections that are common across most universities (such as "teaching philosophy"), and other sections that are unique to each university.[4] The Tool will have a **dossier template** facility, allowing instructors to choose a starting-point that is appropriate to the purpose of their dossier. The research team will design a **default template** based on effective practices research. Participating universities can **customize and extend** this template with their own sections. After creating a new dossier from a template, users may further customize the structure of the document (e.g. adding new sections) to fit their individual needs.

This "document structure" is central to the function of the Tool. A novel feature of the system is the pervasive presentation of a **Dossier Development Guide** which leads the user through the portfolio development process. The Guide is visible on-screen across all sections of the Electronic Teaching Dossier, giving **context-sensitive advice** on what teaching should be documented, and how to gather appropriate evidence. For example, in a dossier section on "teaching philosophy," the on-screen Guide would include an overview of philosophy development, links to sample philosophy statements, and a curated list of articles on philosophy writing. Each university may further **customize** the Guide, e.g., to include institutional criteria on how to prepare the philosophy statement, local experts who can give help, etc.

The user is not restricted to a single teaching dossier. **Multiple dossiers** can be created for different purposes: such as advancement, award application, and hiring purposes. Institutions may create and publish multiple dossier templates for these different uses.

Content and data for the teaching dossier can be incorporated in many different ways. The Tool will allow for direct entry of text through the Web interface. It will also allow the user to upload Word and PDF documents to be included, as well as images and other rich media (such as videos, interactive animations, and audio recordings).

By default, teaching dossiers created in the tool will be **private**: nobody but the author will be able to view or edit the dossier. But the Tool will have extensive **collaboration features**. Users may allow other people to access their dossiers: either read-only, or with the ability to edit and add new content. There will be a **commenting** facility, where reviewers can add notes and suggestions that are only visible to the author, and will not appear in final versions of the dossier. (Commenting will be permitted even if the reviewer has "read-only" access to the dossier.) The tool will also support **collaborative ownership** of dossiers: that is, a dossier may have more than one owner, and each owner will have full control over that dossier within the system. Collaborative ownership will allow the research team to explore the use of the dossier system in experimental scenarios, such as collaborative department- and faculty-level dossiers, as well as submitting dossiers for team-teaching awards.

Dossiers must be **printable**. A survey of current practices across Ontario universities indicates that most promotion and tenure committees still require a printed dossier as part of an advancement application.[5] Therefore, the Tool must be able to generate a high-quality, **professional, printable version** of a dossier. We have received strong feedback that, rather than producing a non-editable print version (such as a PDF file), the Tool should produce an **editable printing version** (such as a Microsoft Word document). This will give users an opportunity to (a) correct any formatting errors introduced by the Tool, and (b)

[4] See the Feasibility Study, pp. 44–45, for a survey of common dossier elements.
[5] For information on current promotion/tenure practices, see the Feasibility Study, p. 46.

make personal decisions about how best to format and style the document before submission. Since rich media such as videos and audio recordings cannot be printed, the Tool will include **hyperlinks to rich media** in the printable output. This will enable users to include all types of evidence in their dossiers, unrestricted by the limits of the printed page.

Designing a teaching dossier system that is **optimized both for printing and online viewing** is a technically challenging issue. Following the *Disciplined Agile Delivery* process, several early development iterations will be committed to stakeholder consultation and architectural work in order to find a viable and satisfactory solution.

The entire system, including the Electronic Teaching Dossier, the Guide, and the dossier templates, will be **brandable** by the adopting institution. Elements such as logos, colour palettes, and headers/footers will be fully customizable at the institutional level.

The Dossier Development Guide will be designed so that it can be used, not only through integration with the Electronic Teaching Dossier, but also as a **stand-alone guide** for use in other contexts. The deep and comprehensive knowledge comprising the Guide will have many applications beyond direct use in the Tool, and we should not limit these uses through arbitrary technical decisions. Therefore the Guide will be prepared in an **independent content management system**, apart from the Tool. Although the Guide is in a separate system, topical information in the Guide will be **indexed and cross-referenced**, so that the Guide can be seamlessly integrated with the Tool in an appropriate and context-sensitive manner. (For example, when editing a "teaching philosophy" section in the Tool, only the "philosophy" materials from the Guide will be visible on-screen.) Just as the Tool will be brandable and configurable by the adopting institution, the Guide will also be **editable and extensible**, allowing institutions to include important information and instructions that are specific to their local requirements.

5.3 Mahara as an initial prototype

As an initial vehicle for exploration, we recommend that the research team use Mahara,[6] an open-source application which is popular as a student-learning portfolio tool. It has also been used, though not systematically, in preparing teaching portfolios at the primary and secondary level (Murphy, 2011).[7]

Although Mahara lacks the features required for a fully-functioning Electronic Teaching Dossier, it can effectively serve as an initial prototype and comparator. It can be quickly installed at the start of the project, enabling researchers to begin refining their ideas about portfolio structure and design, and developing a clearer vision of the "user experience" they wish to see implemented in the Electronic Teaching Dossier. This will free up the development team to tackle key architectural issues (particularly: high-quality printing, and the integration of the Dossier Guide) which must be resolved before a proper prototype can be developed. On an ongoing basis, Mahara can continue to serve the project as a baseline comparator for assessing the feature completeness of the Electronic Teaching Dossier prototype.

We anticipate that the research team will make effective use of Mahara for a few iterations of development. By this point, the development team will have addressed the core architectural issues, and can begin development of a prototype which will be used and iteratively refined throughout the remainder of the project.

[6] Mahara home page: http://mahara.org/

[7] For the ECIS standards referenced in Murphy (2011), see: http://www.internationalteachercertificate.com/.

5.4 Research and development process

The participating stakeholder institutions will have already been identified in the project charter. From these institutions, the following teams will be established. The makeup of these teams is heavily guided by the *Disciplined Agile Delivery* process (see p. 5).

1. The **research team** including the principal investigators and the product owner. This team will liaise with institutional stakeholders, write guides and documentation, interact with user groups, and guide the development team. Key activities will include:
    - Establishing and implementing institutional consultation processes to gather input regarding requirements, policies, and desirable features of the Electronic Teaching Dossier
    - Writing an online Dossier Development Guide. This will include original material and sample documents, as well as material gathered through from internationally recognized sources. (As needed, the team will secure permissions to include third-party materials in the guide.)
    - Working with developers on the design of the prototype, and reviewing the prototype with key institutional stakeholders in the partner institutions.
    - Forming a multi-institutional user group who will prepare their dossiers using the prototype; and working with this group and the developers to adjust the prototype design.
    - Key milestone: In conjunction with the funding agency and development team, determine whether to continue the *Disciplined Agile Delivery* process with the current development team, or with a third-party vendor.
2. The **development team** including the team lead and architecture owner, who will design and develop the system. Ideally, the development team will be co-located with the product owner (from the research team). Key activities will include:
    - Setting up Mahara as an preliminary prototype (see p. 10).
    - Identifying, and implementing solutions to, the challenges of generating a professional-quality printable document from a diverse set of input documents.
    - Setting up a content management system (CMS) to contain the Dossier Guide.
    - Designing an indexing/cross-referencing scheme to integrate the Dossier Guide CMS with the Electronic Teaching Dossier.
    - Identifying requirements for Electronic Teaching Dossier data migration and inter-institutional portability.
    - Contributing to decision regarding the ongoing development of the system (using the same team; or hiring a third-party vendor) by reviewing the outstanding technical challenges with the stakeholders.
3. The **user groups** at each institution will test iterations of the software, and give feedback on the user interface, user experience, effectiveness of the system, and quality of the written documentation and guides.

6 Project #2: SRI Visualization and Analysis Tool

The goal of this project is to build a set of data visualization and analysis tools for use by individual instructors and department heads. The tools will enable users to aggregate and disaggregate their data in a variety of ways. For example:

- aggregating data (e.g., instructor and course scores, as well as sub-scales if viable) across courses within a department, and across courses for an instructor
- contextualizing data within identified norm groups
- disaggregating data (at the program chair level) based on demographic data (e.g., required vs. elective course, student year, student program)
- representing multiple data points in one visualization (e.g., course and instructor scores across time, organized by course, and historically)
- representing score distributions
- creating and visualizing descriptive and statistical calculations, including frequency counts, measures of central tendency, bimodal variations, contingency tables and significance tests, and regression analysis
- SRI data annotation
- Preventing statistically unfounded comparisons and alerting users to weak bases for conclusions

Visualization is a critical tool: it is nearly impossible to identify patterns from raw SRI scores, and even summary tables can be very difficult to interpret. Visualization tools can be of significant value to faculty as the basis for ongoing critical inquiry into practice, growth, and areas for further development.

The underlying premise of this project is that Student Ranking of Instruction (SRI) data are different across universities, but are similar enough that a common tool can be designed to visualize, analyze, and help interpret them. Universities do not need to replace their current SRI infrastructure to take advantage of this tool; they simply have to integrate the tool with their campus information systems — a "behind-the-scenes" activity that will leave their current SRI infrastructure otherwise unchanged.

We underscore that the purpose of this project is *not* to aggregate and analyze data across universities. Under our model, each institution maintains full and independent control of its data; the Analysis Tool can be installed and administered independently at each institution.[8]

6.1 Intended outcomes of the project

- Design a Web-based SRI analysis tool that is independent of any specific SRI delivery system.
- Design a method for mapping institutional SRI (and related) data into a standard format, upon which reusable visualizations and analyses can be based
- create a Web-based collection of reusable visualizations and reports tailored to support teaching evaluation, inquiry, and improvement
- Design a data annotation system that enables users to mark external events that may explain variations in SRI results

[8] During the research phase, anonymized data from participating institutions may be stored in the same research database. This data will only be used in accordance with the research ethics agreements signed with each participating institution, and then it will be deleted.

Key non-technical outcomes:

- Write an ethical practices guide for SRI exploration and SRI-based decision making
- Identify and document common errors in SRI analysis
- Determine a set of effective practices for using data annotations on SRI data to improve interpretation
- Develop a preliminary glossary and support materials for use of the visualizations and the Analysis Tool.

6.2 Description of the SRI Visualization and Analysis Tool

The Common Data Framework The majority of data relevant to SRI analysis falls into three broad types:

- categories into which SRI questions can be classified,
- student responses, which are mostly numerical responses to Likert-style questions,
- institutional data about the departments, faculties, instructors, and students who participate in the SRI process.

At a typical university, these data may be spread across three data systems: a survey management system, a student information system, and a human resources system (for instructor information). In order to build an SRI analysis tool that can function across the sector, we must define a common format for SRI data, and related institutional data, so that they can be imported into the Analysis Tool.

We propose to develop a universal data format and import method, which we call the *Common Data Framework* (see Figure 5). The Framework is the single "coupling point" between an institution's diverse data systems and the Analysis Tool.

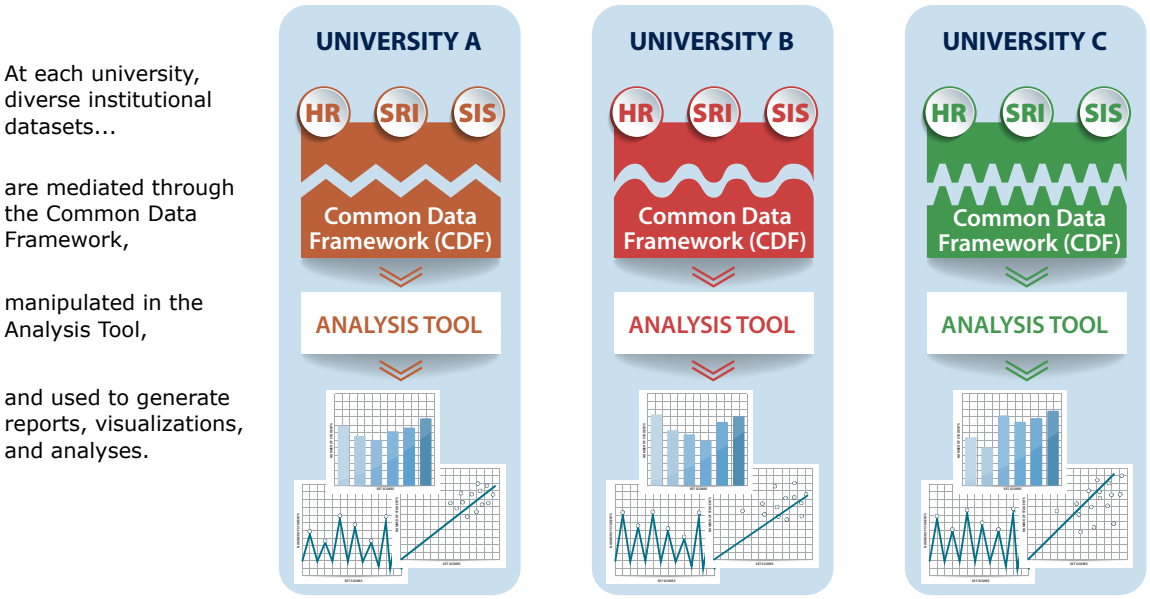


Figure 5: Conceptual diagram of the Common Data Framework.

Key features of the Common Data Framework include:

- **Built-in anonymization features**. The Framework will be designed to prevent personal identifying information from being exposed in the Analysis Tool. For example, student identifiers (such as student numbers) could be "scrambled" or stripped out so that they cannot be used to identify respondents.
- **A taxonomy of common question types**. While specific questions will vary among institutions, most SRI questions fall into common categories, such as:
  - questions about the instructor vs. about the course
  - "overall impression" questions vs. detailed questions
  - common "detailed" questions such as instructor preparedness, availability, topical knowledge
  - institution-wide questions vs. questions designed and added by individual instructors

  The Framework will provide mechanisms for *defining* these categories, and *mapping* questions into them, so that common reports and visualizations can be designed for use at all institutions. The system will permit new categories to be created, to accommodate questions that do not map into "universal" question types.
- **Accommodating different response formats**. Likert-style questions are common in SRI's, but the scales often vary (e.g., 1–5 vs. 1–7), even within an institution. The Framework will accommodate any Likert-style range, as well as other "discrete" question types (such as binary "yes/no" questions).
- Long textual answers to open-ended questions. Open-ended answers will be importable, and viewable from within the Tool. (However, our reports and visualizations will focus on analyzing short-answer, "numerical" responses.)
- **Detailed information about courses and sections** (such as year, class size, faculty, department, campus and building, evening/weekend, blended or distance, lecture or other format).
- **Demographic data** about students and instructors, such as:
  - for instructors: rank, career year, discipline; role and reporting relationships
  - for students: program, year of study, full-time/part-time
  - for all people: department, faculty affiliations
  - student's relation to course: required, elective
  - By design, the Framework will *not* support importing of protected demographic data that may exist in some systems, such as gender, country of origin, disability status, etc.
- **Access control directives**. While access permissions can be configured in the Analysis Tool itself, some institutions may also have "data-level" restrictions in their databases. For example, a survey in an SRI survey system might be configured such that only Instructors A, B, and C may view the survey results. The Framework will allow for these access restrictions to be imported alongside the survey data itself.

The Common Data Framework will be designed iteratively and collaboratively. The development team will establish an initial design, and adapt this design until the stakeholder institutions' data can be correctly imported. We expect the Framework to evolve further as new institutions (with unique, local requirements) adopt the Analysis Tool. We will strive to keep the Common Data Framework as technically and conceptually simple as possible, using established and widely-used technologies, to minimize adoption costs. Extensive technical documentation will be written on an ongoing basis as the Framework evolves.

We emphasize that the adoption of the Common Data Framework is a **one-time cost** for an institution. Once the "mapping" of institutional data into the Framework is complete, the Analysis Tool will be available for use going forward. Institutions will only need to "re-map" their data into the Framework if they replace a key campus information system (such as their HR system).

Data security and access control The Feasibility Study indicates that rules governing access to SRI data, both aggregated and disaggregated, vary widely across the sector. The Analysis Tool does not prescribe or mandate any access control rules; rather, it will have a **fine-grained access control model** which each institution can adapt to meet its local needs. We will not assume, for example, that department heads always have access to their departmental SRI data, as some institutions require instructors to formally grant these rights to administrators on a case-by-case basis. (We accommodate this scenario by allowing these "releases" to be imported alongside the data via the Common Data Framework). That said, the Analysis Tool will have a default set of access controls that map onto typical rules found across the sector, thus simplifying installation and configuration.

Access control permissions are *transitive*: if an administrator only has permission to view certain SRI results in the department, but not others, the administrator can only perform aggregate analyses on the permitted results; the others will be excluded from the analysis (and the Analysis Tool will notify them of this fact). To accommodate cases where aggregated data should be more widely accessible, the Analysis Tool distinguishes between different **levels of access**: the right to view data in detail, and the right to view it in aggregate. For example, an administrator could be allowed to run a full aggregate report across the department, but not allowed to "drill down" into the data.

Visualization and Analysis Tool Having established a method for importing the data, and a model for managing permissions over the data, we can build a shareable system for analyzing that data. The Visualization and Analysis Tool will be a Web-based application that can be installed independently at each participating institution. Based on the Common Data Framework, the research team (and users of the tool) can collaboratively build useful, shareable SRI visualizations and reports.

Key features of the Analysis Tool include:

- Web-based tools for quickly and easily creating a wide variety of **charts, tables, and visualizations** from available datasets
- Built-in support for **common statistical functions** such as measures of central tendency (mean, median, mode, variance), significance tests (χ^2), regression analysis, banner tabulations (pivot tables), etc.
- **Active prevention of comparisons that are statistically unfounded** (e.g., small sample size) and providing cautions where appropriate (e.g., weak correlation). For example, if there is a "button" in the tool to produce a certain report, but there is insufficient data to produce a statistically significant report, the button will be disabled and an explanation will be given to the user (see Figure 7).
- **Easy exporting** of tables, reports, and visualizations (e.g., into the Electronic Teaching Dossier)
- Fine-grained, **customizable access controls** over SRI data and analyses

Using prototypes of the SRI tool, the research team will produce a collection of **reusable visualizations and reports** that will be incorporated into the Tool as recommended reports for instructors and administrators. These tools will be informed by effective practices research, and will be reviewed for statistical and ethical validity.

When a user first logs into the Analysis Tool, their **user dashboard** (Figure 6) will display recent SRI activity (e.g., "last term's courses") as well as a list of **recommended starting points** based on the

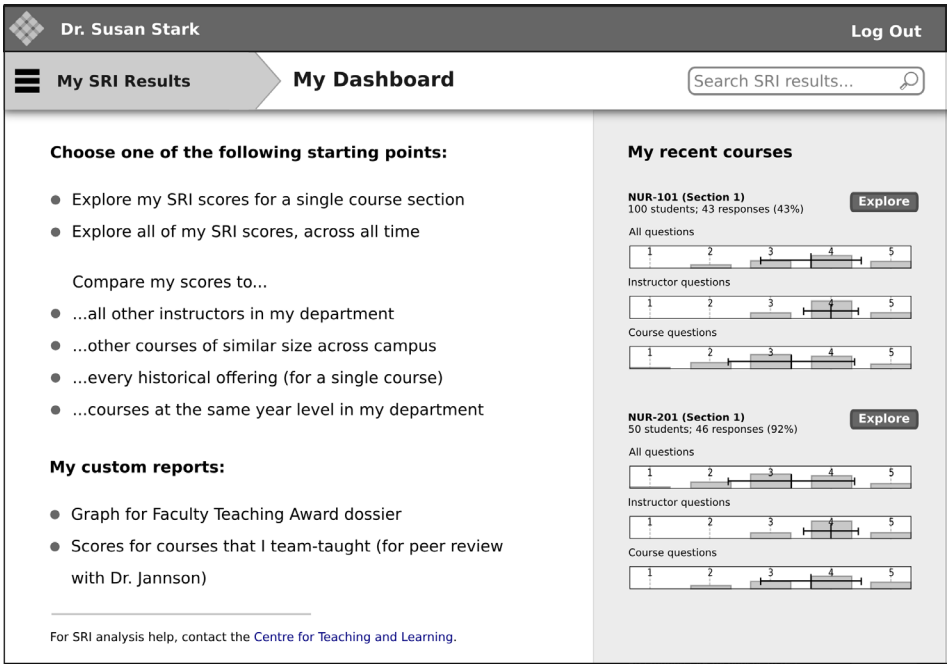


Figure 6: The "instructor dashboard" of the SRI Analysis Tool. A menu of "best practice" reports are offered, on the left, as starting points for inquiry. Recent SRI results are shown on the right. (See Section 9 for more user interface diagrams.)

best-practices reports and visualizations created by the research team. The user can immediately view the recent SRI activity, in more detail, by clicking on the appropriate "Explore" buttons near the visuals; or they can select one of the starting points to begin their inquiry. (They can also save their own "custom" starting points, based on analyses they have completed in the past.)

After selecting a starting point, the user can **adapt and customize** the analysis in a variety of ways. They can change the **dataset** for the report: e.g., changing the timeframe, selecting specific course numbers or course sections, filtering the data based on demographics, selecting which SRI questions to analyse, etc. The Analysis Tool interface will provide intuitive controls for making these dataset selections. Users can select the **analysis** to perform; this could be as simple as "just show me the raw data," or a more sophisticated analysis such as finding correlations between responses to different SRI questions. Finally, users can select an appropriate **visualization** for the analysis, such as a table (e.g., of raw SRI responses, or a contingency table to explore correlations) or a graphical interpretation (histogram, scatterplot, line or bar chart).

The pre-populated list of "starting points" is the first step in **guiding instructors** toward sound interpretations of their data. The system will further guide instructors as they continue their exploration. For example, the system will never invite users to select datasets that the users don't have permission to access. More significantly, it will disallow comparisons that are *statistically flawed*: for example, when the response rate for a survey is too low for a statistical test to have a meaningful result (see Figure 7). A key goal of this project is that **a deep understanding of statistics is not required** for using the Analysis Tool. The system will assist and prompt users toward meaningful statistical comparisons, and actively discourage users from drawing false conclusions from their data.

At any point, the user may choose to save their modified view of the data as a **custom starting point** which will appear on their user dashboard. These custom reports can be shared among users, encouraging reuse, experimentation, and discussion of SRI analysis practices among instructors and

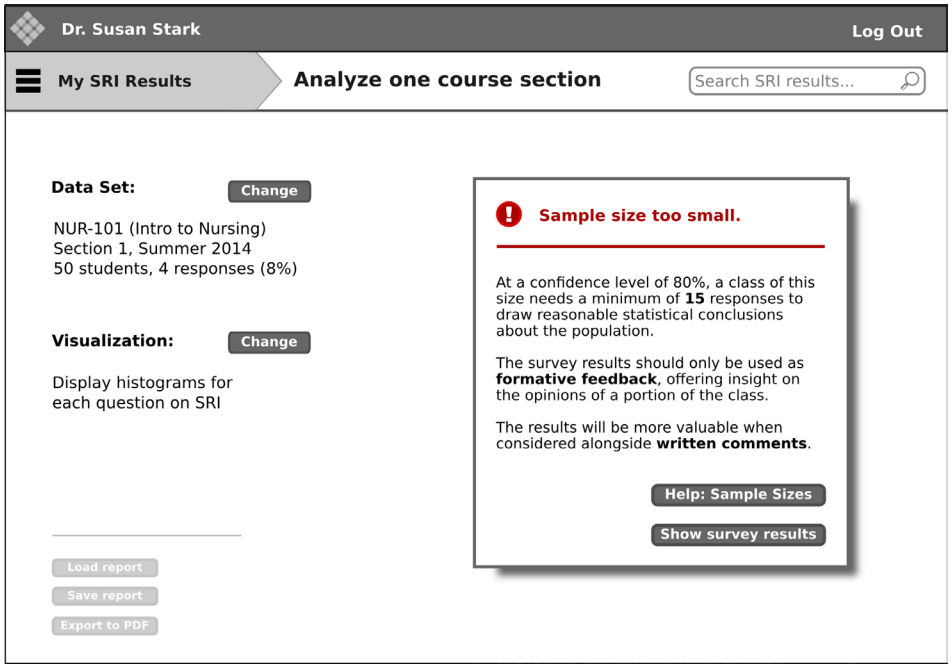


Figure 7: The SRI analysis tool will prevent users from drawing invalid conclusions: for example, when sample sizes are too small to produce a statistically significant result. (See Section 9 for more user interface diagrams.)

administrators. (Sharing a custom report does not imply sharing permissions over the underlying data; if a user does not have permission to share the data as well, then only the report settings — such as the selected analysis, and the visualization method — will be shared.)

Visualizations and reports will be **interactive**. For example, clicking on a data point in a scatterplot will "drill down" into the data, showing detailed information about that point (see Figure 8). Clicking on a "total" in a summary report will let the user explore the raw data that resulted in that total.

All tables and visualizations prepared in the Analysis Tool will be **exportable** in a variety of formats (PDF, JPG, etc.) for use in other contexts such as the Electronic Teaching Dossier (see Figure 12, p. 23).

Data annotations Variations in SRI results are sometimes attributable to known events surrounding teaching activities. For example, an instructor may try an innovative teaching method in a given term; or a building may undergo noisy renovations during the course. To assist users in keeping track of these external factors when interpreting data, the SRI Tool will provide an **annotation system** for documenting external events, when and where they occurred, and (by extension) which users and courses were influenced by them. When visualizing and reporting data, the SRI Tool will automatically include these annotations: for example, as visual interactive markers in graphs, or as footnotes to tables and reports (see Figure 8). The display of annotations can be disabled as desired.

Both instructors and administrators may add annotations to the data (see Figure 9). Each annotation will have a **description**: a short textual explanation acting as a title, and optionally a longer, detailed description. A list of annotation **types** will be provided, allowing users to classify annotations, and select which annotations to include in reports and visualizations. (For example, a "problem" annotation might describe a noisy renovation; an "experiment" annotation might indicate the use of a novel teaching method.) Each annotation will have a **timeframe**, measured in months or semesters, indicating the period of time over which the event occurred, or a point-in-time at which a significant change

was made (e.g., the introduction of a new curriculum). Finally, each annotation will have a **scope**, which determines which course sections are impacted by the annotation. Scopes can be described using locations ("all courses in the Nursing Building"), instructors ("All courses taught by Dr. Stark"), academic programs ("Nursing courses"), academic course codes ("NUR-203, Sections 1–5"), or any combination of these.

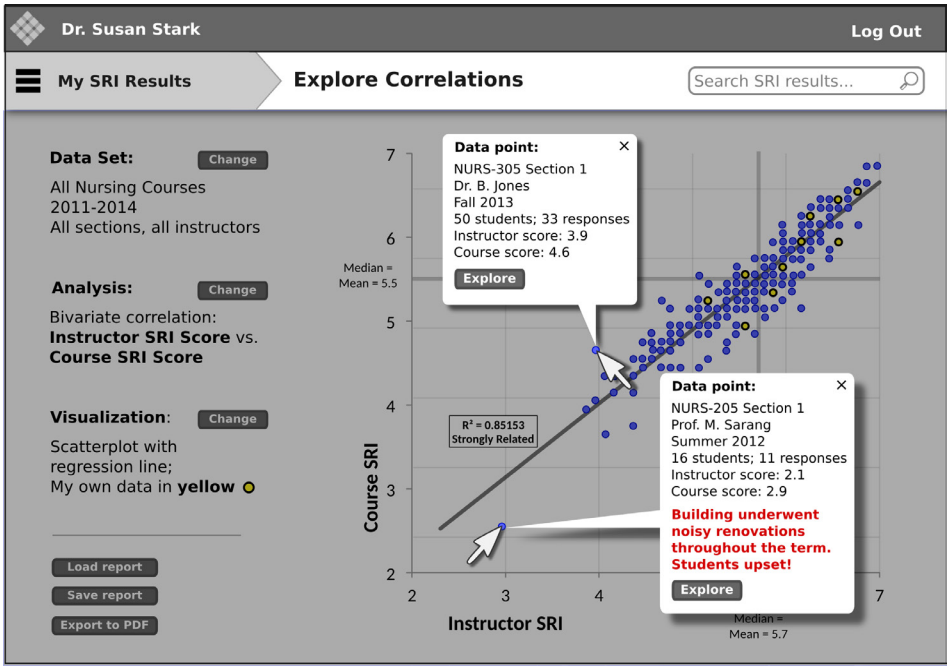


Figure 8: Interactive visualizations allow the viewer to drill down into specific data points, driving further inquiry. Data annotations, in red, allow users to note external factors that may have affected teaching and learning. (See Section 9 for more user interface diagrams.)

We believe the data annotation system to be a novel contribution to the analysis of SRI data. The research team will explore and report on their findings of the benefits of and best practices for annotating data in support of interpretation and decision making.

Key non-technical outcomes In addition to the development of the Analysis Tool, there are several key non-technical deliverables that will be completed by the research team:

- Writing a guide for the ethical use of SRI data and SRI-based decision making;
- Writing a guide identifying and documenting common errors in SRI analysis;
- Writing a guide describing the effective use of data annotations in SRI interpretation.

These guides will be available through (and will influence the design of) the SRI Visualization and Analysis Tool.

6.3 Research and development process

Groups and roles The participating stakeholder institutions will have already been identified in the project charter. From these institutions, the following teams will be established. The makeup of these teams is heavily guided by the *Disciplined Agile Delivery* process (see p. 5).

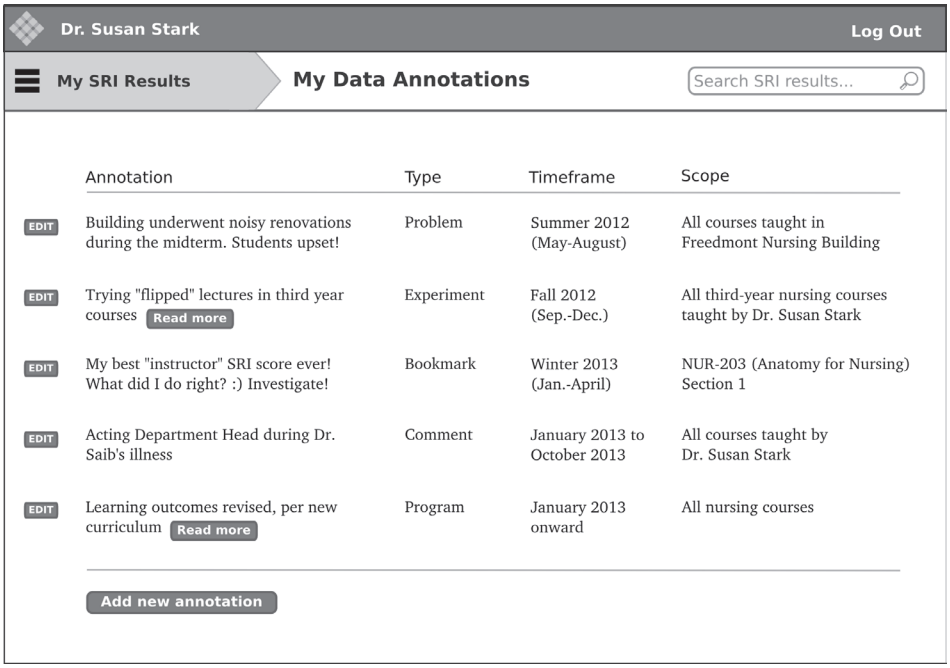


Figure 9: Data annotations. Instructors and administrators can annotate SRI data to facilitate SRI interpretation. Annotations may appear as footnotes in tabular reports, visual indicators in visualizations, or can be hidden if desired. Annotations identify events that might affect teaching and learning within a specific time-frame and scope: instructor, schedule, course, program, etc. (See Section 9 for more user interface diagrams.)

1. the **research team** including the principal investigators and the product owner. This team will liaise with institutional stakeholders, write guides and documentation, interact with user groups, and guide the development team.

2. the **development team** including the team lead and architecture owner, who will design and develop the system. Ideally, the development team will be co-located with the product owner (from the research team).

3. **institutional developers** at each stakeholder institution, who will work with the development team to (a) perform the initial export of anonymized research data into the system, and (b) implement the the Common Data Framework specification, through multiple iterations, at their institutions.

4. **user groups** at each institution will test iterations of the software, and give feedback on the user interface, user experience, effectiveness of the system, and quality of the written documentation and guides.

Centres of activity There are six major centres of activity in this project:

1. **Stakeholder relations.** The research team is responsible for establishing all of the other teams, and for identifying key stakeholders across the partner institutions. It will be their responsibility to inform stakeholders of development progress and, through the product owner, communicate stakeholder feedback to the development team. The research team must also acquire REB clearances in order to access anonymized SRI research data from each institution.

2. **Common Data Framework** specifications and implementation. The initial specification for the Common Data Framework will be written in an early development iteration. Institutional developers at the partner institutions will be involved in this process; this is essential, as each institution must implement the Common Data Framework themselves, thus providing the "glue" that binds their institutional data to the shared Analysis Tool. The initial import of anonymized SRI research data from each institution might not be performed using the Common Data Framework, as it may not yet be implemented; this data import will be handled on a "one-off" basis, and lessons learned from the import will guide development of the Framework. The final outcomes of this activity centre are (a) a detailed technical specification of the Common Data Framework, and (b) conforming implementations of the Framework specification at each partner institution.

3. **Core system development.** The development team will spend an early development iteration designing and implementing the core architecture of the Analysis Tool. This will include the internal components of the Common Data Framework, the system data model, the permissions/access-control model, the data annotation system, and the core of the visualization/analysis system. Development of the user-facing tools will begin only once the architecture has been proven. The research team will provide input through requirements and responses to questions, and will track and report on development progress.

4. **Visualization and analysis engine** and user interface. Once the core architecture is proven, the visualization and analysis system will be built. This will include all aspects of reporting, including the user interfaces for selecting datasets, analyses, and visualizations (see p. 16), and for adding data annotations and exporting reports. During these iterations, third-party visualization and statistics frameworks may be selected and incorporated into the tool. The research team and user groups will play a central and ongoing role in testing and guiding the user experience of the Analysis Tool.

5. **Best-practice reports and visualizations** will be developed by the research team in consultation with the user groups and key institutional stakeholders. The research team will conduct testing on the effectiveness of these reports, refine their design, and finally produce a set of best-practice visualizations that will be incorporated into the Analysis Tool as "starting points" on the introductory user dashboard (see p. 15).

6. **Guide and documentation writing.** The end-user documentation, the guide on the ethical use of SRI data, and the guide on data annotation best practices will be written primarily by the research team. The guides and documentation will be incorporated into the Analysis Tool by the development team.

7 Conclusion

The two focused projects that have grown out of the original Teaching Evaluation Toolkit vision are both technically feasible and likely to effect positive and lasting change for the project stakeholders. The *Electronic Teaching Dossier* takes a quantum leap beyond existing e-portfolio tools by introducing an interactive Dossier Development Guide. The *SRI Visualization and Analysis Tool* promises to bring cutting-edge SRI analysis tools to all instructors across Ontario, without requiring deep infrastructural or institutional cultural changes.

The principal research team has concluded that technology-based teaching evaluation practices must be carefully *grown*: institutional cultures must have time to evolve in tandem with technological innovation. The *Disciplined Agile Delivery* methodology, with its fundamental commitment to stakeholder success and a process of iterative evolution, is especially well suited to this challenging development scenario. We are confident that the development process we have described will rapidly deliver two high-quality, cost-effective, strategic innovations for the Ontario teaching evaluation context.

8 User interface diagrams: Electronic Dossier Tool

The following sequence of diagrams describe a possible user interface for the Electronic Teaching Dossier tool.

- We show an instructor working on her dossier, using the integrated Dossier Guide, adding textual content and SRI results, and preparing to print the dossier.
- A reviewer reads and critiques the online dossier using an interactive rubric supplied by the Dossier Guide.
- Finally, a system administrator configures the structure and branding of Electronic Teaching Dossier tool.

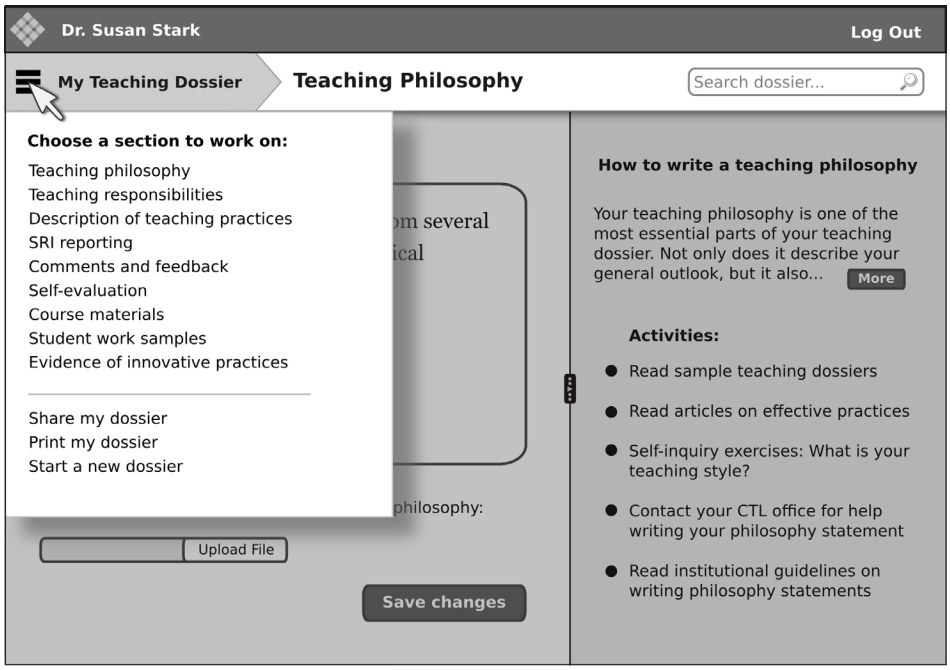
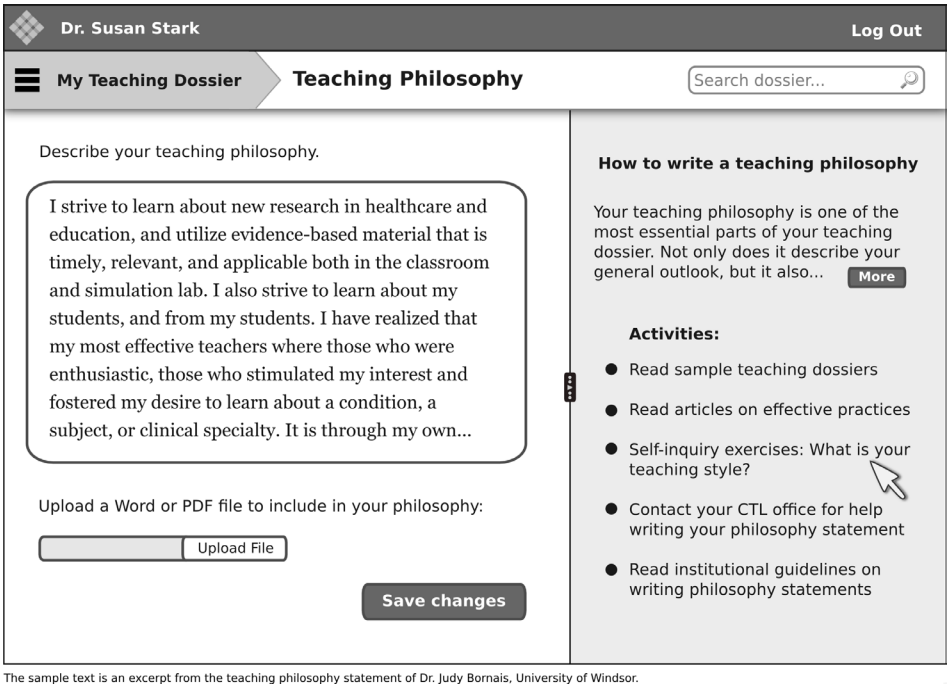


Figure 10: Navigating among the sections of a teaching dossier. Institutions and users can customize the dossier structure, adding and removing sections according to their needs.



The sample text is an excerpt from the teaching philosophy statement of Dr. Judy Bornais, University of Windsor.

Figure 11: Adding content to a section of a teaching dossier. The integrated Dossier Guide, on the right, guides instructors through the dossier development process.

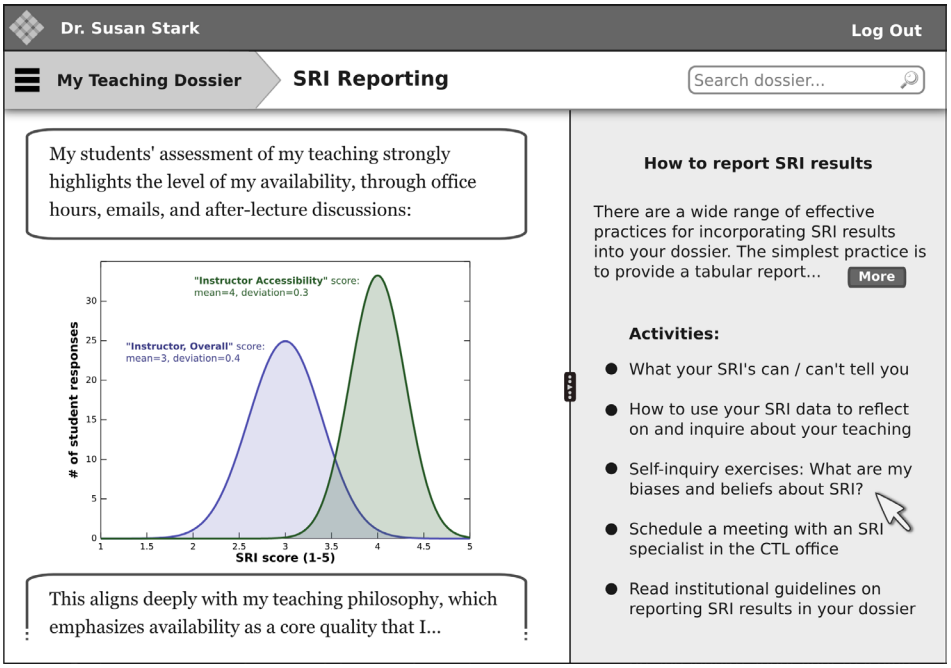
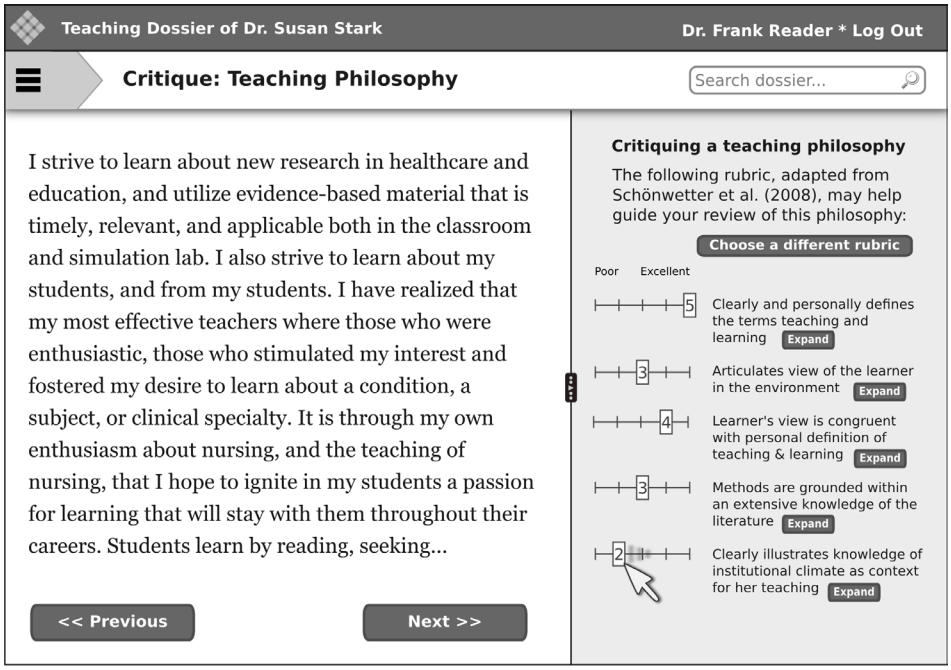


Figure 12: Incorporating SRI results in the dossier. SRI results can be exported from the SRI Analysis Tool in a variety of formats, including tabular reports and visualizations such as this comparison of two distributions. SRI tables and graphs can be embedded directly in the text of the dossier, allowing instructors to present both the data and their interpretation together in a natural fashion.



The sample text is an excerpt from the teaching philosophy statement of Dr. Judy Bornais, University of Windsor.

Figure 13: A reviewer reads and critiques a teaching dossier. The integrated Dossier Guide, on the right, provides evaluation rubrics and other aids to the reviewer.

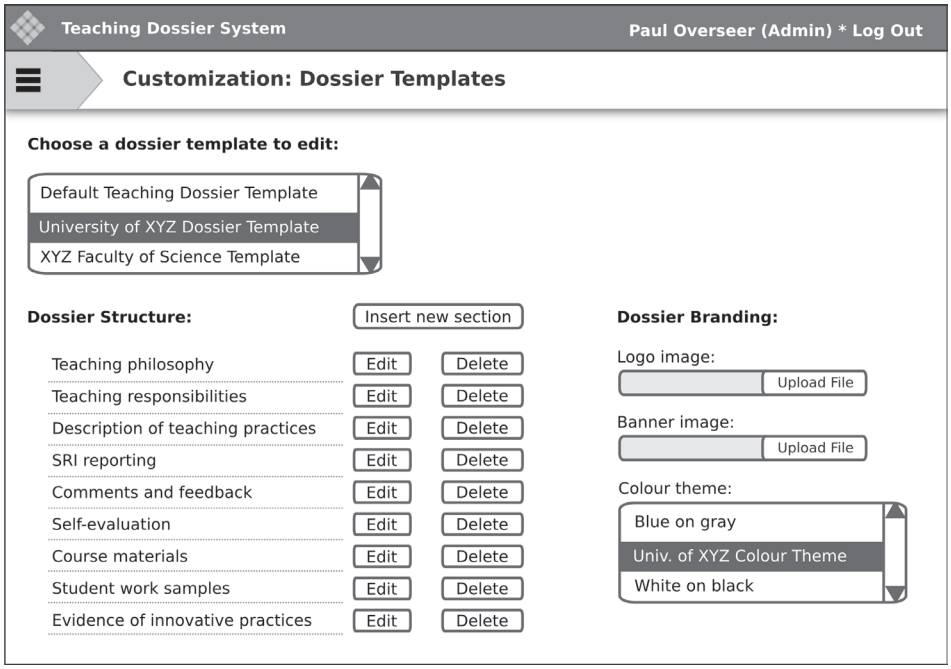


Figure 14: The system administrator can configure many aspects of the Dossier Tool: creating and editing dossier templates, adding institutional branding, and more.

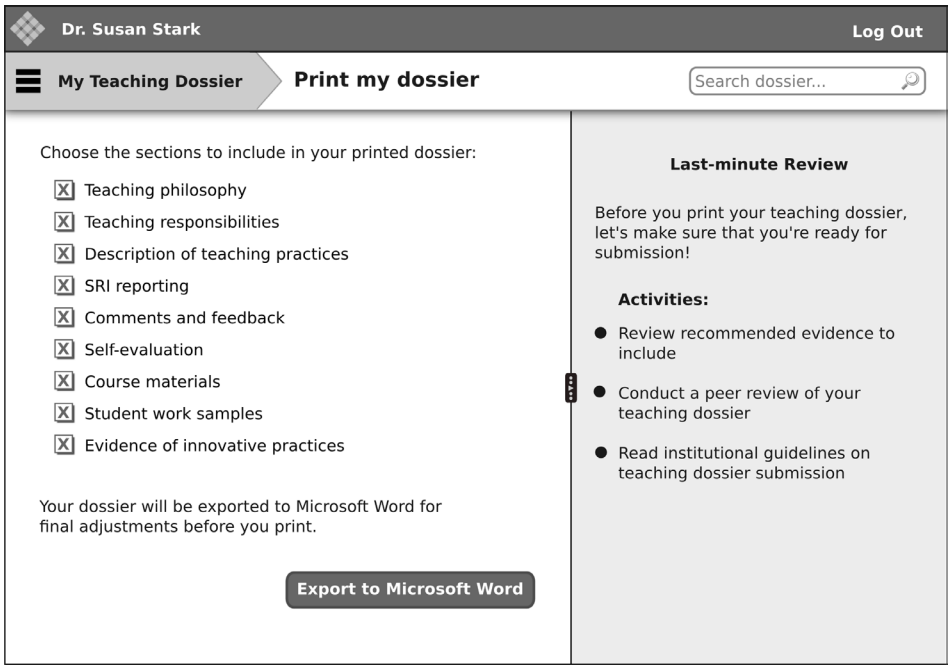


Figure 15: Printing. Instructors can export their "finished" dossier into a word processor for final style adjustments before printing and submitting their dossier for formal review.

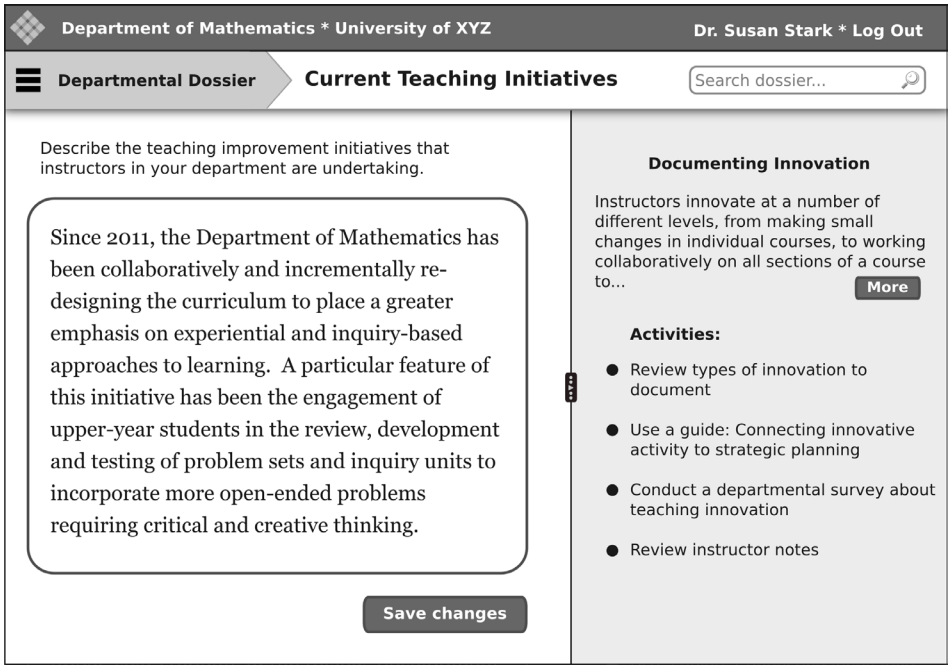


Figure 16: Collaboratively-owned dossiers enable experimental use of the Dossier Tool, such as this "departmental teaching dossier" which reports on the teaching activities of a whole department.

9 User interface diagrams: SRI Visualization and Analysis Tool

The following sequence of diagrams describe a possible user interface for the SRI Visualization and Analysis tool.

- We show the instructor dashboard, from which the instructor can start exploring their SRI results from a variety of recommended starting points
- a departmental Head's dashboard, showing reports and analyses of department-wide SRI data
- several detailed views of SRI data
- interactive visualizations of correlations between questions in the SRI data
- the use of *data annotations* (p. 17) to visually identify external events that may affect teaching and learning

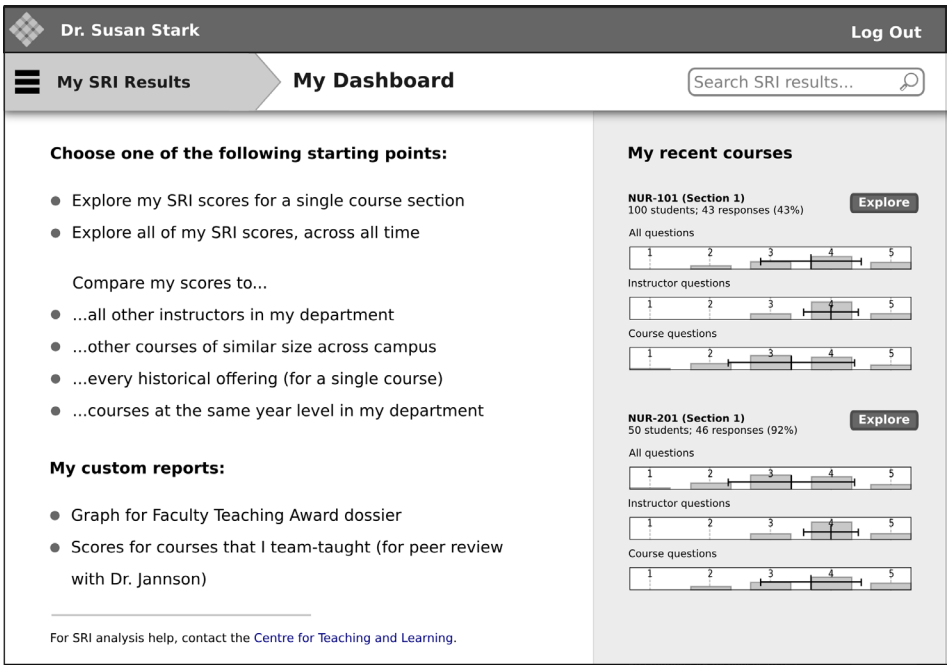


Figure 17: The "instructor dashboard" of the SRI Analysis Tool. A menu of "best practice" reports are offered, on the left, as starting points for inquiry. Recent SRI results are shown on the right.

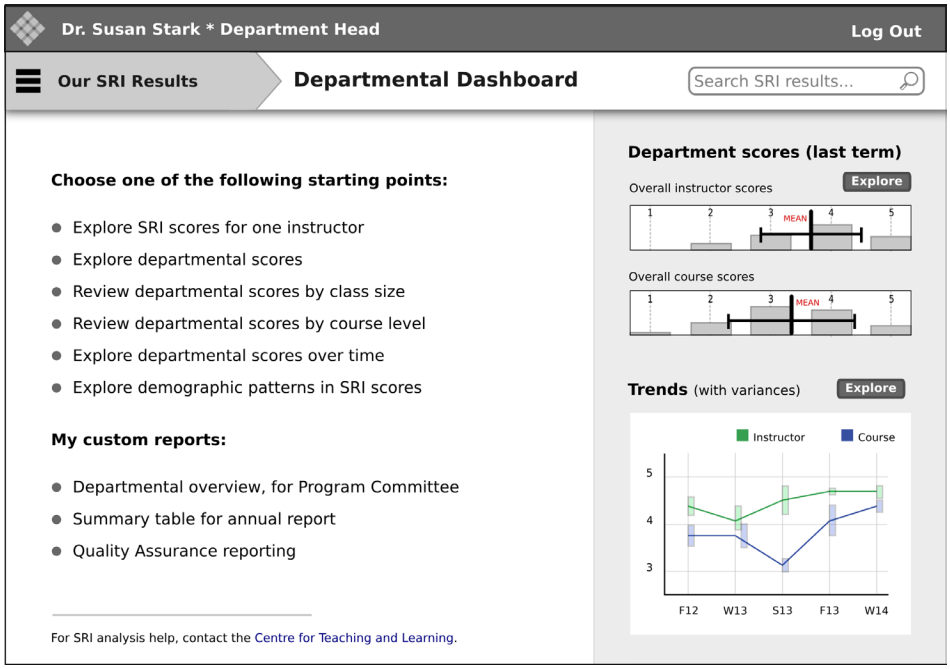


Figure 18: The "departmental Head's dashboard" of the SRI Analysis tool. A menu of "best practice" reports are offered, on the left, as starting points for inquiry. Recent departmental SRI results are shown on the right.

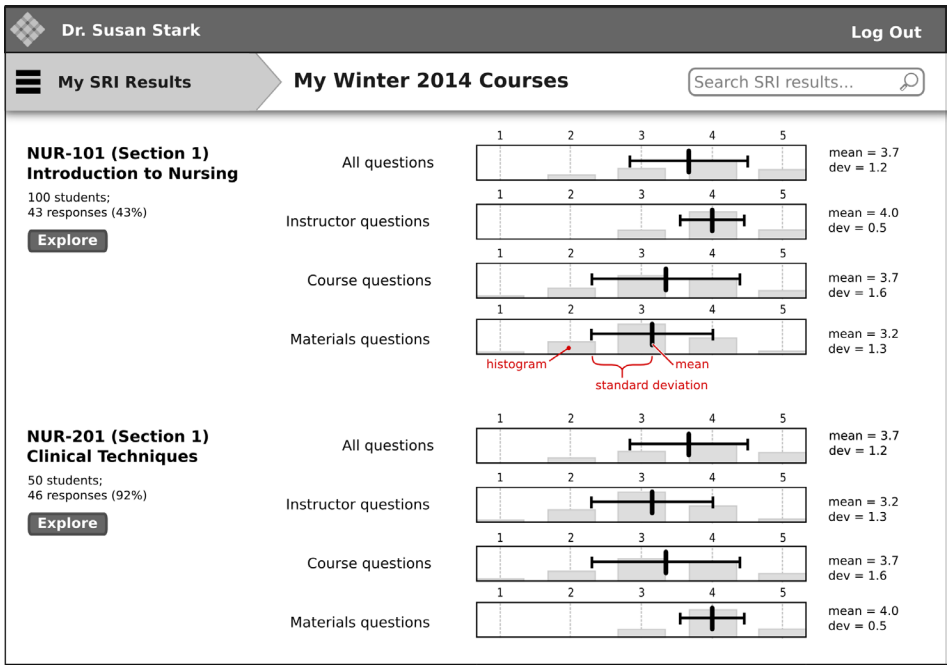


Figure 19: Drilling down into the SRI results for a single term. Effective visualizations speed up interpretation and help the instructor to discover patterns.

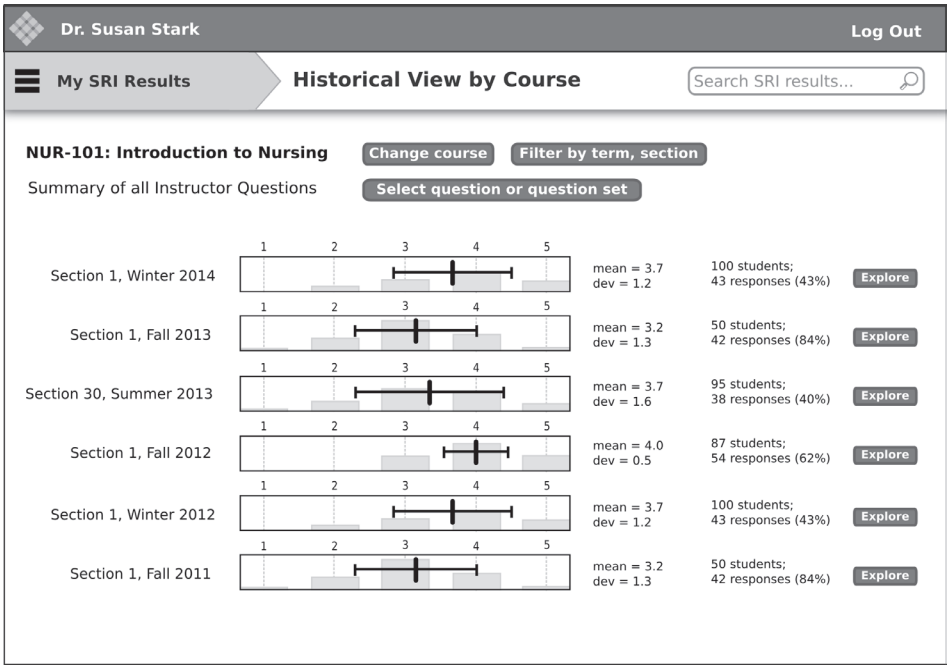


Figure 20: Reviewing SRI results over time for a single course. Effective visualizations speed up interpretation and help the instructor to discover patterns.

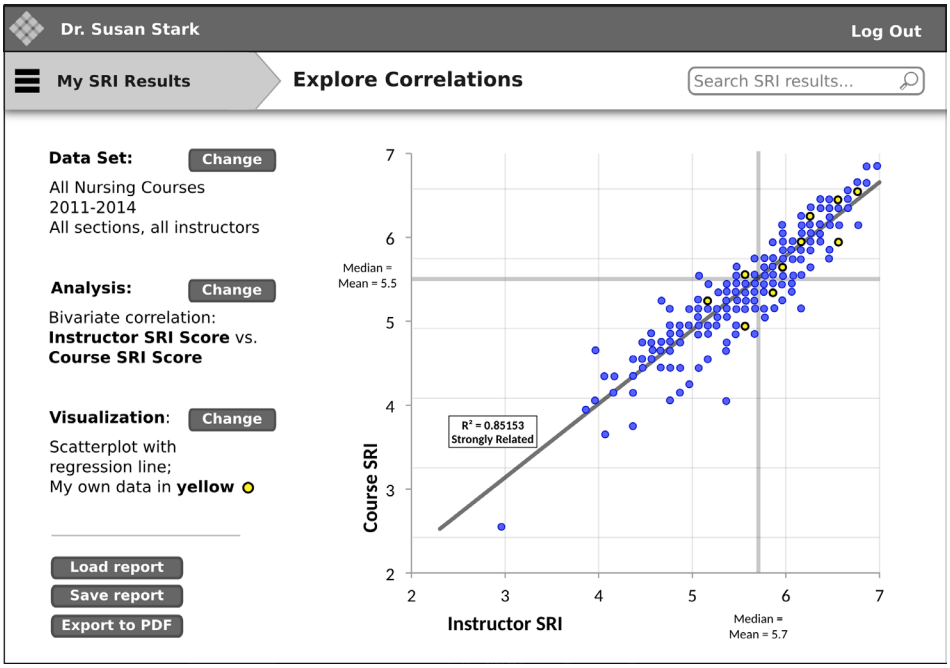


Figure 21: Exploring correlations between questions in the SRI results across a whole department. Instructors and administrators can easily select which data to explore, and can select from a number of effective visualizations such as this scatter plot.

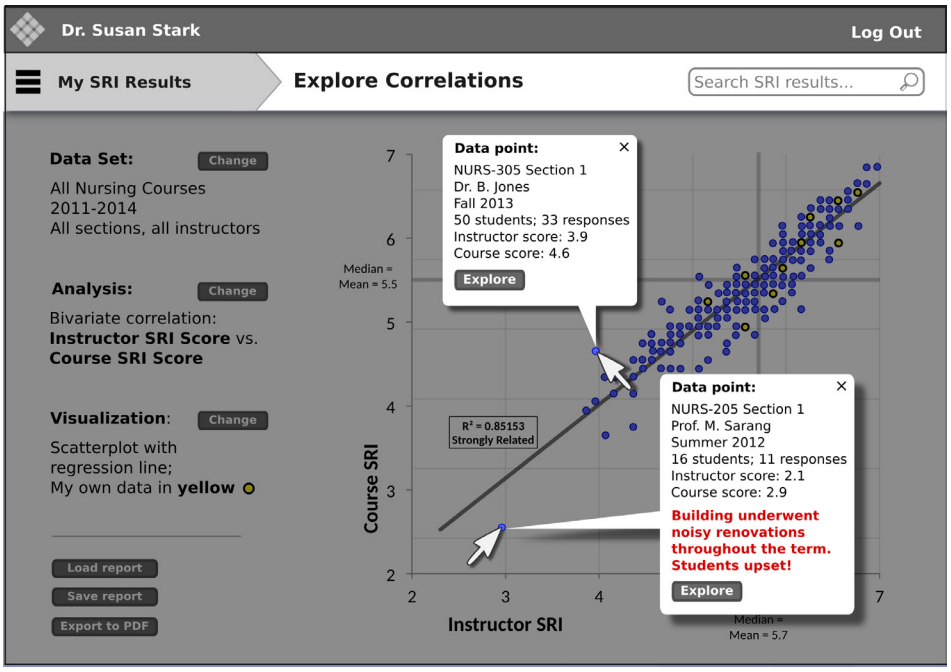


Figure 22: Interactive visualizations allow the viewer to drill down into specific data points, driving further inquiry. Data annotations, in red, allow users to note external factors that may have affected teaching and learning.

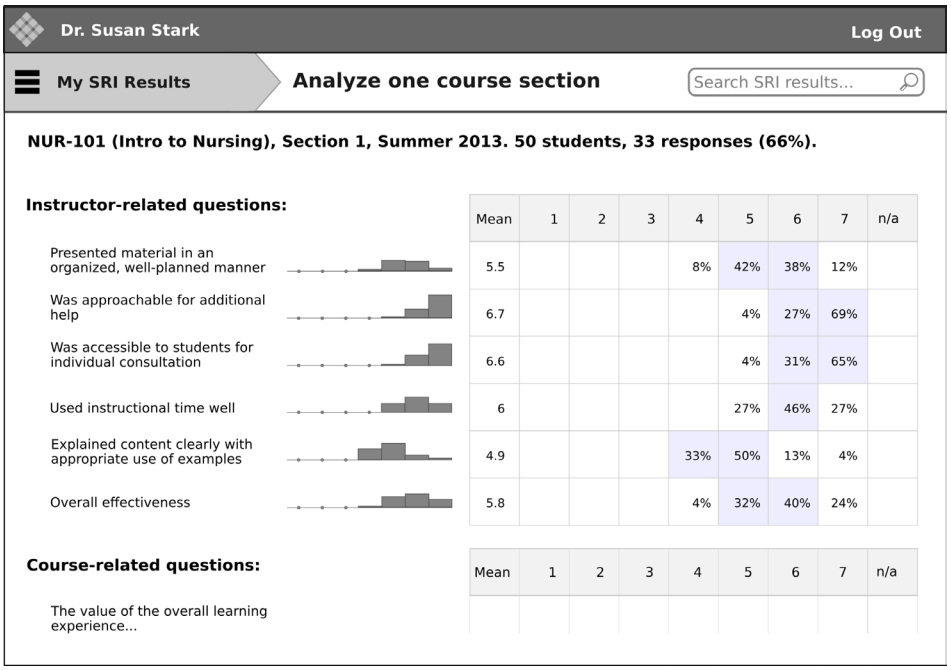


Figure 23: A question-by-question review of SRI results for a single section. Sparkline charts, in the middle, visually describe the "shape" of the response distribution. On the right, coloured cells indicate the largest response groups, helping instructors to discover modalities in the response sets.

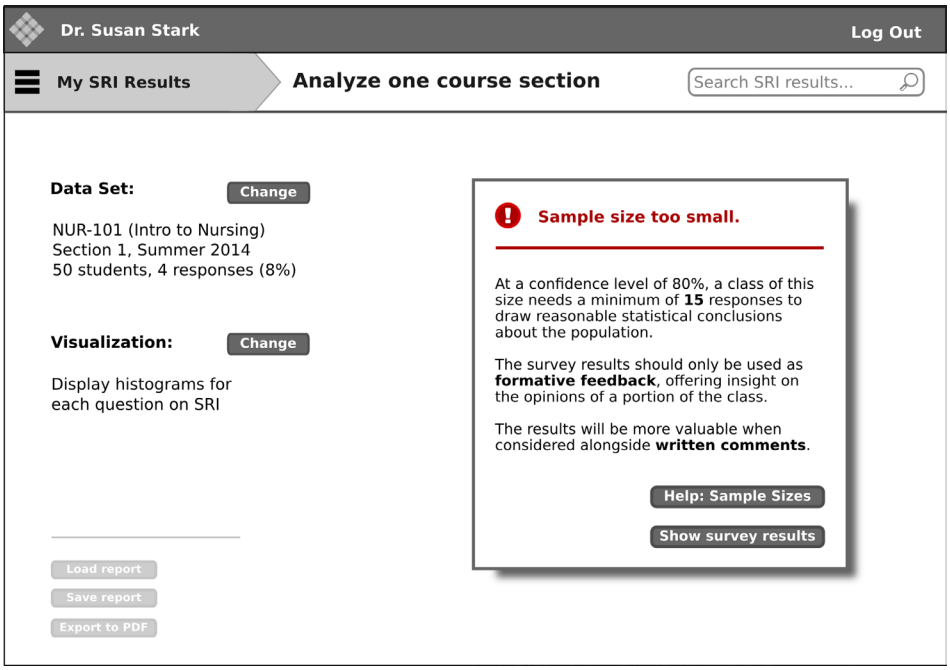


Figure 24: The SRI analysis tool will prevent users from drawing invalid conclusions: for example, when sample sizes are too small to produce a statistically significant result.

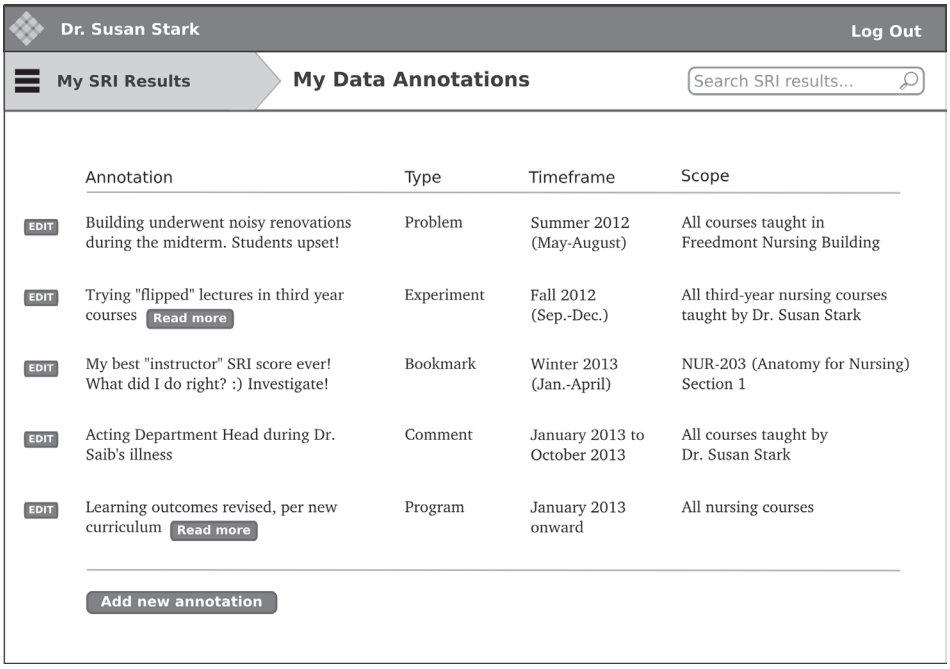


Figure 25: Data annotations. Instructors and administrators can annotate SRI data to facilitate SRI interpretation. Annotations may appear as footnotes in tabular reports, visual indicators in visualizations, or can be hidden if desired. Annotations identify events that might affect teaching and learning within a specific time-frame and scope (instructor, schedule, course, program, etc.).

10 References

Ambler, S. W., & Lines, M. (2011). *Disciplined agile delivery: An introduction* (White paper). IBM. Retrieved from http://public.dhe.ibm.com/common/ssi/ecm/en/raw14261usen/RAW14261USEN.PDF

Ambler, S. W., & Lines, M. (2012). *Disciplined agile delivery: A practitioner's guide to agile software delivery in the enterprise.* Upper Saddle River, NJ: IBM Press.

Murphy, E. M. (2011). *Mahara 1.4 Cookbook.* Birmingham, UK: Packt Publishing.

Appendix H

A Report on Third-Party
SRI Management Systems

A REPORT ON THIRD-PARTY SRI MANAGEMENT SYSTEMS

Dr. Gordon Joughin
University of Queensland, Australia

Kristin Boujos
University of Ontario Institute of Technology

Introduction and Overview

This report provides an overview of the advantages of third-party systems for the management of student responses to instruction, and the integration of these systems with related corporate systems. It includes descriptions of the services offered by third-party providers, reports the experiences of three universities that have recently or are currently going through the process of evaluating and selecting a third-party system, and one university that has retained an in-house system. It provides responses to these questions:

- Why are institutions looking to third parties for help?
- What dimensions of systems need to be considered in making an evaluation?
- What are some examples of these systems?
- What are some examples of institutional usage and what is their feedback?

A survey of Ontario universities regarding their SRI systems was conducted but produced responses from only two, one of which was using eXplorance Blue and one of which was considering this system. A small number of personnel at Canadian and Australian universities were interviewed, along with a senior manager at the IDEA organisation based in the USA.

While many institutions continue to use in-house systems or use the limited functionality of other systems (such as learning management systems), there is a range of third-party management systems on the market. One significant player in this market is the Montreal-based company eXplorance, with its Blue management system, which provides a comprehensive and effective Canadian product with very good customer service.

The following were consulted in preparing this report:

- McGill University: Laura Winer, Interim Director, Teaching and Learning Services
- University of Toronto: Marden Paul, Director, Planning, Governance and Assessment
- University of Ontario Institute of Technology: Bill Muirhead, Associate Provost, Academic
- University of Arizona: Jennifer Franklin
- Dalhousie University: Brad Wuetherick, Executive Director, Centre for Learning and Teaching
- IDEA: Sally Garvin, Client Development Specialist
- University of Queensland: Deanne Gannaway; Jon Edwards; Marcel Lavrencic
- University of Western Australia: Sid Nair

What are Third-Party Student Ratings of Instruction Systems?

A 'third-party' system for student ratings of instruction (SRI) refers to an SRI management system which is either: a) located outside the institution and provides a direct service to the institution as an external contractor; or b) has been developed outside the institution and has been purchased by, or licensed to, the institution.
An important distinction must be made between an SRI tool and a survey management tool. An SRI tool simply facilitates the process of students providing feedback to their instructor on the quality of

teaching and the course. A plethora of such tools are available, including Survey Monkey and the SRIs that are part of most learning management systems. A survey management tool, on the other hand, will support or enhance this basic SRI function while also facilitating a) integration with other core institutional data systems such as student information systems and HR systems; b) the aggregation of data that can be analysed from various perspectives; and c) the output of data into reporting systems, including institutional dashboards. Survey management systems support not only individual instructors and the evaluation of teaching, but also provide strong support for institutional and educational research for a variety of other purposes.

Why Opt for a Third-Party System?

The case *for* third-party solutions (and *against* in-house solutions) was argued strongly by several respondents:

- In-house solutions to other common corporate needs are becoming increasingly rare in the information technology (IT) and business sectors as organizations come to understand the greater efficiency and benefits of focusing on their core business mandate and outsourcing service needs that, while supporting that mandate, are not fundamental components of it.
- Developing internal applications is expensive and carries unacceptable risks, and while an institution may claim that it can best develop an application that will meet all of its specific needs, this rarely eventuates.
- In-house systems require considerable expertise and time to develop and are costly. One estimate of the initial cost to build a system for one Australian university was AUD 500,000 (about 510,000 CAD).
- Requirements can quickly outgrow the capacity of a single university.
- Where an application is successfully developed, its ongoing use and improvements depend on ongoing funding, ongoing positions, and distributed expertise. In contrast, external providers have a body of experts and the capacity to constantly update their products as client needs change and the market evolves.
- One large university had experience building its own system – with limited success.
- Reliance on key individuals or a small team for what is usually a mission critical function carries considerable risk.
- SRI processes are highly complex, involve multiple stakeholders, and serve multiple needs: acquiring the degree of expertise required to develop an effective system across these multiple functionalities is a considerable challenge. An effective third-party system offers institutions an opportunity to benefit from others' experience very cost-effectively and to work on a much shorter timeline for similar or better results.

eXplorance published a useful paper summarising key issues to be taken into account when considering a move to a third-party provider. Titled *To Build or To Buy? Seven cautions on procuring enterprise software*, this paper provides a clear overview of the issues involved, supporting the views expressed above. The paper summarises three reasons to buy and three reasons to build in-house:

Reasons to buy	**Reasons to build**
• The functionality you seek will not help you achieve any mission critical business goals. • The functionality you seek is easy to find on the market. • Your (institution) is not primarily in the business of creating software, and it does	• Your (institution) absolutely requires some unique system functionality to compete or to survive. • The solution you want is not available on the pre-packaged software market. • Your (institution) has the in-house expertise to build and maintain the

not maintain a cadre of in-house software development experts.

solution.

An In-house System: McGill University

McGill University developed its current system ten years ago. At that time they had two primary reasons for not considering a third-party option – the need for their system to integrate with their own particular student information system, and a concern for data security if data were to be stored outside the University. They also believed that with a locally developed system, buy-in from faculty and staff would be more likely. The development of their system is estimated to have taken a little under the equivalent of two full-time employees for three years, while maintenance required one to one-and-a-half full time employees. Compared to the previous paper-based system, considerable savings are made in the much less labour-intensive administration processes. The system is integrated with the student information system but not other systems, no tools are currently used for analysing comments, and the system does not link to a dashboard. While the system meets current needs, it is unclear whether the University would opt for a third-party system or an in-house system if they were faced with this decision at the present time.

Selecting a Third-Party Supplier

While there appears to be a large number of third-party systems available, the selection of a system that will be appropriate for a given institution has, in the experience of a number of institutions, quickly narrowed to a very small range. The steps typically followed in selecting a supplier include:

- Establishing the rationale for opting for a third-party solution.
- Engaging internal stakeholders in the selection process, including faculty, senior administrators, and information technology services units.
- Identifying the essential requirements or specifications. A short list of absolute essentials allows for a wider pool of potential suppliers to be considered. A longer list may ensure that the respondents to an RFP (request for proposals) are better informed and more likely to meet requirements.
- Scanning the higher education environment. This can be done through listservs such as the Australian listserv of higher education evaluation units or other existing networks, or surveys of higher education institutions similar to one's own. It can also be approached from the opposite direction, identifying providers then contacting institutional users regarding their experiences with those providers.
- Issuing an RFP, either to the world at large or to companies who seem to have the capacity to meet institutional requirements.
- Developing a shortlist and inviting demonstrations from the shortlisted companies.
- Making a final decision based on agreed criteria.

Some common specifications

Specifications are at the heart of decision-making with respect to third-party systems. The following specifications are commonly used and typically result in a rapid culling of potential providers.

- Intuitive and accessible interfaces that demonstrably enhance response rates, including mobile compatibility, easy log-on processes, and integration with learning management systems.
- Data location, whether inside the institution or hosted externally. In many jurisdictions (e.g., Australia), this is a legal requirement under privacy legislation. For many Canadian

institutions, this would also include a requirement that data remain in Canada. Many suppliers, for example IDEA, will be excluded by this requirement.

- Institutional control of questions and the capacity to include faculty-, department-, and instructor-specific questions.
- Institutional control over administrative processes, which allows for centralized coordination and distributed engagement.
- Ability to track response rates in real time.
- The capacity to handle both paper-based and online surveys. (Paper-based surveys may be required for very small classes and/or to ensure an acceptable response rate in other instances.)
- Flexible access to data – the capacity to nominate categories of personnel for access purposes.
- Analytics capacity, including the capacity to analyse text-based responses and create new reports based on emerging institutional requirements.
- Ease of integration with learning management systems, student information systems and human resource systems.
- An existing survey product rather than a stated capacity to build a product from scratch.
- A sound, established business with a track record in similar contexts.
- Flexibility and willingness to adapt to client needs.
- Quality and accessibility of support.
- Cost.

Some conclusions about selecting a third-party provider

- While there many generic survey and education-specific survey tools that can handle individual instructor and course SRIs, tools that allow for the aggregation of data, integration with other institutional information systems, and detailed analysis are limited in number. An Australian review found only two viable options for that particular institution, eXplorance Blue and Evasys. A Canadian review examined EXplorance, EvaluationKit, Campus Labs, and CCI (see Table 1).
- No single system in its current form will meet all of an institution's requirements without modifications or further development. The University of Toronto's experience with Blue suggests that eXplorance has been willing and able to adapt its system to their requirements.
- While the selection of the most appropriate system is crucial, equally important is the allocation of internal resources to manage the implementation of the new system, to oversee its ongoing use, and to manage changes as needed.

The section that follows provides information about the characteristics of various potential systems as well as illustrations of decision-making processes institutions have employed in assessing systems.

An Evaluation of Third-Party Systems for the Canadian Context: The UOIT Process

A recent review of possible third-party systems at the University of Ontario Institute of Technology (UOIT) provides a representative model of how evaluation can be approached as well as a comparative summary of three providers' services and characteristics (Table 1). UOIT identified drivers for change and specific requirements in order to create a criteria matrix for decision-making purposes. They are still in the final decision making stages of their process.

Table 1: University of Ontario Institute of Technology SRI Decision-Making Matrix

Product Ranking		1st	2nd	3rd	Current
Criteria		EXplorance	EvaluationKIT	Campus Labs	CCI
1	Server location is in Canada.	Yes	No	Yes	Yes
2	Stores the data in a hosted environment.	Yes	Yes	No	Yes
3	Tool has been implemented at Canadian institutions.	Yes	No	No	Yes
4	Provides mobile compatibility.	Yes	No	Yes	No
5	Provides an App for students to download and access their course evaluations.	No	Yes	No	No
6	Annual cost to administer and manage course evaluations and institutional surveys. (Thousands)	$25.0	$17.5	$35.0	$16.0
7	Initial cost to implement the tool(s) (Thousands)	$96.0	$17.5	$43.0	$16.0
8	Provides secure access to the course evaluations for students.	Yes	Yes	Yes	Yes
9	Offers multiple student access points. (Log in page, LMS, email links)	Yes	Yes	Yes	No
10	Provides access to the course evaluations through a web browser. (Internet Explorer, Chrome, Safari, etc.)	Yes	Yes	Yes	Yes
11	Allows access to the course evaluations through Blackboard.	Yes	Yes	Yes	No
12	Integrates with other Learning Management Systems	Yes	Yes	Yes	No
13	Allows pop-up student reminders in Blackboard to complete outstanding evaluations.	No	Yes	No	No
14	Integrates with Banner to send and receive information.	Yes	Yes	Yes	No
15	Allows exporting of data in different formats (PDF, .cvs, excel)	Yes	Yes	Yes	No
16	User friendly software - easy to implement, maintain and administer within UOIT.	Yes	Yes	Yes	N/A
17	Allows instructors to monitor the course evaluation response rate in real time without requesting it from a third party.	Yes	Yes	Yes	No
18	Allows for modifications or revisions to course evaluation enrolment information.	Yes	Yes	Yes	No
19	Includes advanced reporting features to allow for longitudinal analysis and queries.	Yes	Yes	Yes	No
20	Create ad hoc reports as required.	Yes	Yes	Yes	No
21	Own and manipulate the data at any time by authorized users at UOIT.	Yes	Yes	Yes	No
22	Support available from vendor.	$	Yes	Yes	No
23	Additional services to use the software for institutional surveys.	$	Yes	$	$
24	Ability to have an exit strategy and export all the UOIT data for storage and/or migration to another solution in the future.	Yes	Yes	Yes	No

All vendors offered secure web-based access to surveys. Prices vary: initial start-up costs for an institution the size of UOIT was between 16,000-96,000 with an annual licensing fee of between $16,000 and $35,000 per year. Criteria met by 3 of 4 providers included:

- hosting on Canadian servers;
- multiple ways for students to access surveys;
- LMS integration;
- ability to export data in different formats;
- user-friendly software on the administrative side;
- real-time monitoring of response rates;
- advanced reporting features including querying and ad hoc report development; and
- an exit strategy for institutional data migration should the contract be terminated.

This review indicates the range of considerations that must be taken into account and the existence of strong providers in the Canadian context. It should be noted, however, that while Campus Labs indicated the ability to provide these services, they do not currently provide them at any Canadian institutions, while eXplorance Blue has a growing Canadian portfolio, which currently includes the University of Toronto, Durham College, and Mohawk.

Adoption of eXplorance Blue: University of Toronto

The adoption of EXplorance Blue at the University of Toronto followed an extensive planning and decision-making process. A Vice-Provostial Committee was formed to consider the wide range of practices across the University and to establish goals for evaluations of teaching at the university. Specifically, they identified the development of a more standardised approach across the university, with a common set of empirically based questions, as well as the capacity for additional questions at the levels of divisions, departments and individual instructors. They also required stipulated limitations on who could access what data. The University undertook extensive research in course evaluation processes worldwide to identify effective practices to adopt.

The decision to seek an external provider rather than develop a system in-house was based on a number of considerations, including: the known difficulties of building in-house systems based on prior experience; the time and the expertise that would be required; the difficulties of managing and resourcing ongoing development over time; and the recognition that the market could provide a proven, viable product. The initial RFP drew 27 expressions of interest with 6-7 companies submitting proposals. Eventually only two companies were considered to have viable solutions based on the University's established criteria. Several companies had no experience with evaluation systems but claimed they could build one.

Ultimately, eXplorance Blue was viewed as being 'head over heels' above every other proposal. Its Montreal location was considered to be a major advantage. The University has been pleased with Blue's performance and have not experienced any significant difficulties or limitations in its use: however, while impressed with the support provided by eXplorance, the importance of the local University team in implementing the system, both initially and in ongoing use, was emphasized.

Adoption of Evasys: University of Queensland, Australia

In 2010, the University of Queensland moved from the tri-annual teacher-determined use of SRIs to a system in which each instance of the course would be evaluated, resulting in a significant increase in the scope and magnitude of data collection and processing at the University, without growth in the resources available to accomplish the task. Pressures on the system highlighted a number of

challenges and risk factors. The system was not integrated, meaning different tools and people were responsible for preparing for SRI, data capturing, analysis, and reporting. The system suffered from inefficiencies related to manual paper handling and to the use of a secondary system for online SRI with an entirely different set of business processes. The database housing SRI data was reviewed and deemed to need re-design. The SRI systems were not integrated with the University's other corporate systems, resulting in a process that was resource intensive, inefficient, and vulnerable to failure. In order to address these challenges, the University sought a complete survey management system that would incorporate a database, detailed reporting, and survey administration. They were seeking capacity for broad data gathering and analysis on multiple fronts, not a simple online survey tool.

In 2013, the University undertook an exhaustive evaluation of third-party SRI systems with the goal of implementing an improved institutional management system for SRI data. This included: a detailed analysis of integration requirements and data access needs, the development of a business case, and an investigation of possible solutions. The University assessed the marketplace, current practice in the Australian university sector, and internal considerations. They evaluated possible solutions and identified costs and timelines. Only three vendors, eXplorance Blue, EvaSys, and CATEI were deemed to meet the University's core requirements. These were then evaluated based on their reporting capacity, their extensibility, capacity for LMS integration, flexibility, delivery models, and capacity for system integration. While Blue was considered superior, Evasys met requirements and could be accommodated within a limited budget. Process analysis determined a possible estimated savings of approximately $90,000 from the implementation of the new system. More critically, however, the new system would enable the redeployment of current unit staff to focus more fully on high-level analysis, extension of survey offerings at the institution, and consultation and support of quality assurance and teaching inquiry processes. The new system will also allow for greater SRI customization by individual instructors and programs, and improve the accuracy and reporting of data.

IDEA

IDEA Education is a non-profit organisation based in Boston and focused on student ratings of instruction since 1975. They also provide feedback systems for administrators and chairs. They seem to be a highly regarded organization whose work has a strong research base. While it is unlikely that IDEA could become a provider for Canadian universities and colleges, its approach is described in some detail as it is highly professional and informative. Their work has a number of aspects not commonly found in institutional SRI systems and from which some important lessons can be learnt.

IDEA resources

The IDEA web site (http://ideaedu.org/) offers a wealth of useful resources on SRI in general, their own approach, and sample forms and reports. The following are noted in particular and provide an excellent overview of the IDEA approach:

- *IDEA Paper #50: Student ratings of teaching: a summary of research and literature*. An important feature of the IDEA approach is the extent to which the survey instruments are informed by the research and literature.
- The three survey forms discussed below (http://ideaedu.org/services/student-ratings/sample-forms-student-ratings-instruction)
    - Sample Faculty Information Form
    - Sample Diagnostic Survey Form
    - Sample Short Survey Form
- The sample reports: http://ideaedu.org/services/student-ratings/sample-forms-student-ratings-instruction

- Diagnostic Form Report
- Short Form Report
- Group Summary Reports
- Benchmarking Trend Report
- Benchmarking One-Year Report
- Benchmarking Discipline Report

The IDEA survey forms

The IDEA survey system includes three separate questionnaires:

Faculty Information Form – This form is completed by the instructor who provides information on:

i. Objectives (e.g., 'gaining factual knowledge'; 'learning to *apply* course material'; 'learning to *analyse* and *critically evaluate* ideas, arguments, and points of view'
ii. Contextual questions, which IDEA use for research purposes and to improve their interpretation of student ratings. These include (a) 'the primary approach to this courses ' (e.g., lecture; discussion; field experience); (b) course requirements with respect to features such as writing, oral communication, critical thinking, memorization; (c) whether a number of circumstances were considered to have a particular impact (positive, negative, neutral) on learning; and (d) the principal type of student taking the course (e.g., first year/sophomore seeking background for a specialisation; first year sophomore seeking to meet a general education requirement.
iii. The discipline, using a discipline code based on Classification of Instructional Program codes.

Student Reactions to Instruction and Courses Form – This survey form has seven sections:

i. The instructor: Twenty standard questions on the frequency of the instructor's teaching procedures (e.g., 'displayed a personal interest in students and their learning; 'related course material to real life situations'). These questions are fixed by IDEA and cannot be modified.
ii. Progress: Twelve questions on the learning objectives listed on the 'Faculty Information Form'. The student responds only to those, which apply to the course. These questions are fixed by IDEA and cannot be modified.
iii. The course: Three questions comparing the course to others taken at the institution in terms of amount of reading, amount of work and difficulty of subject matter.
iv. Seven questions addressing motivation (e.g., 'I had a strong desire to take this course') and overall ratings of the course ('Overall, I rate this course as excellent') and instructor ('Overall, I rate this instructor an excellent teacher').
v. Five questions on student effort, instructor standards and use of teaching methods and technologies.
vi. Up to ten questions created by the instructor.
vii. A field for free comments.
    - The survey includes a minimum of 47 questions plus up to ten instructor created questions.

Short Form – Student Reactions to Instruction and Courses: This contains the 12 questions from section ii and five of the questions from section v of the Student Reactions to Instruction and Courses form plus one additional question.

Key features of the IDEA approach

- The research basis of the questions asked.

- The inclusion of contextual and motivational questions to allow for adjustments of ratings based on variables known to affect student responses but over which the instructor has no control.
- The focus of the questions on student learning.
- The reporting of teaching effectiveness in relation to three aspects: teacher excellence; course excellence; and progress on objectives.
- The reporting of raw scores and 'adjusted scores' (based on ii above).
- The use of 'converted averages' that "take into account the fact that average ratings for items on the IDEA form are not equal" – converted scores all have an average of 50 and a standard deviation of 10.
- Detailed reports on teaching methods and styles and suggestions for improvement.
- Comparison of instructor scores with national discipline means.
- Trend reports for instructors, departments, whole institutions or other subsets.
- Controlled access to reports as determined by each institution.
- IDEA keeps all data to allow for comparisons with national responses. All data is also provided to the institution for their own use.

The IDEA approach is detailed, research-based, highly professional, applies important statistical techniques that are typically ignored elsewhere, and is well documented. A particular strength is the use of 'adjusted' and 'converted' scores. A further strength, the use of fixed, research-based standardised questions, is likely to be seen as an unacceptable restriction by many institutions. IDEA's retention of data may raise privacy issues that would preclude its use by Canadian institutions.

The Experience of Working With Third-Party Providers

Third-party providers are (usually) business ventures needing to optimise profit. Leading suppliers are at the cutting edge of the field, constantly refining their programs and developing new services. Respondents for this report made the following general observations:

- University or college personnel dealing with the vendors need business acumen.
- Some vendors may have a tendency to over-promise, stating that they can meet all of their potential client's needs. While the leading programs seem to have the capacity to meet most of the needs of most institutions, additional work is often needed to ensure that all needs are met. Accurate specifications are needed, and all verbal agreements should be confirmed in writing. Leading vendors should be expected to work with their clients to develop the capacity to meet specific needs – such needs are likely to be shared by other current clients or future clients.
- It is important to factor institution-side costs into the planning, customization, and implementation process.
- Some leading vendors are expanding rapidly, leading to potential delays in responding to institutional requests.
- Time zones are important. With suppliers and client support located around the world, there are significant advantages in having 'local' support. Extreme time zone differences can result in up to 24 hours delay in responding to problems, which may need an immediate solution.

The specific experience of working with third-party providers in implementing their systems is limited in this report to three universities with recent experience: University of Toronto, University of Western Australia, and University of Queensland, and is limited to two suppliers, eXplorance and Evasys. In summary:

- No university reported significant problems at the implementation stage. Suppliers provided support as promised and the systems worked as expected, but it should be noted that the implementation of these systems is not simple and requires a significant but manageable amount of joint problem solving as the system is integrated into each university.
- The University of Toronto described an excellent level of support from eXplorance – the company was keen to incorporate required changes/developments into its core product (knowing these would benefit other clients), provided support that exceeded contractual obligations, responded rapidly to requests, and readily sent staff to the University when required.
- The University of Western Australia implementation went smoothly and the system is working well. However some significant delays in support were experienced. This is believed to be a result of the rapid expansion of the eXplorance company rather than a lack of willingness to help.
- The implementation of Evasys at The University of Queensland has proceeded without difficulty. The Evasys consultant spent two weeks at the University assisting with integration into other systems. Support is based in Europe, leading to some delay in communication. This can be a significant issue if a problem arises at the point at which surveys are about to be administered. In survey matters, rapid solutions to unexpected problems are often needed, making local support particularly valuable.
- The propensity of some companies to 'over-promise' should be kept in mind. All verbal agreements should be confirmed in writing.

Conclusion

For an individual institution considering a new SRI management system, the case for a third-party solution is strong. For Canadian universities and colleges, the availability of a Canadian product, eXplorance Blue, which is rapidly establishing itself as one of the leading international products, makes this case even stronger. As the eXplorance report to St. Mary's College (CA) about their product puts it, "The essential element of the Blue architecture is that it allows for the scheduled synchronization of staff/instructors/course/student demographics and relationships directly from your information systems...This information can also be imported directly into a hosted environment." Blue's features include very strong reporting features, support for both online and paper and pencil implementation, full AODA compliance, real time response rate monitoring, access across multiple kinds of devices, strong LMS integration, and easy management of administrative structures from within institutions. Although eXplorance Blue is a strong presence in the market, any determinations regarding the selection of third-party providers must inevitably follow a more systematic, needs-based process in order to determine suitable future directions.

This report has focused on the needs of individual institutions, has noted the limitations of an individual institution in developing and maintaining a system, and highlighted the successful adoption of third-party systems in several institutions. The paper has not considered the possibility of a consortium of universities developing a common system, or of a consortium entering into partnership with an existing provider: both of these require a degree of collaboration that would require considerable contextual change if a common instrument is intended.

Documents

Extensive documentation is available on the web sites of the three providers highlighted in this report. While each provider is motivated by commercial considerations, the sites provide

informative papers on SRI and SRI management systems as well as information about the specific services provided.

- eXplorance. http://www.explorance.com/
- Evasys. http://www.evasys.co.uk/start.html
- IDEA http://ideaedu.org/

A further document, the St. Mary's College (CA) proposal to adopt an electronic course evaluation system, may also be of use, as it provides an overview of the characteristics of 11 third-party SRIs, as well as eXplorance's compact summary of Blue's functionality and the services provided by the company. St Mary's College of California (June 2012):
http://www.stmarys-ca.edu/sites/default/files/attachments/files/Proposal%20to%2012sep6%20Adopt%20an%20Electronic%20Course%20Evaluation%20System.pdf

eXplorance, n.d. 'To Build or To Buy? Seven cautions on procuring enterprise software' http://www.explorance.com/resources/whitepapers/attachment/build_or_buy_enterprise_feedback_management_software_system/

In addition to the above resources extensive documentation from the University of Queensland's selection and implementation of Evasys has been provided to the Technical Team.

Appendix I

SRI Data Aggregation and Visualization: An Evaluation of Potential Uses

SRI Data Aggregation and Visualization: An Evaluation of Potential Uses

Phil A. Graniero, Beverley Hamilton, & Ken Cramer
June 2014

The authors would like to acknowledge the contributions of the following people to this report:

Graham Fawcett
Peter Marval
Ashlyne O'Neil
Lisa Plant
Bethany Paper
Jessica Raffoul
Behdokht Rekabdar

For more information, please contact Phil Graniero at graniero@uwindsor.ca

Introduction

This report provides an overview of some of the ways that aggregate student ratings of instruction (SRI) data, combined with data visualization tools, can significantly improve the use of the data that we already collect, in order to enhance institutional decision-making, improve instructional practice, and inform program assessment and management.

Student ratings of instruction are one of the many kinds of evidence that should inform teaching evaluation practice. That said, we can accomplish much more with the information that SRIs provide, including the following:

- We can use SRI data to enable effective individual inquiry into student perceptions of teaching and courses. Ludlow (1996, 2005) demonstrates that this approach can have positive effects on instructors' engagement with SRI data, their willingness to make changes to practice, and their capacity to effectively represent their work for promotion, tenure, and performance review.
- We can use SRI data in more varied and comprehensive ways to draw out patterns and commonalities that might inform decision-making (Office of the Auditor General, 2012), enable more targeted instructional development intervention, and potentially inform curriculum and course re-design efforts (Joughin & Winer, 2014).
- We can use aggregate data to parse variations and commonalities in response patterns in order to identify norm groups, address bias, and inform decision (Joughin & Boujos, Appendix H).

In order to maximize the potential of SRI data, we must provide faculty and administrators with better tools and more expertise, so that they can understand SRI data more easily, effectively, and carefully.

There are many misconceptions about the nature, potential, and limitations of SRI data. A review of the literature has shown significant variations regarding the use of various terms. This has led to sometimes unintentional misuse of data to produce reports that are not methodologically sound, potentially leading to incorrect decisions and an overall lack of confidence in SRI systems. Appendix A provides an overview of the most relevant terms, principles, and challenges in SRI data analysis, also summarized in the body of the report. After a brief contextualizing section, this report explores a variety of examples of visualized aggregate data use in order to assess its potential and recommends initial approaches to developing and implementing tools for aggregate SRI data use in the Ontario context.

Method

To date, research on SRI data aggregation has been limited, but some institutions and companies provide models for study. Our approach has been informed by some of those examples, but has been primarily exploratory. The initial phase of the research was undertaken by a psychology professor with expertise in psychometrics who, with a team, explored possible approaches to SRI data aggregation for instructors and programs using Excel spreadsheets and pivot tables. This work was then reviewed and extended by individuals with expertise in data management and programming who, in collaboration with the project lead, developed a variety of prototype tools that make it possible to pull data from standard spreadsheets to produce more extended visualizations. The team conducted a brief literature review and developed preliminary ethical guidelines for aggregate SRI data use, intended as a starting point for proposed Phase II discussions with partner teams. They also explored technical approaches to limiting viewers' ability to produce inaccurate or misleading visualizations, as well as visualizations based on insufficient data. Based on feedback from the lead team, the approach was adjusted to ensure that it would fit with current institutional cultures and sensitivities.

Aggregation Methods

For the type of applications discussed in this report, the results from all respondents to a single SRI survey (i.e., the results from all students in a single course offering) are aggregated, forming the basic unit of analysis. Aggregated items, scales, and scores are used, not individual responses. These course-level survey results are then aggregated in various ways: aggregated in the sense of using a collection or selected group of surveys together; and aggregated in the sense of representing specific survey items in that collection in different ways.

Ways to select surveys for a collection include:

- all offerings of a particular course;
- all courses taught by a particular instructor;
- all courses from a particular department;
- all courses from a particular faculty;
- all courses taught in a particular semester; and
- all courses taught in a particular year or time range (e.g., past ten years).

Collections may be created by selecting surveys using more than one of these characteristics, either by limiting the collection to results that match all criteria, or by creating sub-groups within a collection. For example, one might be interested in examining all offerings of a particular course taught by a particular instructor, a collection limited by instructor identity and specific course. Another example would be collecting all scores for a specific department, over ten years, grouped by course. In this case the collection is limited by department and time range and grouped by course identity.

Ways to represent an item or scale in an aggregate collection include:

- tabular, statistical, or visual representations of a response distribution;
- a summary number representing the distribution of scores;
- tabular, statistical, or visual representations of the distribution of scores;
- a time series of scores (scores placed in chronological order); and
- a single response distribution aggregating the response distributions for an item from all surveys.

Aggregating Several Items Into a Single Scale

A *scale* is the collective responses from a set of items with identical response formats, where the set of items are designed to make up a reasoned, cohesive, and complete examination of a single multi-faceted subject such as overall

Terminology Review

A *scale* is the combination of several *items*, usually by summing the items.

The *response distribution* is a count of the number of respondents who chose each response category for an item or a scale.

Terminology Review: Four Primary Levels of Measurement

Nominal-Level Data

Nominal-level measurements are distinguished by their membership in a qualitative, named category: there is no meaningful ordering of the categories based on rank or magnitude.

Meaningful mathematical operations: related to set membership, e.g. sizes of groups. **Meaningful measure of central tendency**: mode, or most common value.

Ordinal-Level Data

Ordinal-level measurements are distinguished by their membership in a qualitative category that may be sorted meaningfully by a rank ordering. It is meaningful to compare a category value as *bigger* or *better* than another, but impossible to state the *degree of difference* between them.

Meaningful mathematical operations: related to set membership, e.g. sizes of groups.
Meaningful measure of central tendency: mode and also *median*, or the category of the middle-ranked item in an ordered set of values.

Interval-Level Data

Interval-level measurements are quantitative or numeric measures that distinguish the *difference* and *relative direction* among items; one unit of difference between two items means the same thing regardless of the size or magnitude of the items. However, the *ratio* between two items is not allowed since the zero point is arbitrarily chosen from more than one possible value.

Meaningful mathematical operations: Addition and subtraction
Meaningful measure of central tendency: *Mode* (if integers), *median*, and *arithmetic mean* or *average*. The spread or dispersion of the distribution can be described using the *range* (the difference between the maximum value and the minimum value) and the *standard deviation* when it is statistically valid, for example when the set fits a *normal* ('bell-shaped') distribution.

Ratio-Level Data

Ratio-level measurements are quantitative or numeric measures that distinguish the *difference* and *relative direction* between items, similar to interval-level measurements, but they also have a unique and meaningful zero value so the *ratio* between two items is allowed.

Meaningful mathematical operations: All
Meaningful measure of central tendency: *mode*, *median*, and *arithmetic mean*. *Range* and *standard deviation*, as well as other statistical measures such as the *coefficient of variation,* when valid and appropriate.

instructor effectiveness. Also, a subset of the items that are included in a scale may be used to form a *subscale*, which represents one component or dimension of the overall scale such as presentation skill, organization, or approachability. When looking at SRI data aggregated for an entire class instead of an individual response, the scale is a *response distribution*, but it is typically collapsed to a *score*, a single summary of the distribution (discussed more below). Correspondingly, a subscale would become a *subscore*. In the rest of this report, scale and score, and subscale and subscore, are used interchangeably to conform to usage elsewhere in practice.

Scales are a highly useful aggregation method. Not only is the practice sound from the perspective of instrument design and statistical validity, it is helpful to get people to read and use SRI results. The large volume of numerical information from a list of many detailed items can be overwhelming, and grouping the items into scales or sub-scales means the instructors and administrators receive fewer category scores, which facilitates evaluation (Algozzine et al., 2004).

Aggregating the Response Distribution of an Individual Item Into a Single Summary Score

The response distribution for an aggregated item or aggregated scale represents the range and variation of opinion within the class. The distribution is usually represented by a single summary number (a *score*) to provide a simpler, more easily interpreted measurement.

The most common summary description is the *central tendency* of the distribution. The appropriate method for reporting it depends upon the level of measurement for the item. If the item uses interval-

level or ratio-level measurements, then either the median or the arithmetic mean, or average, can be used. If the item uses ordinal-level measurements, then only the median can be used. The variation of a distribution may also be of interest. If the item uses interval-level or ratio-level data, then the most common measure is the standard deviation. If the item uses ordinal-level data, then quantiles or percentiles may be used.

Likert and Likert-type items often use integers as the response encoding for each category. It is common practice to calculate the mean and standard deviation of these distributions. However, this is not mathematically valid. Likert-like items are *ordinal-level* measurements, not *interval-level*. The integers are only substitutes for the qualitative labels. The calculation of the mean and standard deviation, and the mathematics underlying them, fundamentally rely on a uniform degree of difference between adjacent values. It is rarely defensible that the 'distance' or 'number of units' between values is equal – that is, the distances between *poor* and *adequate*, *adequate* and *good*, and *good* and *outstanding* are all the same – especially because the meanings and differences among the choices are judged by each student according to their own personal standards, experiences, and frames of reference.

Although the practice is not strictly valid from a mathematical standpoint, Carifio & Perla (2007) point to statistically grounded empirical research that shows that many of the commonly used tests are sufficiently robust that the results may be used with the recognition that assumptions are being bent and the results do not have the same strength as they otherwise would. Hativa (2013a) also notes that, technically, ordinal-level data should not provide means or averages as measures of central values. She argues, however, that means, averages, factor analyses and regressions are widely used in the social sciences for ordinal scales, and that many well-established and widely used survey systems like SEEQ do employ these approaches, concluding that "the policy is widely agreed upon in practice" (Hativa, 2013a, p. 64).

This must always be kept in mind wherever SRI scores are reported or interpreted: **It is not strictly valid to do statistical tests on means of Likert-type scales and scores, but these methods may be used with caution. The exact value of such scores, the differences between two scores, and summary statistics describing groups of scores, are all less precise than they appear. Likewise, the degree of certainty for statistical tests conducted on these scores will be less than what is reported.**

Comparisons: Statistical Tests of Similarity

The purpose of an SRI instrument is to measure the range and variation of opinion of students in a course with respect to instructor or course effectiveness. It is not possible to determine one single value that completely represents every student's rating: it is expected that students will have a range of experiences in the course, and that not everyone will arrive at the same rating. It is therefore important to represent the class' response as a *distribution* of values. This makes it difficult to make comparisons among different groups or course instantiations to determine whether the distributions representing them are the same or different: there will always be some degree of overlap, and some degree of divergence. The challenge is to decide if there is a *sufficient degree* of overlap to decide that the groups are effectively the same. There are a number of statistical tools that can be employed: each must be used with a good understanding of its particular function and requirements. Otherwise it may be used in situations where it is not the appropriate tool, resulting in inaccurate conclusions. Appendix A provides detail regarding a variety of tests used to make statistical comparisons, and the conditions under which those tests can properly be used.

When comparing one collection of scores to another, it is crucial to remember that summary SRI scores were originally derived from ordinal data, not interval data, and therefore the values and their

differences are not as precise as they seem. The results cannot be used for hard 'scientific proof', only as indications of whether differences likely exist. They cannot be used to definitively and unarguably state that an instructor has exceeded a threshold, but rather for guidance regarding which cases easily pass the threshold and which cases near the boundary need more careful examination.

Determining what can, and cannot, be compared and how to make methodologically sound comparisons is probably one of the most difficult aspects of creating effective SRI reporting and facilitating decision-making informed by SRIs. The reality is that most of the users of SRI data do not have sufficient expertise to make statistical design decisions, to identify elements in a set of data that do not provide a valid basis for conclusions, or to determine if the data are representing differences that are not just numerically different, but statistically significant (Hativa, 2013b; Winer et al., 2012).

Even with sufficient statistical design expertise, it can still be difficult to design an analysis that does not violate the varied assumptions that statistical tools rely upon. The typical practicalities involved in administering SRIs in a course tend to violate many assumptions regarding representative samples. Furthermore, comparative analyses rarely have a sufficient number of data points due to the wide number of factors impacting the similarity – and therefore comparability – of courses and offerings.

Using Numerical Data in Contexts of Uncertainty

Statistical tools for making comparisons and other analyses can be incredibly powerful for making decisions in a complex, heterogeneous world, but only if the measurements are made in ways that are accurate, sufficiently representative of the population being examined, and both mathematically and statistically compliant with the tools' calculation methods and assumptions about the world that the measurements describe. There are many inherent obstacles to designing and administering SRI instruments that meet these requirements, as have been discussed above. In summary:

- Students interpret the subjective questions and define or understand the available qualitative response categories in different ways, which introduces **measurement error**.
- Only a portion - and often too small a portion - of the class responds to the survey, so there are questions of how well the sampled group's responses represent the class as a whole, which introduces **sampling error**.
- Many statistical tests assume a random sample, but typically classroom response to an SRI constitutes a **convenience sample**. This violates assumptions about the sample used in statistical analysis.
- **Scales may involve too few items** to allow for statistical calculation with certainty.
- A variety of factors appear to impact the validity of student ratings of instruction. A number of differences between respondents and non-respondents have in fact been noted (Goyder 1987; Richardson 2005), in particular for students, in their attitudes and behaviour (Goyder 1987), and in their study behaviour and academic attainment (Astin, 1970; Neilsen et al., 1978; Watkins & Hattie, 1985). **Non-random samples may be biased by differences in the individual characteristics of students**, such as disciplinary differences in response patterns, gender differences in response patterns, and student year might affect the representativeness of a given sample (Hativa, 2013b). **Comparison among different courses may be affected by differences among the courses that impact ratings, such as level, class-size, or delivery mode**. Issues of bias in student ratings of instruction, however, are hotly debated, with considerable evidence on each side of the debate (Hativa, 2013b). One critical element of establishing the validity and reliability of SRI is establishing regular analysis at each institution of context-specific data to identify or disconfirm theories about bias within student responses (Winer et al., 2012; Joughin & Winer, 2014), a practice that is not widely employed.

- SRI instruments rely heavily on the use of **Likert and Likert-type items, which produce ordinal-level data, but the responses are used as interval-level data even though the degree of difference between response choices is not uniform**, and is very difficult to characterize. Most of the calculation and statistical methods we use to describe and make comparisons between distributions rely entirely on having uniform differences between values.
- **Statistical tools must be used in ways consistent with their particular functions and requirements, a practice, which is not always ensured**. Some tests, for example, assume a normal distribution of scores, which is certainly not the case for many SRI distributions. Other tests should not be applied to small populations.
- **Statistical measures are frequently provided without an indication of whether they are significant**, or whether the population was sufficient to justify the calculation involved.
- The use of **statistical measures of central tendency** (means, medians, and modes), even when valid for the level of measurement involved, **can mask important information about score distributions**: a bimodal distribution with many students at each extreme can produce the same mean score as a tight clump of scores around the middle, but these distributions have significantly different implications, both for decision-making and for instructional improvement.

Essentially, in typical SRI implementation and analysis, necessary mathematical and statistical requirements of the tools are not met, which means that results are simply not as accurate as our faith in statistics tends to lead us to believe. Nor are they as accurate as the statistical measurements of accuracy tell us they are. Given these challenges, the results of statistical analyses of SRI data must be used with informed caution, and with an understanding that their results are not as precise, accurate, or certain as they are in other fields or applications where the requirements of statistical calculation are more easily met. It doesn't mean that the results are unusable and any analysis is pointless. Rather, the results should be used as signposts to broader patterns, trends, or potential differences: persuasive, not conclusive, evidence. Unfortunately, their appearance of numerical precision can be beguiling.

While the guidance offered by statistical information can be helpful if used appropriately, other visual tools can significantly enhance our understanding of what the data and accompanying analyses tell us. For these reasons, we have adopted an approach to tool design that uses visualization methods to document the nature of the patterns within the data and to display statistical measures in a more accessible fashion. In many respects, the visualizations present a complementary view of what the statistics articulate with numbers (which in many cases are less easily comprehended). While, for example, SRI data reporting may provide the reader with a mean and standard deviation, actual study of the scatterplot that those numbers represent can concretize the information, and offer a more nuanced representation than the standard deviation. Further, the use of visualization acts as a check for the appropriateness of the statistical shorthand. Finally, visualization is more democratic: it allows all users to reflect more effectively on their data, regardless of their level of familiarity with statistics. The visualizations afford an opportunity to provide other contextual information that simply cannot be captured in a numerical fashion. These contextual factors play an important role in uncovering and telling the teaching narrative that the numbers summarize, a summary that often has limited effectiveness. The visualization tools employ fundamental statistical concepts to ensure that the visual story is compatible with the statistical story to aid the users' understanding of statistical results and their accuracy, and help users avoid drawing inappropriate conclusions. Future work may further explore effective ways to integrate more advanced statistical practices with visual tools for exploring aggregate data.

For a more detailed exploration of the fundamental terms, methods and requirements of statistical calculation, please see Appendix A.

Ethical Principles in Data Aggregation

There are some obvious potential risks involved in the wholesale adoption of aggregate SRI data analysis, risks that are in essence a magnified version of the many challenges and tensions involved in all SRI activities. Firstly, the data involved are highly sensitive, and impact people's professional lives in significant ways. Just because it is possible to calculate, for example "the ten worst SRI scores on campus" does not make it constructive, statistically valid, or ethical. Further, the creation of tools to facilitate comparisons must be accompanied by mechanisms that guide people towards effective and appropriate data use, and limit the capacity for misinterpretation, bias, and the drawing of inappropriate conclusions. Finally, different institutions have significantly different understandings of how SRI data can and should be used, as reflected by the variability in collective agreements and policy guidelines across the province: these must be taken into account in considering how to approach data aggregation and the uses to which it can, and cannot, be put. These variations mean that in practical terms, tools built for actual institutional use must allow for customization at the institutional level.

It is clear that there are ethical issues to be addressed in the development of aggregate data tools, but a fully articulated set of guidelines is well beyond the scope of this study. We must find ways to establish practices that are collegial, appropriate, respectful, and beneficial. Some preliminary fundamentals are:

- practices must be in keeping with the ethical principles of the university as well as all policies and collective agreements;
- practices should be consistent with the stated purposes for which data has been gathered;
- practices must describe, clarify, and emphasize the limitations of data and tools, and limit user capacity to draw invalid conclusions where possible;
- practices must be respectful of instructors as central agents in teaching, and in many cases, as the owners of the data;
- drawing comparisons among and ranking individuals should be discouraged without extremely good reason;
- practices must be based on classification of data in terms of access rights and ability to drill down, and must also respect the need for confidentiality;
- data used in the aggregate should be anonymized, and under no circumstances should it be possible to disaggregate data in ways that make the identification of individuals possible; and
- practices must be in accordance with the Freedom of Information and Protection of Privacy Act.

There is a substantial and evolving body of literature in cognate fields such as business intelligence, health analytics, and learning analytics that could be drawn upon for the further development of appropriate guidelines. Slade and Pinsloo (2013), for example, identify the following in a discussion of the ethics of learning analytics (the large-scale use of student data for predictive purposes):

- learning analytics is a moral practice which should focus not only on what is effective, but on what is morally necessary;
- learning analytics should engage students as collaborators, co-interpreters, and agents, rather than as mere recipients of interventions;
- data should be understood as a snapshot view at a particular time and place, and identity and performance should be understood as dynamic and changing;

- student success is complex and multi-dimensional. Data are incomplete and analyses are vulnerable to misinterpretation and bias;
- there should be transparency regarding the purposes for which data will be used, who will have access to data, the conditions under which data will be used, and how and under what conditions privacy will be protected; and
- higher education cannot afford not to use these data (p. 12-13).

These kinds of principles appear to resonate well with the possible concerns that might arise in pursuing aggregate SRI data analysis.

It is one thing to build tools founded on and intended to promote ethical, methodologically sound data use: it is another for stakeholders to put faith in them. In general, aggregate data analysis is most likely to be effectively integrated into institutional practice if its use is of value to faculty members in pursuing their own goals and needs, and if their rights are protected through careful, consultative and incremental development of approaches to data use (Alderman & Melanie, 2012; Joughin & Winer, 2014). A process where tools are designed with faculty and administrators, and with sustained and proactive consultative processes with faculty associations (Alderman, in conversation, June 18, 2014) is more likely to produce a system that is sustainable, uncontroversial, and effective. Instructors, who are described by the data and also often own them, should also have opportunities to annotate the data so that contextual factors – first courses, introduction of innovative practices, team teaching, illnesses – can be introduced to support accurate interpretation of the narrative. It is impossible to predict all of the possible ways that tools like these need to be framed and delimited in advance of development and testing with real populations, so a thoughtful and responsive approach is necessary.

Committees, administrators, and faculty members are already making decisions based on data. Whether they are doing this well, with a strong understanding of what the data mean and do not mean, is a completely different question. There are two approaches to addressing this challenge: the first is education, and the second is the simplification and improvement of data and data reporting. While there is some evidence (Villascusa, Franklin, & Aleamoni, 1997 cited Hativa 2013b; Ludlow, 2007) that training can significantly improve facility with the use of statistical information for decision-making in teaching evaluation, there is also evidence that faculty members are not pre-disposed to engage with this kind of professional development (Ryan, 1997, cited Hativa, 2013). Visual representations, with clear markers of significance and limits, and tools, which disallow inappropriate or insignificant disaggregations or comparisons, can provide decision makers with clearer and more compelling data to work from. The goal of better analytical and visualization tools should also be to enhance equity, accuracy, and fairness. It is possible that access to data tools such as the ones described in this report, would impact the data culture and pre-disposition towards data use at Ontario universities, a much-desired outcome.

Applications of Aggregated SRI Data

Visualization Tools

As part of this study we developed prototype visualization tools to explore ways to present and manipulate existing SRI data drawn from its typical published formats. The purpose of such tools is to more clearly see the rich information embedded in existing SRI data, put it into context to improve its meaning, and improve decision-makers' ability to interpret, compare, and evaluate the data. The tools were developed in Microsoft Excel workbooks. For each tool, data tables are dragged-and-dropped onto a worksheet tab from standard reports published electronically in data files or on the Web. Charts and

tables on other worksheet tabs then automatically manage and display the data, needing only minor manual data manipulation in some cases.

SRI Course Report Generator

This tool uses the SRI survey results for one course and displays them in a user-friendly, visually oriented manner. The numerical data are presented with improved, graphical layout. Several types of graphs show the data in context, and some guiding interpretive statements are generated. An example of the tool in use can be seen in Appendix B.

Scatterplot Context Tool

This tool uses a table of summary SRI results for a large collection of surveys, for example all courses taught in a department over ten years, and plots each course as a point on a two-dimensional graph with the overall instructor score and overall course score on axes. Lines representing the departmental median for each score and some explanatory annotations help with interpretation. Interactive input boxes select the scores for specific instructors, courses, or combinations of the two, and those points are highlighted on the graph against the background of the department as a comparator group. Examples of the tool in use can be seen in Figures 7a, 7b, and 13.

Longitudinal Teaching Profile Tool

This tool uses a table of summary SRI results for an instructor over a period of time and plots overall instructor scores as several bar charts in one display, either one for each year or one for each course. The year of each course offering is shown in course groups, and the mean score for each group is shown to aid interpretation. The user may add text annotations to groups or to individual scores to provide additional context. Examples of the tool in use can be seen in Figures 9a, 9b, 10a, and 10b.

Departmental Teaching Timeline Tool

This tool also uses a table of summary SRI results for all courses taught in a department over ten years, but rather than lumping all results together for the time period like the *Scatterplot Context Tool*, the distribution of scores is displayed as a boxplot for each individual semester, and the boxplots are shown in sequence over time. Examples of the tool in use can be seen in Figures 14 and 15.

The sections that follow demonstrate the prototype tools built to date, and illustrate the range of their potential for instructors, program chairs, and other institutional purposes: while these live, operating tools provide proof of concept, a further phase of this project involving extended consultation with stakeholders would most likely open up further avenues of exploration.

Data Aggregation and Visualization Tools for Instructors

Instructors use SRI data for a range of purposes:

1. For most instructors, SRIs' primary formal function is demonstrating teaching effectiveness in relation to personnel decisions. For sessional instructors or graduate students with teaching responsibilities, reporting on SRI is also a critical element of career development and the job search.
2. SRI data play an important role in teaching award processes for high-performing instructors.
3. Teaching is a reflective practice: SRI is one form of the many kinds of data that feed into instructor reflection on an informal and formal basis.
4. On a formative basis, instructors may study or analyze their data for course and instructional improvement planning.
5. They may also use them as one form of data when considering whether changes they have made in their courses have had an impact on student satisfaction with the learning experience.

Visualization of data enables instructors to more clearly identify patterns in their scores, even at the level of the individual course. Aggregation of data allows instructors to consider patterns in their teaching over longer periods of time or across various groupings of courses. However, without visualization, these data can be difficult to parse. Integrating the two approaches makes patterns more obvious and meaningful, and assists in clarifying contextual factors that may be playing a role in these patterns. Firstly, visualized aggregate data enables instructors to better contextualize their data: they can examine them within the context of a norm group such as their department, and in more nuanced ways than a narrow ranking of overall scores offers. Secondly, visualized aggregate data enable longitudinal explorations and comparisons, where instructors can examine their data, either at the overall "instructor score" scale, or at the level of subscales; within a course they have taught multiple times; or across all of the courses they have taught in a given period of time.

Improving Single-Course Reporting Through Visualization

Table 1 is a standard, numbers-based course evaluation report. It provides a mean response distribution table for all instructor-related and course-related items from an SRI form, as well as overall instructor and course scores derived from each of the two sets of questions. Some basic course information is included: the enrolment, the number of students for whom the course is required, and the number of students expecting various grades. The number who participated in the survey, as well as change in enthusiasm for the course, are also included. As previously noted, the "mean" is in theory a collapsed version of the response distribution for each item. Although the distribution is provided, it is difficult to get a sense of it from what is shown here. Also, each number must be considered compared to the overall number of responses, adding a further layer of difficulty to making meaning from the report.

Table 1. Student Ratings of Instruction Course Report

Student Rating of Instruction for: Course 1 Instructor A 2013 Fall

Instructor Score	Course Score	Number Enrolled	Responses	Course Req'd for	Course Not Req'd	# of Students expecting a grade of A	B	C	D	F
6.1	5.4	41	26	25	1	7	11	1	0	0

Instructor related questions:	Mean	Extremely Poor	Very Poor	Poor	Adequate	Good	Very Good	Out-Standing	n/a
1. presented material in an organized, well-planned manner	5.5	0	0	0	2	11	10	3	0
2. was approachable for additional help	6.7	0	0	0	0	1	7	18	0
3. was accessible to students for individual consultation (in office hours, after class, o	6.6	0	0	0	0	1	8	17	0
4. The overall effectiveness of the instructor was	6	0	0	0	0	7	12	7	0
5. used instructional time well	4.9	0	0	0	8	12	3	1	0
6. explained content clearly with appropriate use of examples	5.8	0	0	0	1	8	10	6	0
7. was a clear and effective speaker	6.1	0	0	0	0	5	14	7	0
8. communicated enthusiasm and interest in the course material	6.3	0	0	0	0	3	12	11	0
9. stimulated your interest in the subject and motivated your learning	5.9	0	0	1	1	5	12	7	0
10. attended to students' questions and answered them clearly and effectively	6.5	0	0	0	0	3	6	16	0
11. was open to students' comments and suggestions	6.5	0	0	0	0	1	10	15	0
12. was sensitive to students' difficulties	6.3	0	0	0	0	3	11	12	0

Course related questions:	Mean	Extremely Poor	Very Poor	Poor	Adequate	Good	Very Good	Out-Standing	n/a
1. How effective was the course outline in communicating goals and requirements of	5.4	0	0	0	5	7	10	3	1
2. How reasonable was the level of difficulty of the course material?	5.7	0	0	0	0	9	15	2	0
3. How reasonable was the volume of the work required in the course?	5.8	0	0	0	0	9	12	5	0
4. The value of the overall learning experience was	5.6	0	0	0	2	10	7	5	0
5. How consistently did the stated course goals match what was being taught in the c	5.3	0	0	1	2	12	8	2	1
6. How appropriate was the course format for the subject matter?	5.5	0	0	1	3	6	13	2	0
7. How well did the methods of evaluation (e.g., papers, assignments, tests, etc.) refl	5.7	0	0	0	2	7	11	4	0
8. How fair was the grading of student work?	5.8	0	0	0	0	9	9	4	1
9. How timely was the grading of student work?	3.9	0	4	6	9	2	3	1	0
10. How helpful were comments and feedback on student work?	5.1	0	0	2	5	10	5	3	0
11. How well did the instructional materials (readings, audio-visual materials, etc) fa	5.5	0	0	0	3	9	8	3	1
12. How well did the instructional activities (lectures, labs, tutorials, practia, field trip	5.5	0	0	0	2	11	10	2	0

	Low	Medium	High
Your level of enthusiasm for taking this course at the time of initial registration:	0	13	13

	Low	Medium	High
Your level of enthusiasm for taking this course at the conclusion of the course:	3	11	11

	Yes	No
Considering your experience with this course would you recommend it to other students:	22	4

In order to improve on this kind of reporting, Hativa (2013a) recommends the provision of an end-of-term, single course instructor report with limited data visualization (Table 2). This provides instructors with their median and mean score on each item, a standard deviation, a frequency count, correlation between their item score and their overall instructor score[1], and a simplified box plot which identifies the first and third quartile of the comparison group, the comparison group mean, the instructor's item score mean. This model also provides linked "teaching tips" in some areas.

Table 2. Single-Course Report Format (Adapted from Hativa, 2013a)

Enrolled	Responded		Gender		Level	Required	Class attendance				Expected grade	GPA
40	**31**	**78%**	**F 57%**	**M 43%**	**Undergraduate**	**76%**	**<25% 3**	**25-50% 6**	**51-75% 9**	**76-100% 13**	**...**	**...**

ITEM	Your Course			Comparison group		Frequencies							Correlation ^	Simplified Box Plot
	Med	**M**	**SD**	**M**	**SD**	**1**	**2**	**3**	**4**	**5**	**6**	**7**		- 1 - - 2 - - 3 - - 4 - - 5 - - 6 - - 7
1. Overall course	6	5.7	1.2	5.8	0.7	0	0	3	1	5	14	8	.92	----------[-* I-] - -
2. Overall teaching	6	6.0	1.0	6.1	0.6	0	0	0	3	6	9	13	1.00	----------[*I-]-
3. Organization	6	5.7	1.0	5.8	0.7	0	0	0	4	9	11	7	.75	----------[*I-]- -
4. Clarity	6	6.1	0.9	5.9	0.7	0	0	0	2	5	13	11	.74	----------[*I-]-
5. Engagement	5	5.3	1.4	5.6	0.8	1	0	1	6	8	9	6	.91	----------*[- -I-]-
6. Intellect. challenge	5	4.8	1.5	5.4	0.8	1	2	1	8	9	6	4	.80	------- -*-[- - I--]- - - -
7. Questioning	7	6.4	0.8	6.4	0.5	0	0	0	1	4	8	18	.81	----------*[I] -
8. Rapport	7	6.3	0.8	6.4	0.5	0	0	0	0	7	8	16	.82	---------[*I-]-
9. Workload	6	5.7	1.0	5.0	1.0	0	1	0	1	9	15	5	.15	-------[- - I - -]*- - - -
10. Difficulty	5	4.5	1.3	4.6	0.7	0	4	2	8	10	7	0	.10	-----[- - *I-]- - -
11. Reading materials	6	6.1	1.2	5.6	0.7	1	0	0	1	2	15	11	.76	----------[I - -]*- -
12. Homework	5	5.0	1.6	5.6	0.9	2	1	1	4	8	9	4	.72	----------[*- -I- -]- -

Box Plot readings: [ ] 1st and 3rd quartiles, repectively (25% and 75% of the population of the comparison group),
|: Comparison group mean,
*: Mean of the teacher/course
^: Correlation with Overall Teaching
Items in bold are linked to teaching tips for improving each behavior respectively.

Although Hativa's model has much to offer, some improvements might be made in its use of visualization. The SRI Course Report Generator provides an automated approach to generating these visualizations.

A first step in improving instructors' ability to parse and understand the data is achieved by visually and numerically focusing the data by expressing the frequencies as percentages, highlighting the top two responses so that consensus or polarity is obvious, and suppressing the zeroes so that the numbers stand out (Figure 1).

[1] Hativa (2013a) notes that technically ordinal data should not provide means or averages as measures of central values, more properly being reported through medians, modes, frequencies, percentiles and so on. She argues, however, that means, averages, factor analyses and regressions are widely used in the social sciences for ordinal scales, and that many well-established and widely used survey systems like SEEQ do employ these approaches, concluding that "the policy is widely agreed upon in practice" (Hativa, 2013a, p. 64). In this instance, a Spearman Rank Correlation coefficient should be used with ordinal data. While our model has not used this approach, it would not be difficult to incorporate into subsequent tools if required: to examine correlations, we have focused on tools to visualize the tightness of scatterplots rather than the provision of a single statistical number, which is more in keeping with our view of the uncertainty of the data set used and its implications for user understanding.

Figure 1. Improving Readability

Mean	1 Extremely Poor	2 Very Poor	3 Poor	4 Adequate	5 Good	6 Very Good	7 Out-Standing	n/a
6.1				4%	19%	37%	39%	

A second step attaches a small visual representation (a "sparkline") of each item score to complement the percentage data (Tufte, 2006). This allows the reader to clearly and easily understand the relative size of each response group and the overall distribution, which can be of crucial importance in distinguishing, for example, between polarized opinions and consensus around a common but unexceptional score, each of which necessitates a different kind of inquiry.

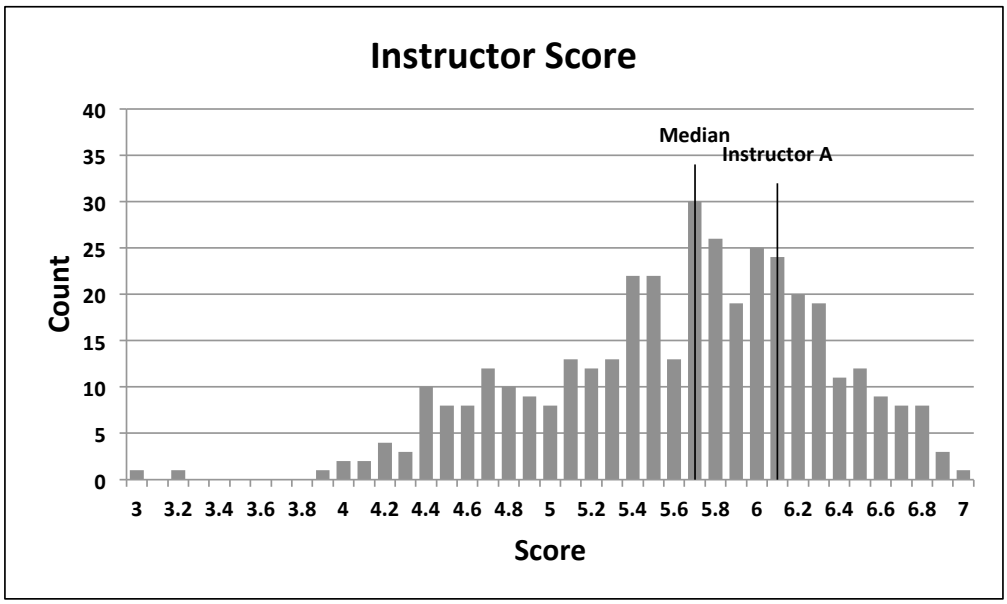
Figure 2. Visualizing Frequency

	Mean	1 Extremely Poor	2 Very Poor	3 Poor	4 Adequate	5 Good	6 Very Good	7 Out-Standing	n/a
Instructor (12 questions)	6.1			0%	4%	19%	37%	39%	

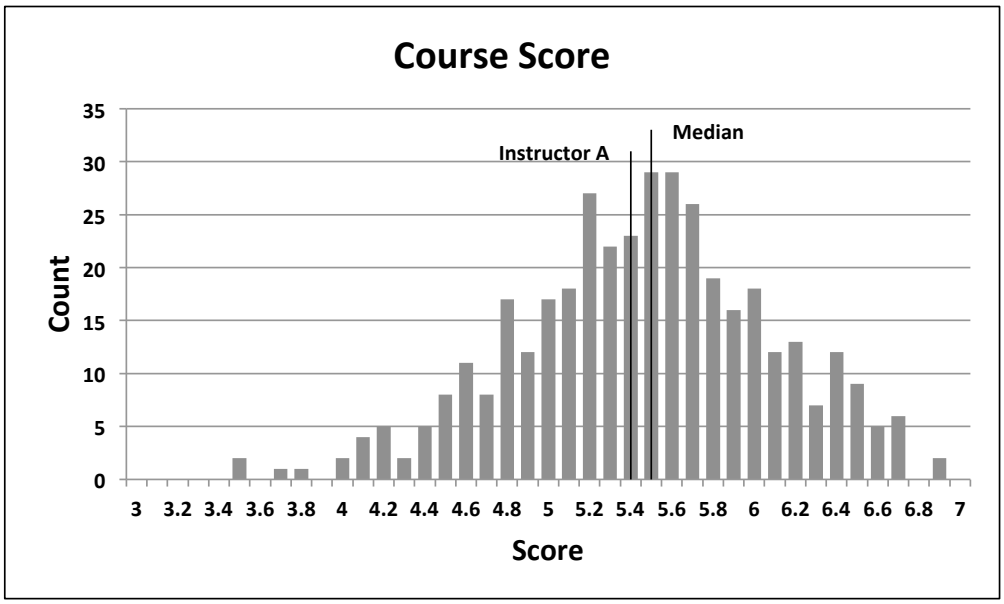
Although Figure 2 is a significant improvement, as it allows for clear visualization of distributions, it is also important to see the scores in context. Figure 3 allows instructors to compare their course score with the historic range and typical scores within their department based on a frequency count of instructor scores, and a second one of course scores, for a ten-year period. This helps give more meaning to the number. The tool generates graphs and comments, which provide very simple interpretation: a median and percentile range. The use of a ten-year aggregate histogram also allows for the provision of context in a way that preserves the confidentiality of other instructors in the department.

Figure 3: Visualizing Context

How do my scores for this course compare to the rest of my department over the past ten years?



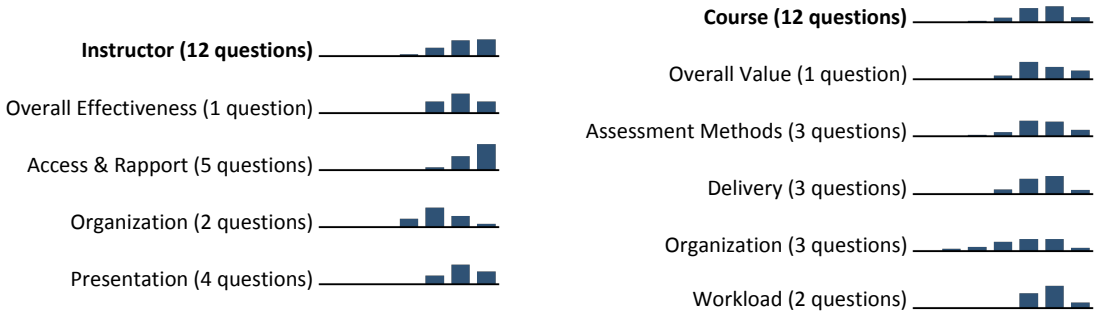
Your Instructor Score is in the 70th-90th percentile group.
This means that you were rated higher on the instructor-related questions than 70% of the other courses taught in your department over the past ten years.
Very good!



Your Course Score is in the 30th-70th percentile group.
This means that your rating on the course-related questions was in the middle 40% of all courses taught in your department over the past ten years.

A fourth improvement can be achieved through the use of thematic aggregation, or alternatively by the "disaggregation" of the composite overall score. The use of an overall instructor score, while statistically and interpretively valid, may also obscure meaningful variations in the dimensions of an instructor's teaching practices that might help them to understand their strengths and weaknesses. Many SRIs, particularly those based on the SEEQ model, are based on subscales – thematic clusters of questions (Marsh, 1982). In the case of SEEQ, these include learning, enthusiasm, organization, group interaction, individual rapport, breadth, examinations, and assignments (University of Saskatchewan, n.d.). Questions can be clustered into subscales based on themes, but they must then undergo statistical analysis to ensure that the subscales are valid. In Figure 4, thematic subscales are combined with sparklines allowing for a quick, clear sense of differences in students' perceptions of the instructor and course effectiveness for the class.

Figure 4. Visualized Sub-scales



In our example, students scored the instructor very highly on "access and rapport" and less highly on "organization" – both instructor- and course-related elements identify this as a weaker area. This nuance would not have been clear at the level of overall instructor score, and might not have been evident even from the distributions of the detailed items. Patterns which occur consistently across courses can identify areas for improvement, and also areas where instructors have valuable insights to share as support for colleagues. This visual representation also makes it easier to informally identify patterns across multiple courses. It can also be useful to apply the same sparkline visualization treatment to individual items: it offers good opportunities for formative feedback and can suggest useful directions for further inquiry.

The full model course report incorporating all of these visual enhancements can be found in Appendix B.

Exploring Multiple Course Data for a Single Instructor

The ability to study SRI data across multiple courses, semesters, and years enables instructors to uncover patterns in their practice for further improvement, and also to contextualize their data, both for their own understanding and for reporting purposes. In order to do so, instructors need access to their data in ways that allow them to see both "collapsed" versions such as means and medians as well as fuller representations of the data, for example, distribution patterns: the latter is significantly improved by effective visualization tools.

Table 3 provides a typical way that SRI multi-year data are reported: in this case, the overall instructor and course scores for each course offered by a given instructor, as well as enrolment and response rate,

organized historically.[2] While an instructor might receive a version with only their own courses, a program chair might receive a complete table for all instructors in the department. Table 4 provides even more concentrated data: instructor and course score aggregated over ten years, with the mean and standard deviations of each instructor's scores.

Table 3: Basic Multi-Course Data Table (adapted from Hativa, 2013a)

Semester	Teacher	Course	Enrolled	Responded		Overall	
						Teaching	Course
			#	#	%	M	M
2011a	A	a	96	70	73	3.9	3.7
2010b	A	b	16	6	38*	5.3	5.5
2010a	A	c	19	6	32*	5.3	4.3
2010a	A	a	17	7	41	5.1	4.9
2010a	A	a	148	87	59	4.5	4.3
2009b	A	a	18	9	50	4.7	4.2
2009b	A	d	124	91	74	3.6	3.6
2009b	A	e	14	13	93	5.9	5.3
2009a	A	c	17	7	41	6.4	5.9
2009a	A	a	134	80	60	4.1	4.0
2009a	A	f	132	87	66	4.4	4.3

Rating scale: From 1 (lowest) to 7 (highest) *Response rate lower than 40%

If the report is provided as a spreadsheet, the instructor or chair can sort for a small number of characteristics: scores in specific courses, or by class size or level. The report lends itself, however, to summary comparisons that may not be constructive: richer data analysis tools can help here.

Table 4: Concentrated Instructor Multi-year Data

Student Rating of Instruction - Department Summary - Fall 2004 to Fall 2013

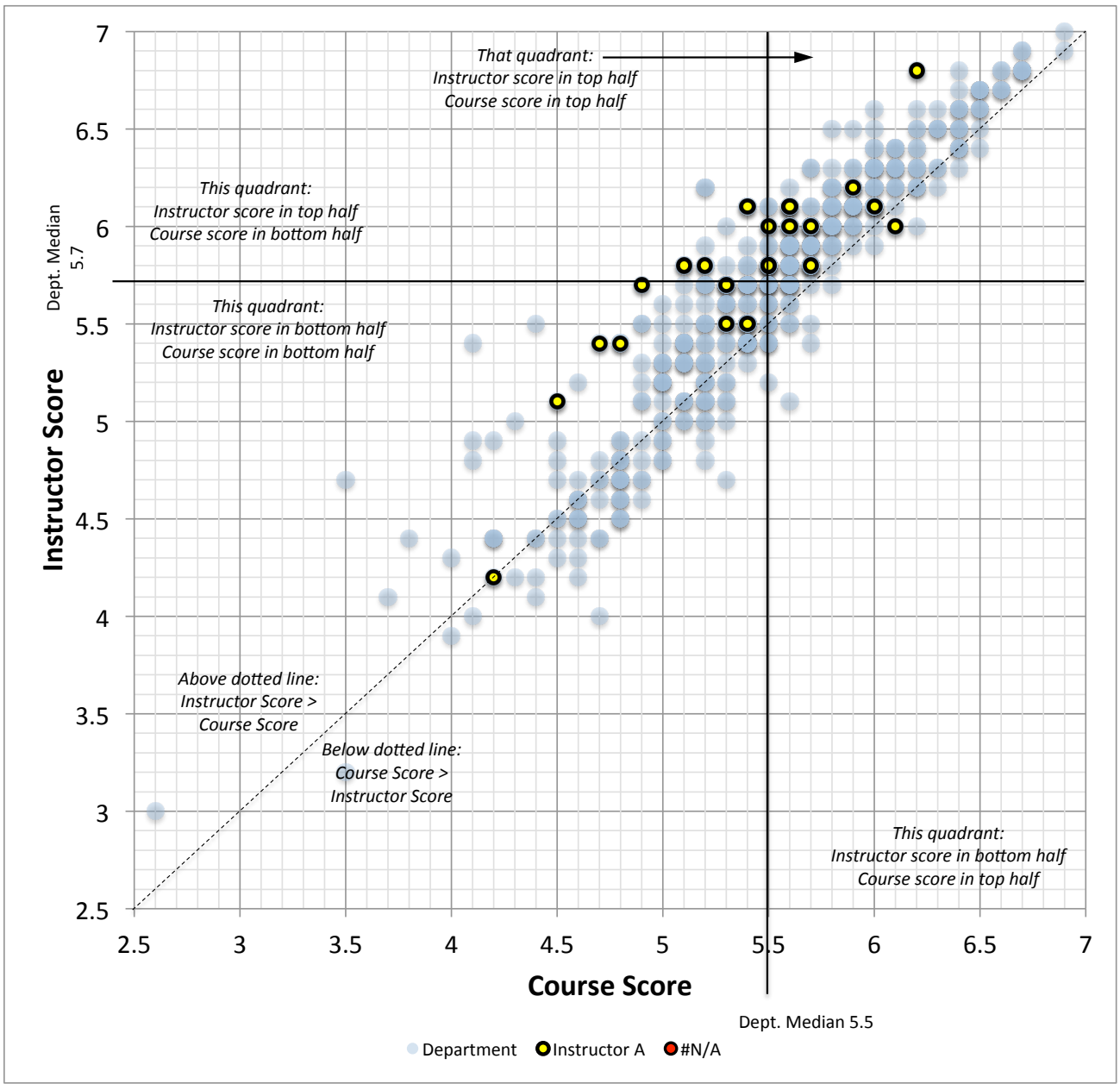
Name	Number of Courses	Instructor Score		Course Score	
		Average	Std. Dev.	Average	Std. Dev.
Instructor A	**23**	**5.8**	**0.49**	**5.4**	**0.50**
Instructor B	**27**	**5.9**	**0.25**	**5.6**	**0.27**
Instructor C	19	5.9	0.40	5.4	0.51
Instructor D	37	6.6	0.25	6.4	0.27
Instructor E	22	4.7	0.49	4.7	0.41
Instructor F	15	5.5	0.27	5.4	0.22

[2] Hativa also recommends an end-of-year, multi-year, all-course report, which allows instructors to compare their ratings on the main items across years, either looking at the same course or looking at different courses. The inclusion of comparator data is recommended, as is the provision of the data both as a report and in a spreadsheet in order to allow for data manipulation. This allows instructors to sort courses by enrolment or overall instructor or course score to examine patterns in the data and to make basic graphs from the data: there is, however no guidance regarding what can and cannot be charted or analyzed. In theory these reports enable instructors to identify strengths and weaknesses, to compare student perceptions of effectiveness across types of courses, and to examine trends, but the process is resource-intensive.

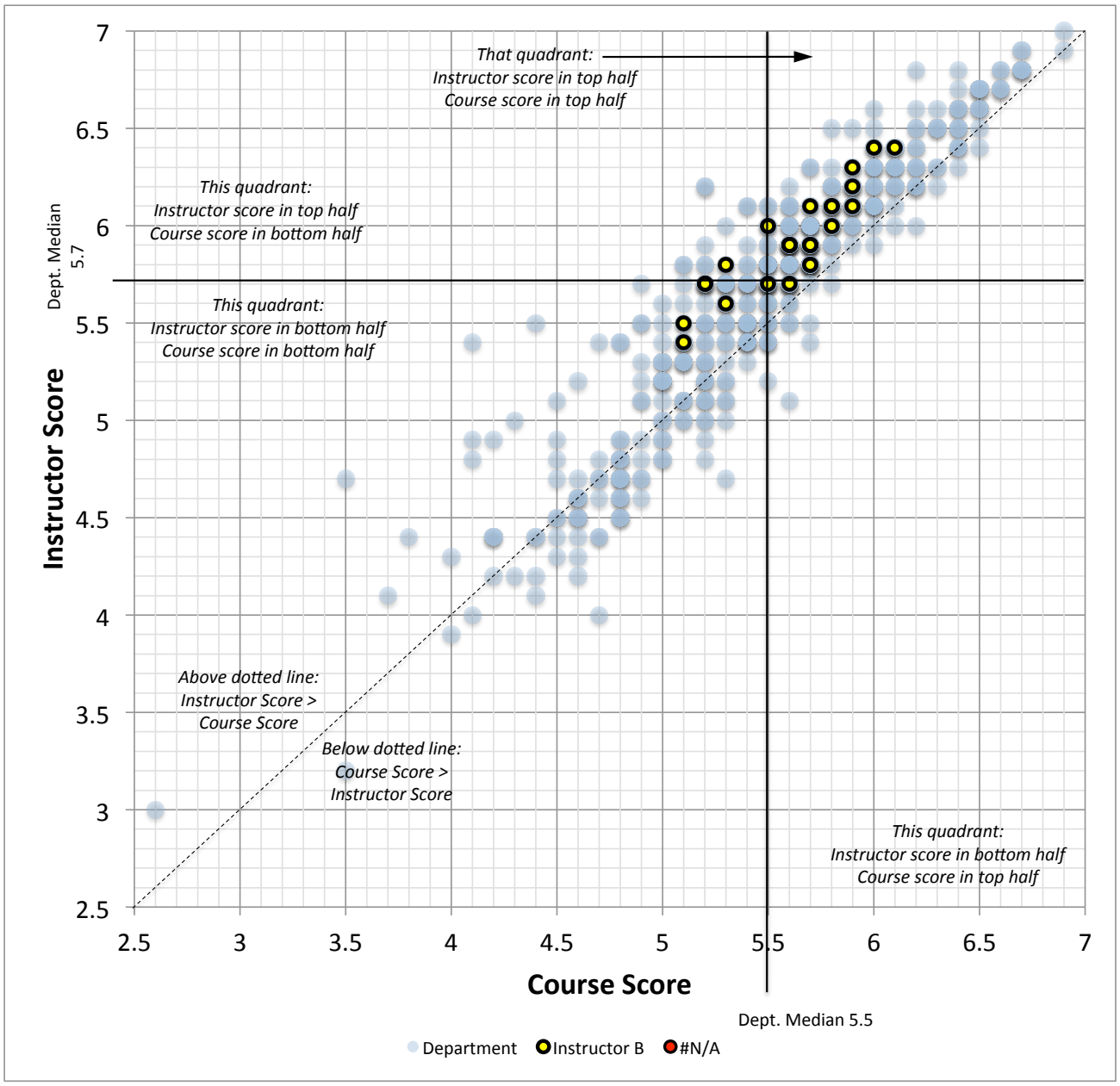
From a first glance at Table 4, Instructors A and B look comparable, with a similar number of courses, and similar course and instructor scores. However, the standard deviation is significantly different. What might that mean? Contextualized longitudinal visualization offers some suggestive insights into the nature of the standard deviation.

Figures 5a and 5b: Instructors A & B: Course Scores Plotted Against Instructor Scores for a Ten-Year Period, using the *Scatterplot Context Tool*

What instructor do you want to see? Instructor A **What course do you want to see?**



What instructor do you want to see? Instructor B **What course do you want to see?**



As noted above, the Scatterplot Context Tool enables viewer to select any instructor, course or combination of the two, and show course offerings that match the selection (highlighted circles) against a backdrop of all courses offered in the relevant department during the time period represented in the

data table. Median departmental course and instructor scores are provided for ease of analysis. The scatter of the dots shows the central tendency and variation across the selected and background scores.

Figures 5a and 5b clearly show much more variation in Instructor A's SRI ratings than Instructor B's, but the centre is essentially the same. This tool allows instructors and administrators to clearly visualize the idea of standard deviation and its actual nature in a given case: Instructor A still has a fairly tight group, similar to Instructor B, but a small proportion of the courses extend further out. Is there anything meaningfully different about these courses compared to the central cluster? Context may provide some insights (see "Longitudinal Teaching Score Analyses" below).

Extending the Potential of the Scatterplot Context Tool

Further iterations of this tool will allow instructors and program chairs to contextualize their SRI data in a variety of ways. It can be set to identify a team of instructors who taught the same course, each identified by a different colour. It can also be modified to employ additional filters, which would allow viewers to look at the distribution of their scores against a context of different norm groups, for example, courses of over 100 or 200 students, required courses, courses at the same level, and so on. The limitation here is primarily whatever attributes can be easily pulled into the spreadsheet from existing databases.

McGill University offers an alternative approach to providing instructors (and administrators) with a range of norm groups for comparative purposes. Figure 6, found in *The McGill Guide To Interpreting End-of-Course Evaluation Results* (Winer et al., 2012) offers instructors their mean score on each item, accompanied by the mean item score for: all sections of the course, the instructor's department, all courses in faculty at the same course level, and all courses in the faculty in the same size category. Each of the histogram bars here could be visualized as a scatterplot, with the instructor's course highlighted within the plot to offer a richer understanding of the data. However, this form of representation may serve a useful summary purpose.

Figure 6 Histogram Representation of Norm Group Comparisons (Winer et al., 2012, p. 7)

Different forms of visual representation can aid in understanding the results; below is an example of course means compared to different groups using a column chart.

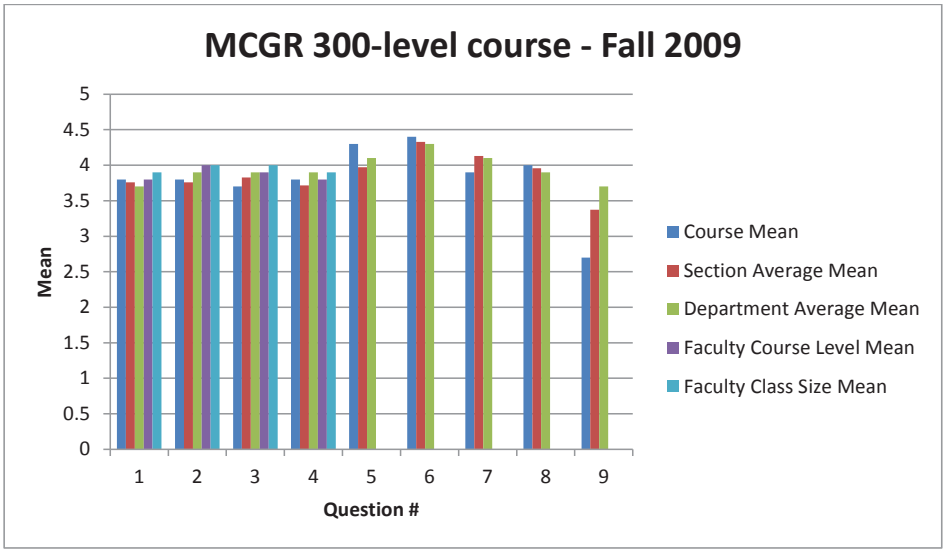
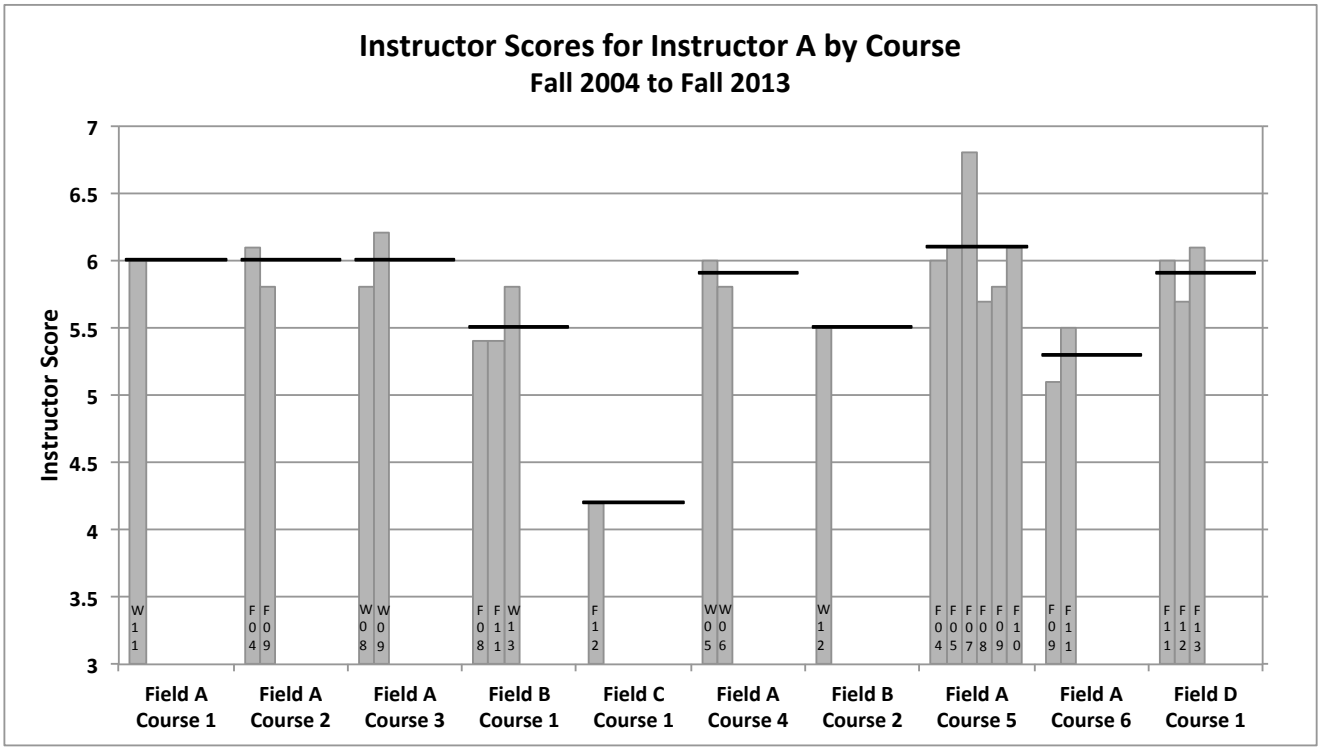


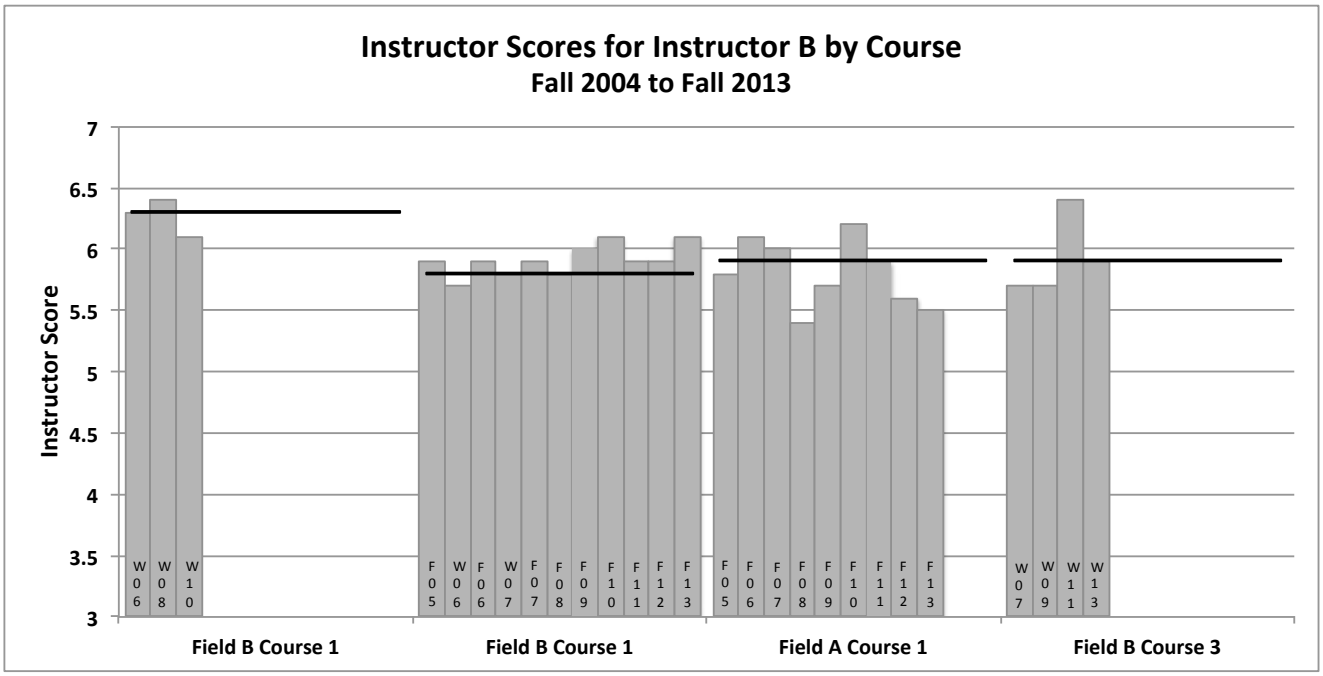
Figure 1: Faculty of Management - 300-Level Course

It would also be possible to adapt the scatterplot tool to show an instructor's scores on specific subscales compared to a norm group's scores on the same subscales. Any statistical comparisons require subscales based on enough items to be statistically valid (Glass et al., 1972). However, visualization for a broad understanding of pattern is reasonable for any subscale. This would enable instructors to see, for example, their "approachability" subscale score in comparison to all instructors' approachability subscale scores in the same norm group. This method can only give a general perspective of trends and patterns, and is not suitable for examining individual items, since they are not intended in the instrument design to be used independently (Carifio & Perla, 2001).

Longitudinal Teaching Score Analyses

Contextual information is important for meaningful interpretation of SRI data. One element of this is the exploration of longitudinal data. The differences in the score distributions of Instructors A and B, discussed above, for example, may be informed by an examination of their historical instructor score trends clustered by course (Figures 7A and 7B). These were generated using the Longitudinal Teaching Profile Tool. The horizontal lines on each course cluster represent the instructor's mean instructor score in that particular course over the time period.

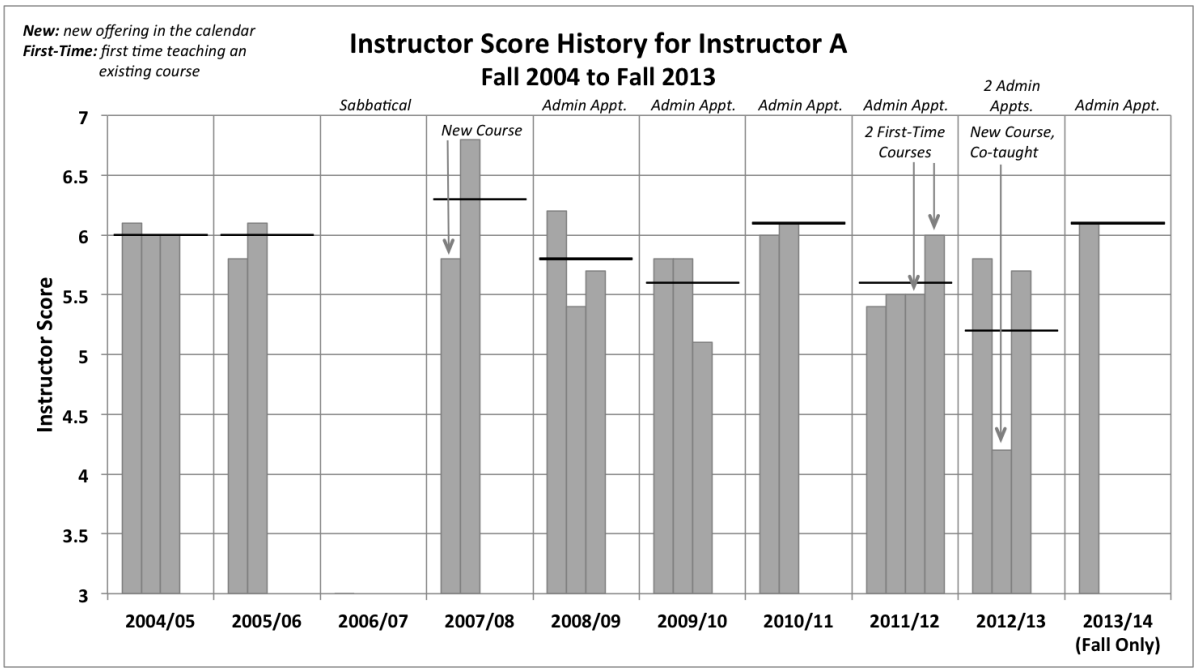


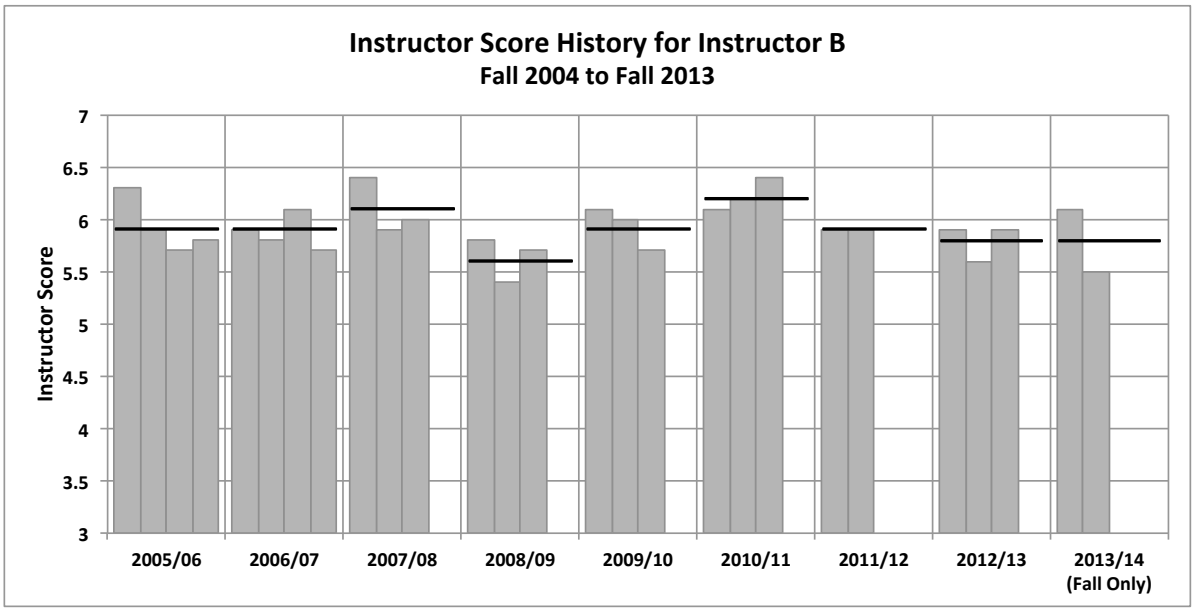


Instructor B has taught four courses very frequently (two have actually been taught more than 25 times over the instructor's career, not just the times shown in this ten-year window) and two others taught in alternating years with a long career history of quite consistent instructor score results. Over the same period of time, Instructor A has taught 10 different courses in four different fields. In most cases, Instructor A has taught each course a maximum of three times, sometimes with significant gaps of time in between. The instructor has only had one consistently offered course, taught 12 times during the instructor's career. This type of information is important in articulating the nature of the teaching role instructors are playing in their departments, and in providing appropriate context for analysis and decision-making: visualization makes the pattern much clearer, limiting the temptation to draw hasty comparisons. While in the cases used here, there is no obvious pattern of strengths in one course or type of course and weaknesses in another, this tool makes such patterns very clear. It would also provide suggestive visual evidence of improvement over time.

There are many other intervening factors that might be included in visual representations to ensure a clear sense of context. Figure 8a indicates that Instructor A has taught a small number of courses per year. However the instructor's annotations (which were selected and input manually) provide context regarding teaching load. As shown in Figures 8a and 8b, Instructor A has taken on several administrative duties, including a double responsibility in the semester of the lowest course score and largest teaching load. Instructor A has also introduced new courses to the undergraduate calendar, and because of the variability of load, taught courses for the first time more often. Two of Instructor A's lowest teaching scores aligned with new courses: the lower of the two was also team-taught. Using visualization and annotation provides a clearer sense of teaching loads and other contextual factors, which may inform a better understanding of the instructor's performance.

Figures 8a and 8b. Instructor Scores, by Year, Annotated as Necessary, Using the *Longitudinal Teaching Profile Tool*





These kinds of visualizations and annotations can be used to enormous effect in teaching dossiers and summative teaching evaluations to illustrate and support claims regarding teaching practice and effectiveness. The capacity to annotate these data would also enable faculty to identify for example courses where specific innovations were attempted, or other intervening factors that may have affected performance in a given course in a given year. They offer powerful tools for formative evaluation and reflection and could be used to good effect in peer review and teaching consultation contexts: one of their real strengths is their capacity to serve as the basis for establishing dialogue and further inquiry (Ludlow, 2005).

Extending the Potential of the Longitudinal Teaching Score Tool

The use of this tool could be expanded by identifying a small number of available course characteristics to create further categories within these visualizations: class size or level might be colour coded for example, or courses in the lowest or highest quartile of the overall distribution of scores. The system could also generate course-identifying labels for the histogram bars. As our prototypes were built with simple spreadsheet software, there were limitations to the graphical sophistication we could employ. An extended version of this tool would also enable instructors to select specific subscales across all courses for visualization, with the usual caveats. In theory, a system like this could also be used to chart all instantiations of a course by different instructor groupings, or all courses of specific types (for administrative purposes).

Data Aggregation and Visualization Tools for Program Chairs

Program chairs engage with SRI data for a number of different purposes, many of which could be facilitated by better access to and guided analysis of both longitudinal and aggregate departmental data.

1. They undertake the **regular and pre-tenure review of instructor teaching**, a process that would be significantly streamlined by access to improved data visualization and analysis tools. Drawing conclusions from instructor data frequently requires considerable manual labour. Chairs may therefore adopt the use of only instructor or course means rather than looking more closely at the kinds of patterns that would be revealed by better visualized frequency and distribution data. Visualizations also provide efficient and compelling evidence to be included in performance reviews, dossiers and other submissions for decision-making purposes, and allow instructors and administrators to discover or document other external contextual influences that may explain anomalous low scores or high degrees of variability in scores. Conversely, visualizations may provide compelling evidence of sustained patterns of weak teaching for both summative and formative purposes. Longitudinal representations of scores for an individual instructor are also useful and efficient, although it is methodologically inappropriate to calculate increments of improvement over time from these kinds of data.

2. Program chairs **work with individual faculty members on teaching improvement**. Better visualized data allows for an inquiry-based and reflective approach that might be very constructive in this area. The visualizations described above are equally useful for generating dialogue, directions for inquiry, and plans for improvement, and may be of real benefit in helping faculty improve, particularly early-career faculty. Over time, chairs are likely to become more adept at seeing patterns in visualized data, as they will see much more of it than the individual faculty member: this may be of assistance in supporting faculty. Centres for Teaching and Learning can be of considerable assistance here, if given permission to review instructor data.

3. Program chairs may use SRI data in **making teaching allocation decisions**. As discussed above, these decisions might be made more strategically if it were possible to look at all course data for the department for a number of years, and then disaggregate by course type to identify instructors' strengths in teaching specific kinds of courses. A version of the Longitudinal Teaching Profile Tool would be useful in visualizing data for this purpose.

4. Program chairs **evaluate courses and programs for improvement and refinement purposes**. The scatterplot contextualization tool would allow chairs to examine all offerings of a given course over a number of years in the context of all departmental offerings, as well as courses in the context of similar courses (See Figures 5a and 5b). It would also allow for the examination of specific types of

courses: all first-year courses, for example, or all courses with more than 100 or 200 students. A version of the longitudinal teaching score tool would facilitate the examination of groups of courses in light of specific sub-scales of the SRI, to examine, for example, whether first-year students generally reported that their instructors were accessible, how that compared with students in other years, and so on. This approach would help chairs to identify targeted areas for improvement and departmental initiatives.

5. Program chairs **articulate the value and quality of their program** for processes such as the Institutional Quality Assurance Protocol and for accreditation purposes. Better access to, use of, and visualization of SRI data may be critical for easier and improved reporting of this nature, and in particular for establishing reporting processes that result in better informed and more strategic recommendations, and better informed instructors and administrators. An underlying principle here is the importance of developing approaches that allow for easy integration of data, and inclusion of evidence of different types and from different sources, to develop richer and stronger narratives describing the program. Methods which go beyond minimum reporting requirements, can leverage the effort that goes into the QA reporting process to create opportunities for more effective, targeted program improvement.

6. Program chairs support **the development of courses and programs**. Aggregate data may be of use here to inform discussions about areas of strength and weakness within a program. Student perceptions of program offerings can complement the faculty members' perspectives in order to determine further directions for development. Of course in all cases, students' open-ended feedback forms an important element of the evidence for consideration: although not an element of the tools developed to date, further exploration of data tools for organizing and analyzing qualitative data might be of considerable value for both instructional and programmatic improvement.

7. They **focus on the success and achievement of students in the department's programs, as well as students from other programs who are supported by the department**. Aggregation tools can be designed to also allow for disaggregation of data to examine student subpopulations within courses. Such exploration can identify whether needs or concerns specific to a given subpopulation are being masked within the overall response pattern. There are numerous ethical considerations to take into account when exploring this possibility: preserving privacy and confidentiality of small subgroups; the risk of removing subgroups from analysis and consequently marginalizing them; and deliberate evidence-seeking to justify removal of access to a subgroup are among the issues that should be discussed. The latter issue speaks to a principle central to the development of tools to explore aggregate SRI data, or truly, any institutional data: they should be designed and used with the intention of disconfirming bias rather than reinforcing it.

Standard Approaches to SRI Data Reporting for Program Chairs

Hativa (2013a) describes several reports that provide aggregate data for administrators: *An end-of-term, all courses and instructors' report* offers concise, easily parsed information about all courses in the department, enabling administrators to identify and address low-rated courses (Table 5). Depending on the degree of detail in the report, it may also provide evidence of low ratings on specific items or scales for all courses or a group of courses, study of which might suggest areas for inquiry, remediation, and professional development across the unit. These data may also be used to inform decision-making about course allocations for future years. As can be seen in Table 5, however, patterns are not immediately apparent and cannot be derived from the data without further effort.

Table 5. End-of-Term All Courses and Instructors' Report (adapted from Hativa, 2013a)

Instructor		Course		Enrolled	Responded		Subscale Means				Overall Teaching		Overall Course	Simplified Box Plot
ID	Name	ID	Title	#	#	%	Communication of material	Interpersonal Rapport	Course Components	Course Difficulty	Mean	Median	Mean	
	A!		m	116	92	79	4.4	6.1	5.6	4.9	4.5	5	4.6	
	B!		n	133	66	50	4.2	4.0	5.6	4.7	4.3	5	4.3	
	C!!		o	125	94	75	6.0	6.5	6.0	4.7	6.5	7	6.1	
	D!		p	86	47	55	4.4	5.8	4.9	4.8	4.3	4	4.6	
	E!!		q	139	103	74	5.5	6.0	6.2	5.0	5.8	6	5.5	
	F!		r	122	74	61	5.0	5.5	5.5	4.7	5.2	5	5.0	
	G!		s	19	11	58	5.1	6.5	5.3	4.7	5.9	5.5	4.4	
	H!		t	35	13	37*	4.7	4.8	5.1	5.4	5.0	5	4.8	
	I!!		u	66	59	89	6.0	6.6	5.1	5.0	6.4	7	5.7	
	J!		u	81	49	60	5.5	6.1	5.9	5.2	6.0	6	5.5	

Response rate <40%; ! Tenure/tenure track !! Primary instructors not tenured or on tenure-track
* Response rate lower than 40%

Hativa also describes a variant of this table for decision-making regarding promotion and tenure issues. It includes all relevant courses and all the times that the instructor taught these courses, and characteristics such as class size and course level. Generally speaking the research indicates that at a minimum, decision-making should be based on all courses taught in at least two years of teaching, that is, a minimum of six to eight courses. Hativa does not recommend aggregating scores for this purpose: although combining the results can theoretically increase the power of statistical tests, substantial differences among courses will mean that combined results may obscure rather than amplify important patterns.

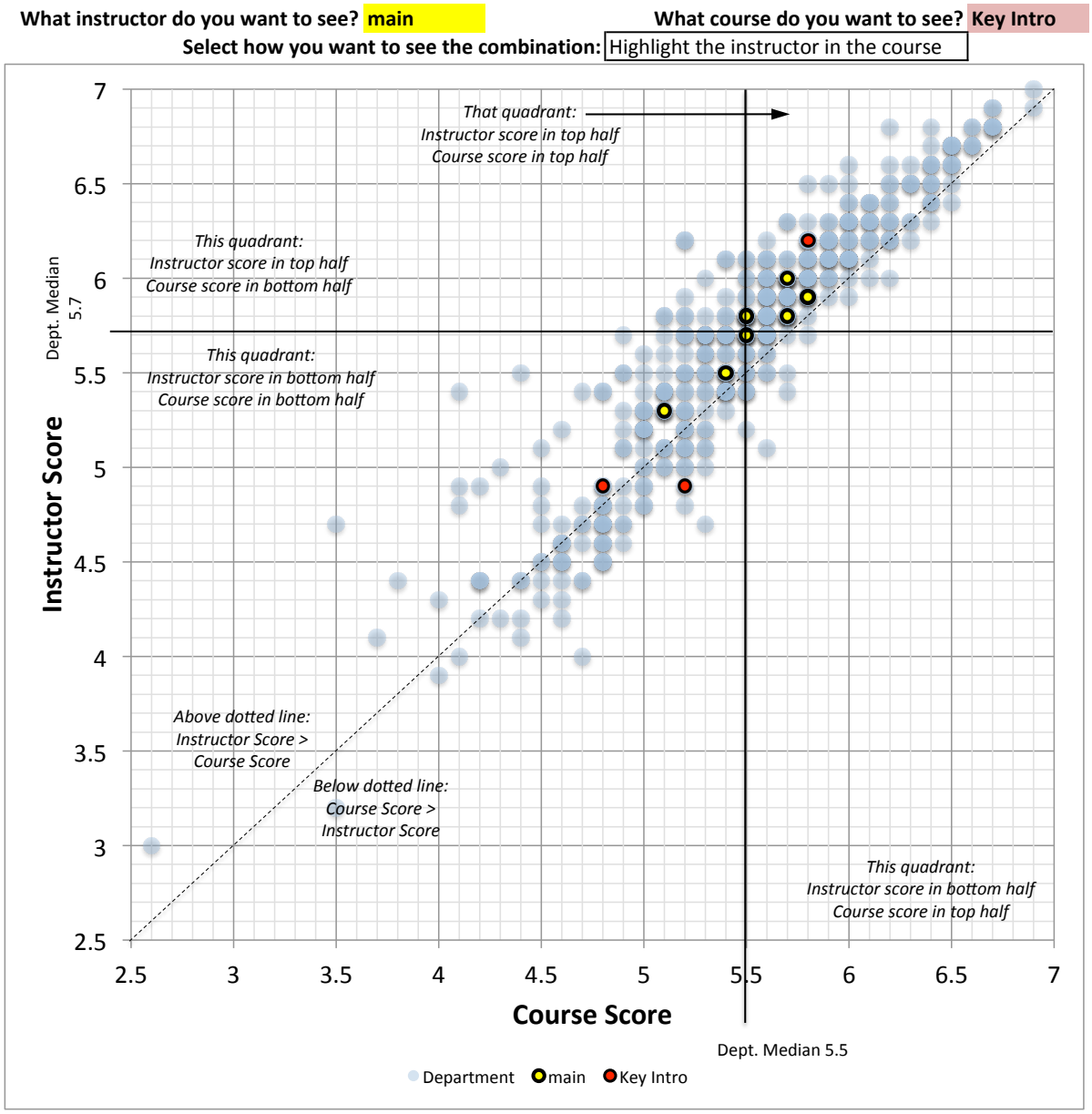
Hativa's models provide a useful starting point for considering the kinds of approaches that might be of value for identifying program-level insights from SRI data. Further visualization of the data would be of assistance to use it for performance review, faculty support, or programmatic analysis: the tools already described, from sparklines, to contextualizing ten-year frequency counts on course and instructor scores, to the more extended use of the Scatterplot and Longitudinal Teaching Profile Tool would be of considerable value in providing more meaningful ways to analyze and contextualize these data.

Visualizing Multi-Course Data for Program-Level Purposes

Example 1: Assessing Students' Perceptions of Courses

In this case, a key introductory course figures prominently in the department's sole program, and also acts as a service course to the university at large. It is a popular science option for students in other faculties. Because of its broad appeal, the course is used as a recruitment vehicle. The effectiveness of the course and students' impressions of it therefore play important roles in the department's strategic planning. By using the *Scatterplot Context Tool* for an individual course, rather than for an instructor, we can see where all instantiations of the course fall within the departmental context (Figure 9). In general they appear in the middle among the majority of the department's scores, however there are outer cases where the scores are worrisome. By highlighting a particular instructor within that grouping (the main instructor in this case) it becomes apparent that the outliers were individuals who taught the course on a one-time basis. This examination led to the realization that while many people in the department can teach the course, continuity of instruction seemed to make a difference for this strategically important offering.

Figure 9: Course Instantiations, using the *Scatterplot Context Tool*



Students' Perceptions of Courses: Extensions

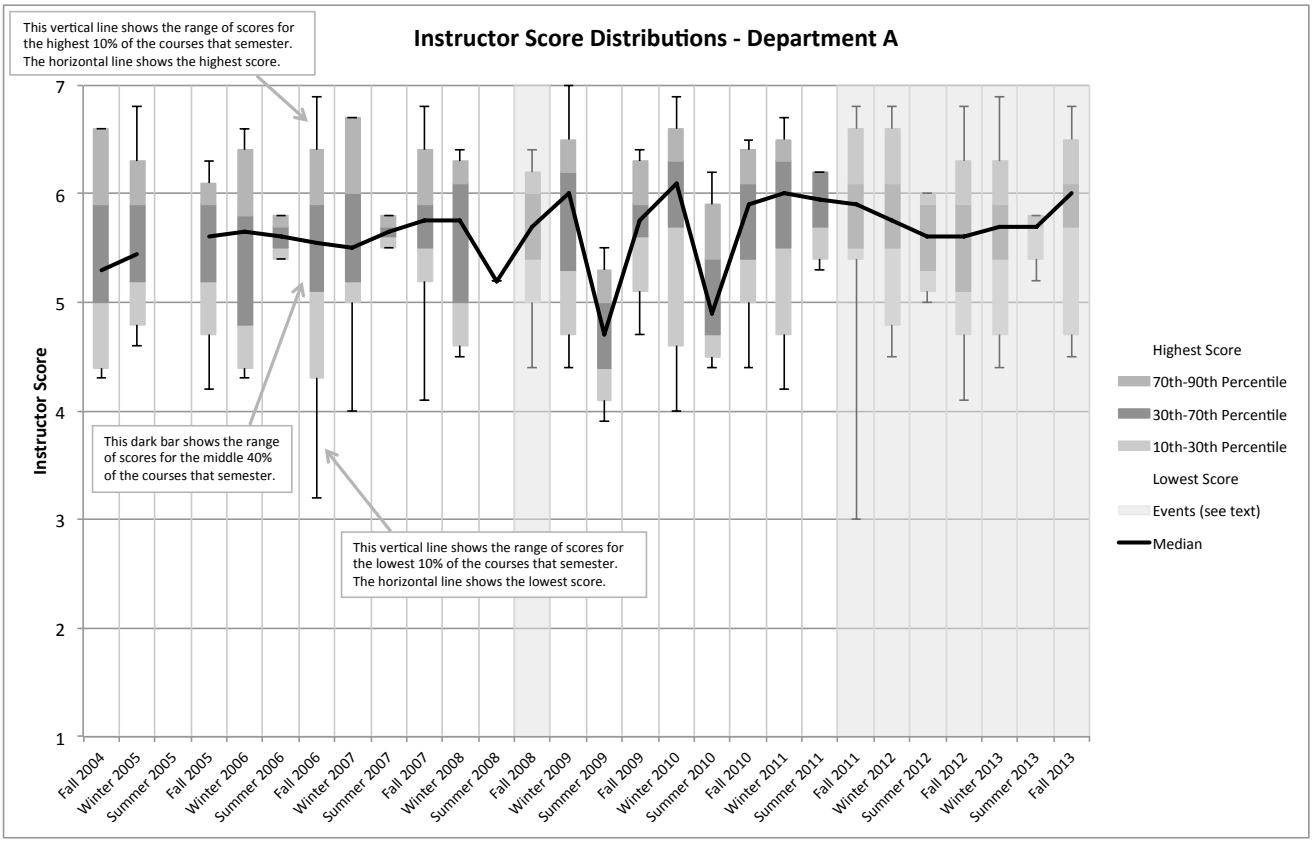
It could also be useful to look at scores for a course (or scores for multiple instantiations of the same course) based on student sub-populations, for example those who require the course compared to those who are taking it as an elective, or even sub-populations based on the home faculties of students taking the course as an elective. This kind of diagnostic might provide important information for course development or re-design, but two important factors must be considered. Firstly, the original survey data must be referenced to divide the respondents into appropriate sub-populations. This does not mean that individual results need to be made available, only a means to request and generate alternative versions of the report with specific filters applied to the collective raw data. The confidentiality and privacy of the individual responses would remain intact. Secondly, if either subgroup contains fewer than five respondents, neither part of the split – neither the students in the subgroup of

interest nor the 'other' students – can be reported in order to prevent identification of specific individuals. Although this kind of analysis can be very useful, it also invites a range of potentially counter-productive interpretations and assumptions about groups within a class. For this reason this filter is not recommended as part of the suite available to individual faculty members, but rather as part of the chair's suite. An individual faculty member can certainly explore the subgroups in collaboration with the chair, but a check-and-balance opportunity should be created to first understand the member's objectives and motivations.

Example 2: Visualizing Historical Scores for a Program or Department

Figure 10, generated by the Departmental Teaching Timeline Tool, is set here to show all instructor scores in a department by semester for a ten-year period, using box plots to represent the distributions. A central dark bar shows the middle 40% of scores, with a line identifying the median score. Vertical lines show the highest and lowest ten percent, capped at the ends with the maximum and minimum score respectively.

Figure 10. Instructor Score Distributions by Semester, using the *Departmental Teaching Timeline Tool*

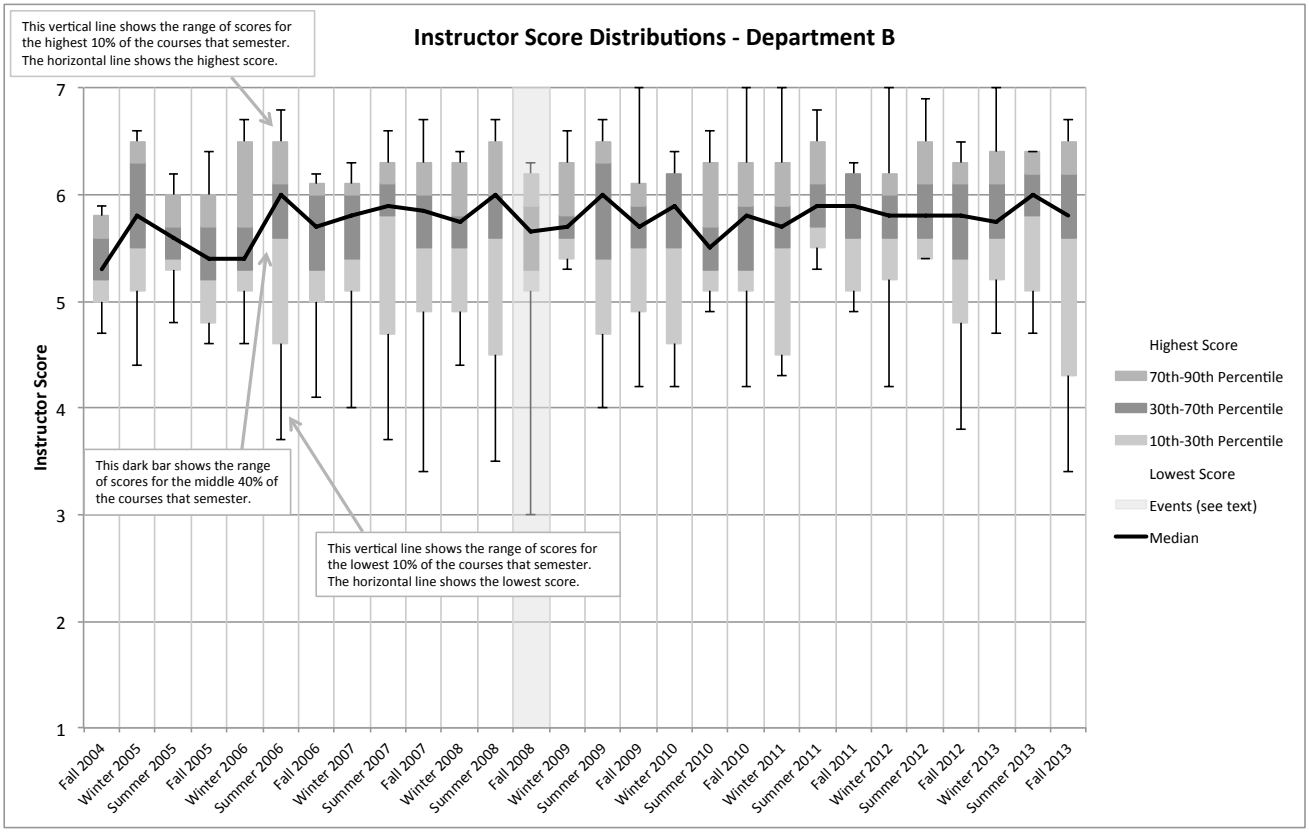


In this department's case the median drifts slightly upwards over the ten years, but the change is not very strong. The 2008-2011 period shows considerable variation which might bear further investigation. One working hypothesis was a connection with a labour disruption on campus in Fall 2008 (highlighted on the figure), but the scores do not seem to be particularly different from other years. The lowest spikes occur in summer: it would be worth investigating the mode of delivery, status of instructors, kinds of courses taught, student demographics, and of course students' insights about summer semesters, to more clearly understand and address this phenomenon.

Visualizations like this could help assess the effectiveness of academic development or program change initiatives. For example, beginning in Fall 2011 the department in question discontinued several programs to focus on only one core program and support for another external program. The core program was entirely re-designed at that time. The first intake to the new program began in 2011, and a year-by-year phase-in of the new program was coordinated with a phase-out of the old program to accommodate course needs for students in their respective program and year. This period of transition is also indicated in the graph. Looking at Figure 10, there appears to be a small temporary dip in scores centered on 2012/13, midway through the transition period when balancing students' course needs was most complex, but it is not very strong. A clearer sense of the impact of the new program will need more semesters of data. In order to verify if there is in fact a shift, other statistical tests are required (See Appendix A, 'Comparing two score distributions'). One reason that a program change may not be visible is that the majority of the courses are not attended by just the students in the program: disaggregation by major would be helpful in this case, or at least extracting the sub-group for whom the course is required. Phased in programming, as seen above, is another reason why trends may not be plainly visible: disaggregation of the population by student-year may provide a clearer pattern. A broader range of data sets, including quantitative data such as attrition rates, student course achievement and enrolment data, as well as qualitative data such as student feedback, open-ended survey information, and graduate interview data would significantly enhance the meaningfulness of this data set. This tool would also be valuable for quality assurance and accreditation reporting purposes.

Looking at this kind of distribution for another department (Figure 11), we can see that the range of scores each semester and the variability among semesters are different from those of the prior department.

Figure 11. Instructor Score Distributions by Semester: Department B, using the *Departmental Teaching Timeline Tool*



For example, the middle scores tend to be clustered around a much smaller range of values, and the median is slightly higher and more consistent. However, in this particular program, when students are dissatisfied they provide considerably lower scores, as shown by the long lower bars and long lower vertical lines. It is impossible to know what factors impact scoring of this kind: it does point out that it is important to make comparisons only to the appropriate norm group, and to do the work required to identify what those norm groups are.

Visualizing Historic Scores: Extensions

The tool used to produce figures 10 and 11 might also provide useful for visualizing subscale data, and for examining satisfaction of students in different years of a program. Such data might provide useful indications of the effectiveness of initiatives to improve instructor accessibility in a given year of a program, or to track student perceptions of workload or other areas that might be identified as problematic through SRI analysis. As in all of these cases, the quality of the original SRI instrument, its use of robust subscales, and the use of multi-faceted data rather than just the SRI, are critical to the meaningfulness of conclusions drawn. However as an initial step in a more inquiry-based approach, this tool could prove fruitful.

The Use of Visualized Aggregate Data for Faculty Development

Although no universities in the survey conducted in conjunction with this report indicated that Centres for Teaching and Learning (CTLs) had access to SRI data, other jurisdictions use these data differently (Joughin & Winer, 2014). Without advocating for a further level of surveillance of faculty practice, it is reasonable to imagine that individual instructors might authorize CTL consultants to review SRI data in the course of personal consultations, teaching dossier development and so on. All of the individual instructor tools above would be of use for guided reflection, the establishment of inquiry and action plans, and goal setting with faculty. These visualizations, while not definitive evidence of classroom practice, can form a strong basis for dialogue and exploration. On a broader and perhaps more ambitious basis, access to the kinds of programmatic tools described here, which would not offer access to individual instructor information without permission, would provide CTLs with mechanisms to identify thematic areas of need for the development of instructional improvement programming. Of course, CTL involvement would be critical to the establishment of any aggregate data initiative on a university campus, in terms of setting priorities, engaging faculty with the initiative, providing leadership and insight, and ultimately working with faculty and administrators on effective use of the new tools.

Institutional Matters: The Use of Data to Improve Validity and Transparency of SRI Processes

There is a broad, long-running, and contradictory literature examining factors that may bias SRI results. A variety of factors may impact student ratings of instruction, including class size, course difficulty, course level, faculty, and delivery method (Clayson, 2009; Galbraith, Merrill & Kline, 2012; Hativa, 2013; Langbein, 2008; Patrick, 2011; Slocombe, Miller & Hite, 2011; Sullivan, Polnick, Nickson, Maninger, & Butler, 2013), a fact that is further complicated by tendencies to misinterpret or misunderstand SRI data (Theall & Franklin, 2001). Many commonly-held beliefs about various kinds of bias have been disproven. However, many myths have shown surprising persistence among faculty members, and these affect willingness to fully engage the potential of SRI data.

Given this context, the proactive institutional use of SRI data is important. McGill's process is underpinned by several years of detailed analysis of various factors that faculty believed produced variations in student SRI response, some of which were shown to have a weak influence, others none at all. Their annual reporting, which provides scoring in the context of several norm groups, continues to

report these factors with no influence in order to strengthen faculty understanding of the legitimacy of the comparisons being made. Factors that do appear to create significant differences in response patterns are also made public, allowing faculty to document their teaching practice with greater transparency and accuracy. Among the uses to which these data can be put for the purposes of institutional fine-tuning:

- validation of SRI instruments;
- validation of scales and sub-scales;
- assessment of significant differences in sub-populations which may differentially impact SRI responses;
- assessment of significant differences in types of courses which may differentially impact SRI responses; and
- further research into patterns within the data which might inform strategic planning, identify groups of students requiring better support across various years or programs of study, and so on.

Core Recommendations

This report reflects the considerable potential benefit of developing tools that would make it easier for instructors and administrators to visualize and analyze SRI data. To date we have built a number of flexible prototype tools and explored their potential use with real, but anonymized, SRI data sets. Although the tools are automated, in that they draw data directly from tables to produce the visualizations, they are not yet automatic. For these to become part of everyday use among regular instructors and administrators, they need to form part of a better-mediated tool suite with a user-friendly interface and a managed connection to the institution's SRI data repository.

Based on our work to date and discussions with the project's technical team, it is reasonable to envision a dashboard-based suite of tools that offer program chairs easier access to programmatic data, and instructors access to visualization tools that allow for better analysis and decision-making. Figure 12 provides one potential approach that would be a significant improvement over the current situation: of note, the dashboard provides two 'thumbprint' representations which could be customizable, as well as suggested starting points written in plain English. Each prompt guides the user to make a further selection from among a range of visualization options, and then presents an interactive interface to explore the courses, instructors, and time period they wish to study. Of course such a system requires customizable hierarchical data access structures to ensure that practice conforms to institutional regulatory contexts, as well as a solid security plan.

Figure 12. A Preliminary Sketch for a Program-Level SRI Data Dashboard

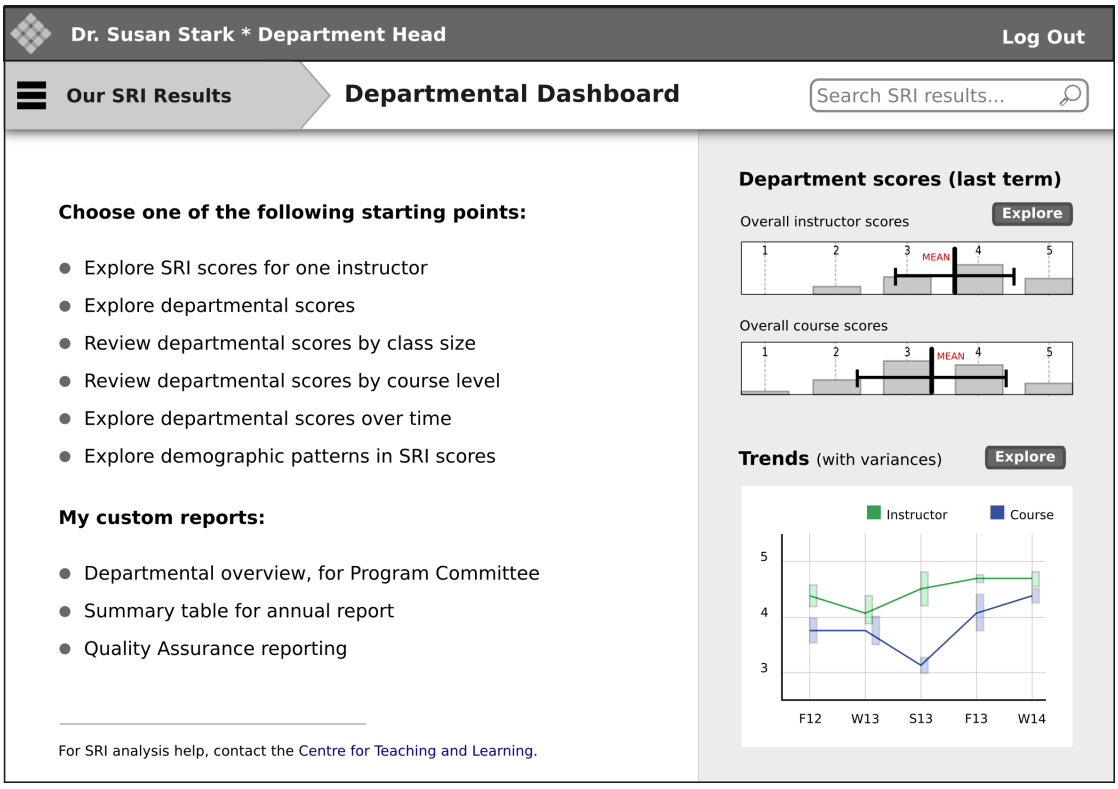
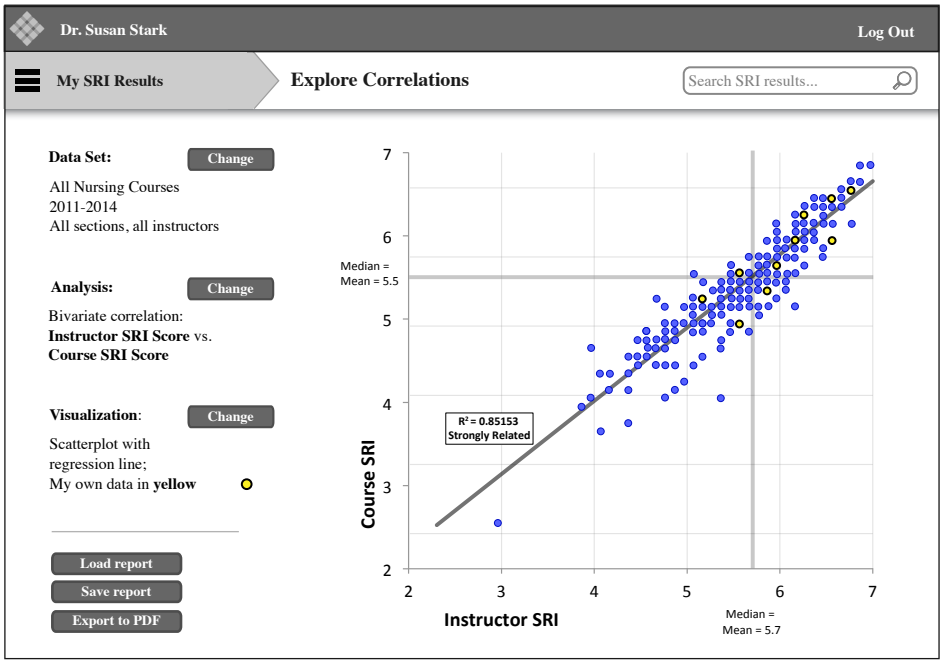


Figure 13a: Aggregate Data Instructor Tools



As can be seen in Figure 13a, data sets, visualizations, and analytical tools can be chosen and changed; the visualizations produced can be emailed, saved, or exported as PDF for further use; and an annotation tool can be used to contextualize data.

Figure 13b: Aggregate Data Instructor Tools

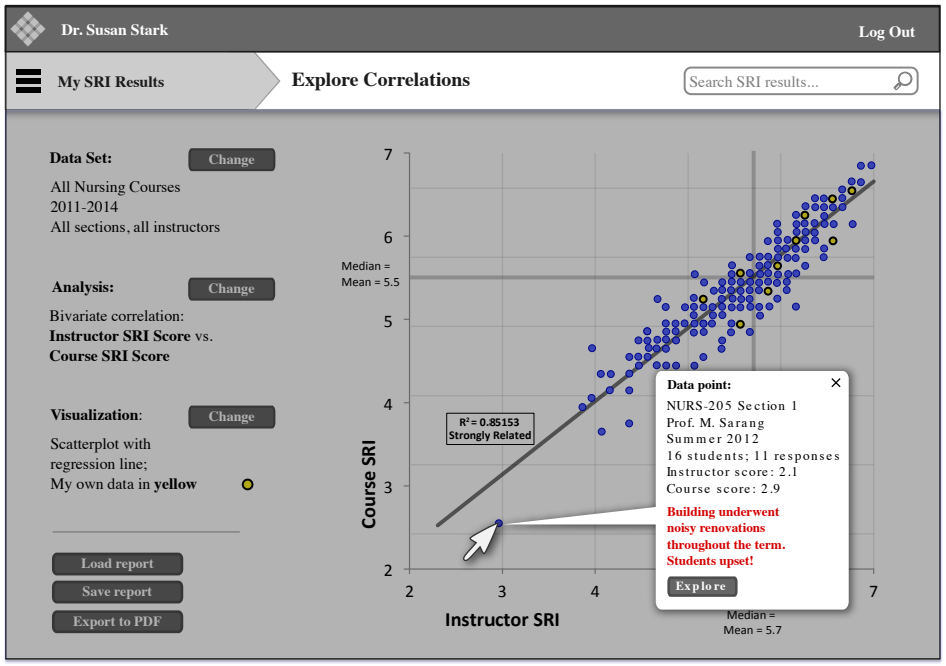
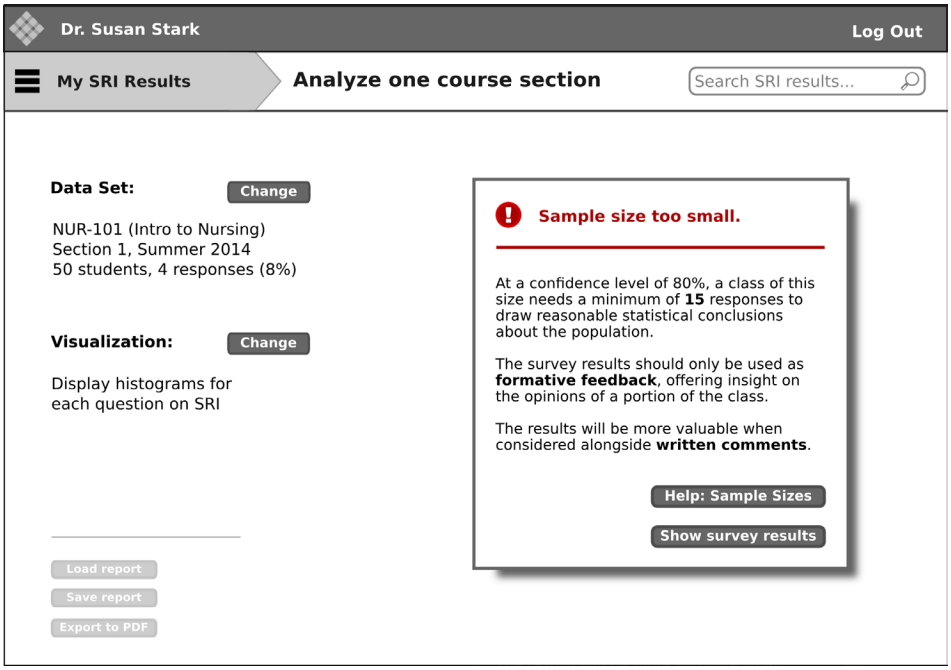


Figure 13b illustrates the use of annotation tools to ensure that instructors and program chairs can both contextualize their course scores and make notations related to points of inquiry or interest.

Figure 14: Reducing the Chance of Misinterpretation



As shown in Figure 14, an important value that a tool of this kind should offer is an ability to limit potential for misinterpretation or inaccuracy, while also raising awareness of the kinds of limitations to SRI data that were raised earlier in this report.

There is much work to be done before such a tool can become a satisfactory reality, from negotiating the political challenges of SRI change in universities, to establishing a broader range of visualization and analytical tools and assessing their usefulness with actual users, to addressing the numerous technical challenges of creating a system robust and flexible enough to work with the varied data sets, information system structures, and regulatory contexts in Ontario universities. An assessment of the technical implications of such a project can be found in the Project's Technical Report (Appendix G). However, there is also much to be gained: our view is that there is a compelling case to be made for developing these tools as a catalytic tool for fostering a stronger culture of inquiry-oriented teaching improvement in the Province of Ontario.

Individual instructors have a lot 'riding on' the outcomes of SRI, and there is considerable distrust of the process. The same will likely be true of efforts to launch aggregate data initiatives. Anecdotal evidence to date suggests concerns that such tools might enable comparisons among institutions, or other comparative activities that are both unsound and appear threatening. It is therefore critical that these initiatives be undertaken with a strong degree of transparency and consultation. Alderman (2013, and in conversation, June 18, 2014) ascribes the success of the Queensland University of Technology's REFRAME program, which completely revamped their approach to teaching evaluation, to its rigorously client-driven approach. While it is possible to build tools that process data in any number of ways, the critical factor that will determine adoption is faculty faith in the usability and helpfulness of the data and an overall (and likely hard-won) belief that the true goal of the initiatives is to help them and to improve, rather than further regulate, teaching. We therefore strongly advocate for the kinds of highly consensus-driven approaches reflected by Winer et al. (2014) and Alderman and Melanie (2012, 2013).

Further Recommendations

This report has outlined ways in which aggregate data and visualization tools can improve instructors' and administrators' capacity for documenting teaching, inquiring about teaching, and improving teaching.

- The potential of aggregate data to improve teaching, courses, and programs requires further research, both technically and in terms of how faculty and administrators can be most effectively engaged in aggregate data use. Although there is suggestive evidence that it can make effective contributions to instructor and program level teaching improvement and decision-making practice, optimizing its potential requires more research and development.
    - Dimensions of this research should include the use of aggregate data to better understand context, to examine longitudinal patterns, and to explore relationships along various dimensions and among various sub-groups. This would include norm groups of similar class types (e.g., large enrolment, by year); disaggregated student populations (e.g., required vs. elective); and different validated subscales.
    - It should also include the question of the potential impact of aggregate data usage on the use and perception of SRIs (engagement, perception, uptake, resistance, misperceptions).
    - It should also include research into user experience interface design based on the principle that users will have varied levels of understanding of statistics and data management.
- Faculty-driven approaches should be adopted: iterative exploration and development will contribute to an understanding of needs, problems, and possibilities that may be critical to the success of such project.

- Under no circumstances should the provision of aggregate data tools be seen as an alternative to the adoption of a more multi-faceted approach to data collection, analysis, and evaluation. Aggregate data analysis can broaden the use and nuance of these data, but they remain a single point of evidence in what should be a range of approaches.
- All tools built must take into account the possibility of their use by people with limited statistical knowledge. Statistically invalid comparisons and conclusions should be disallowed by the system, and an explanation provided. Several critical factors must be kept in mind:
    - Data comparison across faculties or institutions is methodologically unsound owing to significant variations in participant populations and environmental factors
    - Sample size and response rates interact, and are relevant to determining the margin of error and valid use of the collected response data
    - Statistical comparisons regarding 'growth' or 'improvement' in scores are neither valid nor defensible
    - Statistical calculations cannot be employed on single items in SRIs: a subscale of minimally four but preferably eight items are needed to yield usable information
- The system will need embedded learning tools such as glossaries in user-focused language to support the growth of understanding. This system should be designed with actual users as part of the team.
- A legal review of the full implications of FIPPA and other regulations should be implemented as one of the earliest components of first-phase development.

While the mandate of this project was to examine the potential of visualized aggregate SRI data, it is apparent that there is clear potential for broader development using more integrated academic data sets. These could include, for example, student indicators such as achievement, enrolment, success, and perseverance data. Our experience with this study suggests there is much to do within even the relatively narrow confines of SRI data, and that engaging faculty with data tools may require incremental phases. Integrated data access and use are certainly important long-term goals: a further stage of this project may contribute to a longer-term plan of this nature.

SRI Data: An Overview

Phil A. Graniero

Appendix A
SRI Data: An Overview

Terminology

Many of the misunderstandings and misapplications of SRI data and analyses stem from vague understanding of terms, coupled with their ambiguous use in handbooks, guides, and reports (Carifio & Perla, 2007). Another challenge is that many of these terms have precise definitions in statistics, but have other meanings in every day parlance. These linguistic confusions can lead to misperceptions that have serious implications for how data are used and misused. To avoid this problem, we provide a brief overview of relevant statistical terminology for SRIs.

Levels of Measurement

The notion of 'levels of measurement' (sometimes called 'scales of measurement') was first introduced by Stevens (1946). Considerable confusion about levels of measurement has been the root of much of the misapplied methodology and misinterpretation of results in SRI practice (Carifio & Perla, 2007). There are four primary levels of measurement:

Nominal-Level Data

Nominal-level measurements are distinguished by their membership in a qualitative, named category, but there is no meaningful ordering of the categories based on rank or magnitude. Examples include gender, full-time / part-time status, or home faculty.

The only meaningful mathematical operations that can be applied to nominal-level data are related to set membership, for example comparing sizes of groups. The only meaningful measure of central tendency in a distribution of values is identification of the *mode*, or most common value.

Ordinal-Level Data

Ordinal-level measurements are distinguished by their membership in a qualitative category that may be meaningfully sorted by a rank ordering. Examples include level of enthusiasm (low, medium, high), level of seniority (junior, intermediate, senior), letter grade (F, D, C, B, A), and level of performance (poor, adequate, good, outstanding). It is meaningful to compare a category value as *bigger* or *better* than another, but it is not possible to state the *degree of difference* between them; that is, it is not possible to say *how much* bigger or better in a precise way.

Like nominal-level data, set membership operations are the only meaningful mathematical operations on ordinal-level data. The central tendency of a distribution can be described by the *median*, or the category of the middle-ranked item in an ordered set of values, or by the *mode*, the most common item.

Interval-Level Data

Interval-level measurements are quantitative or numeric measures that distinguish the *difference* and *relative direction* among items; one unit of difference between two items always means the same thing regardless of the size or magnitude of the items. However, the *ratio* between two items is not allowed since the zero point is arbitrarily set. An example is date of entry into a program. It is possible to say that a student who started in Fall 2010 started twenty years after a student who started in Fall 1990, but it is not possible to say that the student started 1% later than the other.

Addition and subtraction may be used on interval-level data, but multiplication and division may not. The central tendency of a distribution can be described by the *arithmetic mean* or *average* as well as the *median*. If the measurement uses discrete values (for example, integers but not decimal numbers) then the *mode* can be used as well. The spread or dispersion of the distribution can be described using the *range* (the difference between the maximum value and the minimum value) and the *standard deviation* when it is statistically valid, for example when the set fits a *normal* ('bell-shaped') distribution.

Ratio-Level Data

Like interval-level measurements, ratio-level measurements are quantitative or numeric measures that distinguish the *difference* and *relative direction* among items, but they also have a unique zero value so the *ratio* between two items is allowed. That is, it is meaningful to specify 'how much' or 'how many' of something with ratio-level data, and to say that one item is 'twice as big' as another, or one item has 'twice as many' as another. Examples include number of completed courses, or age.

All usual mathematical operations may be used on ratio-level data. The central tendency of a distribution may be described by the *mode*, *median*, and *mean*. The spread of the distribution may be described by the *range,* the *standard deviation,* and other statistical measures when valid.

Parts of a Survey Instrument

A survey **instrument** is composed of several questions called **items**.

A **respondent** is a person who participates in the survey by giving responses to each of the items on the instrument.

An item has a **response format**, or a defined, structured way in which the person can provide their response. As an example, a common SRI item is a question that asks the respondent whether they were required to take the course; another item asks the respondent about their initial enthusiasm for taking the course. For each item the respondent is presented with a range of ordered **response categories**, such as (*yes, no*) in the first case and (*low,medium,high*) in the second.

A **Likert item** (Likert, 1932) is a statement that the respondent is asked to evaluate according to their subjective opinion or impression. A true Likert item has a specific response format: the response categories should be symmetrical around a neutral value, with an equal number of choices representing intensity of feeling about the item on each side, such as (*strongly disagree*, *disagree*, *neutral*, *agree*, *strongly agree*).

Likert-type items: Most items that are called Likert items are not, since they do not have the Likert response format. Rather, they present a range of choices to rate some quality or characteristic of a subject. Examples include rating performance, such as (*unacceptable, poor, adequate, good, excellent*); rating frequency of occurrence, such as (*never, occasionally, sometimes, often, always*); or rating similarity, such as (*not at all, a little bit, somewhat, very much, completely*). They tend to have five or seven choices, using a set of categories that place a 'middle' label in the center position, balanced by an equal number of increasingly 'negative' items and increasingly 'positive' items on each side, anchored by extreme ratings at the ends. These items are referred to as Likert-like or Likert-type items to distinguish them from Likert items.

The **response coding** is how the response is recorded for storage and analysis. The response format is how the information is *captured* from the respondent; the response coding is how the information is *stored*. In the case of the question regarding whether the course was required, the response format was (*yes,no*) but the response coding in the database may be the letter 'Y' or 'N' respectively. This difference between the response format and the response coding is an important issue later in the discussion.

A **Likert or Likert-type scale** is a respondent's collective responses from several items with identical response formats, where the set of items are designed to make up a reasoned, cohesive, and complete examination of a single multi-faceted subject. A standard response coding for a 5-point Likert-type item that is part of a scale is to use the numbers 1 through 5 to represent each category in rank order. The number 1 is recorded instead of '*very poor'*, 2 instead of '*poor'*, and so on. To calculate the scale, the encoded numbers for each item's response are summed. Sometimes

the sum is divided by the number of items in the scale. It is the same numerical calculation as a mean, but the purpose is to make the magnitudes in the scale similar to those in individual items for easier comparison and interpretation.

The items that comprise a scale can be split into separate **subscales**. The set of items comprising the subscale should also be a reasoned, cohesive and complete examination; in this case each subscale examines one dimension of the larger subject. The items on an SRI capture students' perception of their experience in a course. Although all of the items may be combined into one scale, it is more common that one group of items are combined to address the effectiveness of the instructor, and another group of items are combined to address the effectiveness of the course being taught. Other subscales may address particular elements of instruction.

SRIs are designed to capture a group of students' responses; the class' collective response is of interest, not the individual responses. An **aggregated item** is represented by a **response distribution**: an array of numbers with a count of the number of respondents who chose each response category.

An **aggregated scale** is the collective responses from a set of aggregated items. The values from the response distributions for each aggregated item are summed. Note that the scale for an individual response is a single number, whereas the aggregated scale is a response distribution.

The **score** for an aggregated scale is a single summary number representing the response distribution. Any measure of a distribution may be used, but a measure of central tendency is usually the most useful. The valid measures that may be used depend upon the level of measurement associated with the items' response format. The measure used to calculate the score is often also applied to the aggregated items comprising the scale. The selection of the distribution measure used for the score will also become an important issue later in the discussion.

Types of SRI Items

There are two basic types of items on an SRI: *demographics*, which describe the students who responded to the survey; and *ratings*, which describe the students' perceptions of the course and instruction.

Demographic items identify broad characteristics of the students. Typical items include:

- Required or elective course
- Gender
- Current year in program
- Home faculty or department

Demographic items are used in two main ways:

- to describe the heterogeneity of the students in the classroom, which could provide context when interpreting the scores and range of variation for the scores in one course offering, or when comparing the classroom context of two different courses.
- to divide the class into sub-groups according to one or more characteristics, in order to see if there is a difference in how the sub-groups responded.

Rating items are the core purpose of the survey. Typical items include:

- questions regarding the student's perception of the effectiveness of the instructor
- questions regarding the student's perception of the effectiveness of the course

Some items may be used as ratings or demographics, such as enthusiasm for the course. Course enthusiasm is generally measured using ordinal categories (e.g. *low, medium, high*). It is most useful when students are asked for both their enthusiasm at the beginning of the course and their enthusiasm at the end of the course. Enthusiasm at the beginning measures pre-conceived notions of what *might* happen. Comparing the responses to both items can identify *shifts* in enthusiasm, which is influenced by what happened in the classroom. Initial enthusiasm might be considered a demographic item; together, they produce a rating item.

On most SRI instruments students are required to select one option when presented with several choices for an item. This implies that the student belongs in only one category. This may be a reasonable assumption for some characteristics such as full-time / part-time status in a semester, but may not necessarily be the case for other characteristics. For example, a common SRI item is the home faculty or department of the student's major. In one study (Graniero & Hamilton, 2011), 10% of students were in double-major programs; almost all of these students have two home departments, and for 15% of them, two home faculties. If only one selection is allowed, then the student is forced to choose and indicate only one home. How the student decides to select one is arbitrary, and not all students in the same double-major program will make the same choice; therefore, the power of using the item to compare ratings between students from different departments is weakened.

Common Reporting Formats for SRI Data

Much attention will be given in later sections to the matter of reporting SRI data, but a brief orientation to common reporting practice is given here. Generally speaking (and depending on labour and academic governance agreements), faculty and administrators receive SRI data at the course level, usually including itemized reporting, often with a small number of higher-level scores, for each course (Table 1). Administrators may also receive reporting at the programmatic or departmental level. Scores may be accompanied by means and standard deviations, or the average scores in the department for that semester may be provided. There are wide variations in practice, with some institutions providing ranked scores for the semester, others median scores or means, standard deviations and so on.

Although means and medians can be useful to create summary information for quick review, response distributions are critical to a full understanding of the variation in students' experiences in the class and therefore to proper analytical and interpretive use of the data. For example, a bimodal distribution with many students at each extreme has significantly different implications than a tight clump of scores in the middle, even though they produce the same means or medians. The issue of using distributions will arise regularly through the report.

Appendix A
SRI Data: An Overview

The Importance of Score Distributions

We commonly use - and think about - the result of the SRI as two numbers: a mean score for all of the instructor-related items and a mean score for all of the course-related items, calculated from all surveys items from all respondents. It is easy to forget that those scores are actually just a compact and collapsed representation of the survey results. In reality, the SRI captures a distribution - the range of the class' varied experiences and their assessment of those experiences - and the nature of this distribution is important.

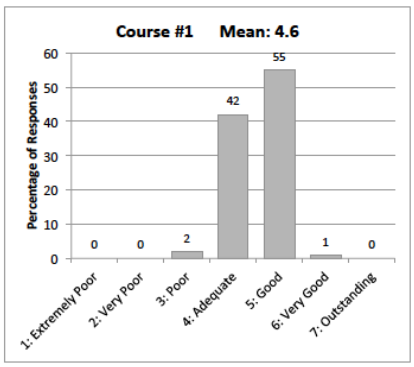
Variation in responses can happen for a variety of reasons directly related to the course and the instructor, given the many dimensions of course design and instructional practice, the complexity of the teaching and learning process, and factors personal to the student and unrelated to the course. Although the single-number mean scores are helpful for certain high-level perspectives and purposes, the range and variation of responses is important and should not get lost. Consider two different courses taught by two different instructors.

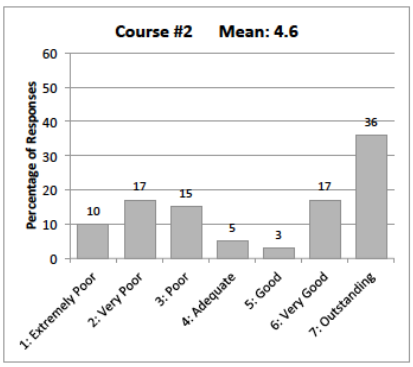
In Course #1, the instructor does a competent job creating a conventional course. It focuses on fundamental facts, delivered in a clear but not inspiring fashion. The class completes standard tests and assignments that are returned quickly, with feedback limited to textbook page references beside incorrect answers. The instructor is friendly but somewhat distant. Students feel comfortable asking questions and getting answers, but don't expect intriguing conversations. Nobody is left behind, but nobody is lifted up either.

In Course #2, the instructor tries to create a unique learning experience. The instructor takes a student-directed approach, regularly seeking input from the class to guide the topics, which some students find empowering. Others, who prefer a pre-determined structure, view it as 'winging it'. The instructor presents the topics through the lens of a controversial real-world problem to create opportunities for critical thinking, discussion and reflective examination. Some students perceive this as an opportunity to develop personally meaningful and intellectually informed thinking; others simply find the topic offensive. Return of assignment submissions is protracted, but margin notes contain detailed, useful formative feedback and suggestions. The instructor has an open, quirky and sharing personality, which students respond to in varied ways. For some students it is the course that they remember as a turning point in their school career; others are just happy when it's over.

The SRI score distributions for the two courses likely look similar to those shown in Figure A1. It is also likely that many of the detailed items on the survey will show similar distributions.

Figure A1: Distribution of SRI scores for Courses 1 and 2.





In Course #1, the experience and rating is fairly universal among the students. Given a 7-point scale with labels like the ones shown here, average scores tend to center on 5 (*good*) or higher across the institution. Students view this course as average, at best. In Course #2, the experience is far from universal. One third of the class gave the highest rating, which is a difficult accomplishment. However, it is rare in practice to see more than a small proportion of students giving the lowest possible ratings, and this is a large concern. The distributions are dramatically different, but the mean score is identical. If only the mean scores are considered and the distributions ignored, then there is no way of distinguishing between these two very different courses, instructors, and student experiences. Likewise, If only the mean scores are used in evaluation, assessment, and development activities, and the distributions are left out, salient information is missed to potentially hazardous effect. The challenge is to find a way to balance the richness of the data with forms that simplify the understanding of patterns.

Principles of Data Aggregation and Comparison

Determining what can, and cannot, be compared and how to make methodologically sound comparisons is probably one of the most difficult aspects of creating effective SRI reporting and facilitating decision-making informed by SRIs. The reality is that most of the users of SRI data do not have sufficient expertise to make statistical design decisions, to identify elements in a set of data that do not provide a valid basis for conclusions, or to determine if the data are representing differences that are not just numerically different, but statistically significant (Hativa, 2013b; Winer et al., 2012).

Even with sufficient statistical design expertise, it can still be difficult to design an analysis that does not violate the assumptions that statistical tools rely upon. The typical practicalities involved in administering SRIs in a course tend to violate many assumptions regarding representative samples. Furthermore, comparative analyses rarely have a sufficient number of data points due to the wide number of factors impacting the similarity – and therefore comparability - of courses and offerings.

These methods and challenges will be described in more detail below.

Appendix A
SRI Data: An Overview

Sample Size and Response Rates

It is unusual that an SRI is completed by 100% of the students in a class: in fact in the survey of Ontario universities completed as an element of this study, student response rate was the most identified concern regarding SRI use. If the response rate is less than 100%, those who did respond are effectively speaking for those who did not. In order to trust the results of the SRI, we must have some confidence that those who responded are likely to be representative of the group as a whole. This concern gets greater as the response rate gets lower. If the majority of the class found the course to be highly valuable but those students were at an on-campus conference the day the SRI was administered, and the half-dozen respondents were those least engaged with the course, the SRI data are of questionable value. Knowing that the effect of response rate on results will always exist, we can try to control our sampling to minimize the effects, or at minimum, to be aware of its impact.

Sampling involves pulling out a subset of a whole population who are sufficiently representative of the population as a whole that any conclusions drawn from the subset can be generalized to the population as a whole. Any sampling will incur some sampling error (the extent to which a statistical measure applied to the sample gives a result that deviates from that of the population as a result of random variation in the sample membership) and possibly some sampling bias (when the result of a statistical measure applied to the sample deviates from that of the population as a result of systematic bias in the sample selection). Any statistical measure also reports measures of likely accuracy given the sample size, which can be compared to a *confidence level* decided upon before the analysis. This effectively helps the decision-maker address two important questions: "How likely is it that I would get the same answer if the entire class had responded instead of just the sample I have? Given the likelihood that I would *not* get the same answer, am I willing to accept and use this result?"

There are three different approaches to sampling governed by who is provided access to the SRI:
1) the entire class (the entire class is the population);
2) a subgroup based on some characteristic of interest (subgroup of interest is the population);
3) a subset of the class based on practicality or size management (the entire class is the population).

Regardless of the method, it is unlikely that all students will respond. In these cases two questions must be asked (Nulty, 2008):

- Do respondents differ systematically from the non-respondents?
- If so, will those differences cause them to respond differently to the questions asked?

If the answer to both questions is yes, the sample is biased, and extrapolation of analysis results from the sample to population is not strictly valid. Systematic differences between respondents and non-respondents have in fact been noted in several different ways (Goyder, 1987; Richardson, 2005), and in particular for students, in their attitudes and behaviour (Goyder, 1987) and in their study behaviour and academic attainment (Astin, 1970; Neilsen et al., 1978; Watkins & Hattie, 1985). Other factors such as disciplinary differences in response patterns, gender differences in response patterns, and student year might affect the representativeness of a given sample (Hativa, 2013b). We must therefore unavoidably deal with bias, which may result in errors in interpretation and decision-making. Issues of bias in student ratings of instruction, however, are hotly debated, with considerable evidence on each side of the debate (Hativa, 2013b). One critical element of establishing the validity and reliability of SRI on a given campus is the ongoing analysis of context-specific data to identify or disconfirm theories about bias within student responses (Winer et al., 2012; Joughin & Winer, 2014).

Appendix A
SRI Data: An Overview

For any of the sampling methods, a hidden source of bias is the way in which the survey is administered. A prime example is offering the survey online rather than on paper: web users are demographically different from other users (Watt et al., 2002), and variance in data from web surveys can be less than for paper surveys (Salmon et al., 2004). For example, those who complete an online survey are more likely to be familiar with and comfortable using online tools. They may be more likely to respond favourably about the online teaching and tools associated with the course. They may be a minority of the class, yet make up a majority of the respondents and therefore the majority of the feedback. The medium here may interact with characteristics of the respondent population, therefore creating a data set that is not representative of the class as a whole. This is an example where institutional data analysis to assess systematic differences in response patterns would be of value.

Nulty (2008) summarizes recommended minimum response rates in social science research in general, educational surveys in particular, identifying acceptable response rates ranging between 50% and 70%. He notes, however, that most of these recommendations are based on a balance of rational, practical, and political factors without a strong grounding in statistical theory. It is possible, however, to derive calculated minimum response rates for groups of different sizes based on statistical theory.

For instance, Dillman (2000) summarizes the theory to calculate the minimum sample size:

- that is required for a specified level of confidence in the result;
- in relation to a population of specified size;
- with a specified degree of sampling error; and
- given a specified level of probability for a particular answer to be provided by a respondent.

In other words, there is a sliding scale of necessary response rate, based on the size of the class, the analysis you wish to do, and the degree of uncertainty you are willing to accept in the results. Although the framework is straightforward, implementation is less so, as it requires institutional review of data to define statistical parameters as well as significant policy decisions.

Nulty (2008) provides a summary of required response rates by class size based on the statistical formulas and discussion in Dillman (2000), using reasoned approximations of the elements above. Table A1 provides Nulty's model, while Table A2 provides examples of how the approach is adapted and employed at two Canadian universities.

TABLE A1: Minimum sample sizes and response rates for selected class sizes, drawn from Nulty (2008). Calculations based on a 10% sampling error and an 80% confidence level that the sample size is sufficient (both policy decisions), and a typical 70:30 split between responses of 4 or 5 compared with 1, 2, or 3 (determined from an institutional study of historical SRI scores).

Class Size	**Minimum number of responses**	**Minimum response rate**
Less than ten	Five	---
10	7	75%
30	14	48%
50	17	35%
100	21	21%
200	23	12%

TABLE A2: Example minimum response rate policies for two universities. Each university adapted different calculation formulas, made different assumptions, and chose different statistical parameters based on the characteristics and reporting needs of their specific institution.

McGill University *Adapted from Nulty (2008)*			**Memorial University** *Adapted from Dillman et al. (2007)*	
Class Size	**Minimum Response Rate**		**Class Size**	**Minimum Response Rate**
5-11	Minimum 5 responses		10-15	67%
12-30	40%		16-20	60%
31-100	35%		21-30	47%
101-200	30%		31-40	40%
201-1000	25%		41-60	30%
			>60	25%

In practice, several requirements or assumptions of the model are typically not met, such as the requirement of a random sample. Although this is not an ideal solution to the challenge of response rate and sampling bias, it is a significant methodological improvement over ignoring the problem altogether.

Aggregation Methods

For the type of applications discussed in this report, the results from all respondents to a single SRI survey – that is, the results from all students in a single course offering – are aggregated together, forming the basic unit of analysis. Aggregated items, aggregated scales and scores are used, not individual responses. These course-level survey results are then aggregated together in various ways: aggregated in the sense of using a collection or selected group of surveys together; and aggregated in the sense of representing specific survey items in that collection in different ways.

Ways to select surveys for a collection:

- all offerings of a particular course
- all courses taught by a particular instructor
- all courses from a particular department
- all courses from a particular faculty
- all courses taught in a particular semester
- all courses taught in a particular year or time range (e.g. past ten years)

Collections may be created by selecting surveys using more than one of these characteristics, either by limiting the collection to results that match all criteria, or by creating sub-groups within a collection. For example, one might be interested in examining all offerings of a particular course taught by a particular instructor, a collection limited by instructor identity and specific course. Another example would be collecting all scores for a specific department, over ten years, grouped by course. In this case the collection is limited by department and time range and grouped by course identity.

Ways to represent an item or scale in an aggregate collection include:

- tabular, statistical, or visual representations of a response distribution
- a summary number representing the distribution of scores
- tabular, statistical, or visual representations of the distribution of scores

- a time series of scores (scores placed in chronological order)
- a single response distribution aggregating the response distributions for an item from all surveys

Aggregating several items into a single scale

As noted earlier, a *scale* is the collective responses from a set of items with identical response formats, where the set of items are designed to make up a reasoned, cohesive, and complete examination of a single multi-faceted subject such as overall instructor effectiveness. Also, a subset of the items that are included in a scale may be used to form a *subscale*, which represents one component or dimension of the overall scale such as presentation skill, organization, or approachability. When looking at SRI data aggregated for an entire class instead of an individual response, the scale is a *response distribution*, but it is typically collapsed to a *score*, a single summary of the distribution (discussed more below). Correspondingly, a subscale would become a *subscore*. In the rest of this report scale and score, and subscale and subscore, are used interchangeably to conform to usage elsewhere in practice.

Scales are a highly useful aggregation method. Not only is the practice sound from the perspective of instrument design and statistical validity, it is helpful to get people to read and use SRI results. The large volume of numerical information from a list of many detailed items can be overwhelming, and grouping the items into scales or sub-scales means the instructors and administrators receive fewer category scores, which facilitates evaluation (Algozzine et al., 2004).

If there is a single item that rates the instructor or course as a whole, such as "The overall effectiveness of the instructor was…" or "The value of the overall learning experience was…", it should not be included in any scale or subscale. The single question and the scale are meant to assess the same thing with two different approaches.

The individual items that get aggregated into a single scale get the student to think about the several different facets of instruction individually and give them independent ratings. The items get summed into a scale to provide an overall view that may be used for further summative analysis while the individual items provide some useful formative feedback.

The stand-alone item addressing overall effectiveness gets the student to reflectively think about the instruction as a whole, implicitly producing a weighted average of the different elements in which the student has intuitively determined the relative importance of the items (Marsh & Bailey, 1993). This intuitive weighting appears to effectively represent all main teaching behaviours that are captured in the detailed items (Apodaca & Grad, 2005; d'Apollonia & Abrami, 1997; Hativa, 2011).

There is a very strong correlation between the response to the single overall item and the overall effectiveness scale (Hativa & Raviv, 1993; McBean & Lennox, 1987). The overall item generally produces a lower rating than the scale. This is often a source of the belief that the SRI has no value, as individual instructors compare the two scores from a tiny sample of SRI results with no supporting analysis of the relationship. However, there is a sound reason for the difference in scores. As noted above, the student implicitly and intuitively weights the relative importance of each dimension of teaching effectiveness. The scale, which is comprised of the average of several items, should also use a weight for each item to arrive at a better result rather than the simple average (Abrami, 2001). Although there are several possible approaches to determine weights (Marsh & Bailey, 1993), in reality it is nearly impossible to do so (Hativa, 2013a; Marsh, 2007).

Appendix A
SRI Data: An Overview

Hativa (2013a) cites research that shows that student responses to overall items have a higher correlation with other measures of teaching effectiveness and student learning than the more detailed items that reflect particular teaching behaviours (Abrami, 2001; Cashin, 1995; Cashin & Downey, 1992). The response to the single overall item may therefore be more appropriate to report and use in analyses, but in practice the scale is used. The overall item can then be used as a consistency check over a large set of results - all courses in a department over several years - to ensure that the scale is in fact highly correlated to the overall item. It is fortunate that there exists a strong correlation between the scale and the overall item, and a good relationship between the overall item and teaching effectiveness, since overall measures - whether a single overall item or a scale - are typically the only ones actually assessed in formal evaluation processes (Beran et al., 2005, 2007).

Aggregating the response distribution of an individual item into a single summary score

The response distribution for an aggregated item or aggregated scale represents the range and variation of opinion within the class. The distribution is usually represented by a single summary number (a *score*) to provide a simpler, more easily interpreted measurement.

The most common summary description is the central tendency of the distribution. The appropriate method depends upon the level of measurement for the item. If the item uses interval-level or ratio-level measurements, then either the median or the arithmetic mean, or average, can be used. If the item uses ordinal-level measurements, then only the median can be used. The variation of a distribution may also be of interest. If the item uses interval-level or ratio-level data, then the most common measure is the standard deviation. If the item uses ordinal-level data, then quantiles or percentiles may be used.

Likert and Likert-type items often use integers as the response encoding for each category. It is common practice to calculate the mean and standard deviation of these distributions. However this is not mathematically valid. Likert-like items are *ordinal-level* measurements, not *interval-level*. The integers are only substitutes for the qualitative labels. The calculation of the mean and standard deviation, and the mathematics underlying them, fundamentally rely on a uniform degree of difference between adjacent values. It is rarely defensible that the 'distance' or 'number of units' between values is equal – that is, the distances between *poor* and *adequate*, *adequate* and *good*, and *good* and *outstanding* are all the same – especially because the meanings and differences among the choices are judged by each student according to their own personal standards, experiences, and frames of reference.

Although the practice is not strictly valid from a mathematical standpoint, Carifio and Perla (2007) point to statistically grounded empirical research that shows that many of the commonly used tests are sufficiently robust that the results may be used with the recognition that assumptions are being bent and the results do not have the same strength. Hativa (2013a) also notes that, technically, ordinal-level data should not provide means or averages as measures of central values. She argues, however, that means, averages, factor analyses and regressions are widely used in the social sciences for ordinal scales, and that many well-established and widely used survey systems like SEEQ do employ these approaches, concluding that "the policy is widely agreed upon in practice" (Hativa, 2013a, p. 64).

This must always be kept in mind wherever SRI scores are reported or interpreted: **It is not strictly valid to do statistical tests on means of Likert-type scales and scores, but they may be used with caution. The exact value of such scores, the differences between two scores, and summary statistics describing groups of scores, are all less precise than they appear.**

Likewise, the degree of certainty for statistical tests conducted on these scores will be less than what is reported.

Comparisons: Statistical Tests of Similarity

The purpose of an SRI instrument is to measure the range and variation of opinion of students in a course with respect to instructor or course effectiveness. It is not possible to determine one single value that completely represents every student's rating: it is expected that students will have a range of experiences in the course, and not everyone will arrive at the same rating. It is therefore important to represent the class' response as a *distribution* of values. This makes it difficult to make comparisons among different groups or course instantiations to determine whether the distributions representing them are the same or different: there will always be some degree of overlap, and some degree of divergence. The challenge is to decide if there is a *sufficient degree* of overlap to decide that the groups are effectively the same. There are a number of statistical tools that can be employed: each must be used with a good understanding of its particular function and requirements. Otherwise it may be used in situations where it is not the appropriate tool, resulting in inaccurate conclusions.

Comparing two response distributions

The chi-squared test may be used to measure the degree of similarity between two distributions, evaluating how likely it is that any difference observed between the two distributions arose by chance. It may be used on ordinal data, so it is well suited for SRIs. It is used for two types of comparison: to test whether or not one frequency distribution differs from a theoretical or expected distribution (test of *goodness of fit*); and to test whether two observed distributions are different or the same (test of *independence*). As such, it is useful for examining SRI response distributions.

A test to see whether two distributions are different is useful in a number of different ways. As examples within a single course offering:

- for a particular course, is the variation of opinion among the students different for the overall instructor item than for the overall course item?
- is a class' pattern of opinion on an instructor's 'approachability and rapport' subscale different than their pattern of opinion on the 'presentation and delivery' subscale?

The chi-squared test is also useful to detect differences based on some significant change made over time. For example:

- An instructor notes that they tend to score weakly on the 'presentation and delivery' subscale. The instructor attends a variety of development workshops to learn new presentation skills and actively puts them into practice. Is there a difference on the 'presentation and delivery' subscale this year compared to two years ago, before they started their development work?
- Survey comments and verbal feedback reveal that students are dissatisfied with the assignments in a course. The instructor significantly revises the structure and assessment of the assignments with the help of their CTL. Does this result in a difference in the 'assessment' subscale?

While this calculation may provide evidence towards confirming a difference, variations among groups can limit the validity of conclusions drawn. For example, when comparing two course instantiations to determine if there is a difference in responses related to a change in practice, unrelated variations in class demographics, scheduling, or classroom might weaken the

comparison. One approach to mitigate this difficulty is to aggregate multiple 'before' and multiple 'after' instantiations to reduce these effects.

If the instructor can request reports filtered to a particular sub-group rather than the full class - for example, only the responses from those who indicated it was a required course - then it is possible to test whether the response distribution differs between two sub-groups. The individual surveys are not required as long as it is possible to obtain two versions of the usual aggregate report - one report filtered to include just the respondents with the particular characteristic, and one report including the other respondents. In order to maintain privacy and prevent identification of individuals, the reports should not be provided if the number of respondents in either the 'selected' sub-group or the 'other' sub-group is less than five (StatsCan, 2005).

Although it is more difficult to do, it is possible to test whether the response distribution for one item matches a particular distribution pattern. For example, is the distribution 'humped' around a central value, indicating a common rating with some lesser variation on either side? Or is the distribution bi-modal, indicating that opinion in the class is polarized?

The chi-square statistic relies on an assumption that all non-zero counts will be greater than or equal to five. If more than 20% of the counts are less than five then the results should not be used. Thus, it is not appropriate for courses with small class sizes or a low number of respondents.

Comparing the relationship between two items

Spearman's rank correlation may be used to measure how strong a relationship there is between how individuals respond to two items or scales on a survey. That is, if a respondent gives a high rating to the one item do they tend to give a high rating to the other as well, and similarly, if a respondent gives a low rating to one do they tend to give low rating to the other?

Spearman's rank correlation may be used on ordinal data, so is well suited for SRI items. When the value of Spearman's rank correlation coefficient is squared, it may be interpreted in the same way as the familiar r-squared coefficient used with in common linear regression. If it is important to determine whether the relationship is statistically significant rather than simply using the coefficient as a guide for interpretation, then a variation of the Student's t-test may be used. The responses from every respondent are needed to calculate the coefficient, so it must be calculated from the original survey data before the aggregate results are published; it cannot be done strictly from the aggregated response distributions, because individual pairs of responses are required, and cannot be derived from frequency counts. If access to the disaggregated survey results cannot be obtained for this purpose, then a variation of a chi-squared test on the response distributions for the two items can give a crude approximation of the relationship. However, there is no way of telling whether the high ratings on one item came from the same respondents as the high ratings on the other item, and so on. All that the test can tell is whether there is a strong relationship between the range of responses the class as a whole gave for the items.

Comparing two score distributions

The Student's t-test may be used to determine whether the mean of one collection of scores is different than the mean of another collection of scores, accounting for the respective size and standard deviation of the two collections. The structure of the test depends on whether we wish to determine if the one mean is *different* than the other mean (a 'two-tailed t-test'), or if the one mean is *higher* than the other mean (a 'one-tailed t-test'). More often than not, it is far more interesting to know whether one set of scores is on average higher than another, and the one-tailed t-test is the appropriate approach to take in most cases. For example, a program chair may wish to determine whether the distribution of all scores in their department during the past semester was higher than

the distribution of scores from the semester two years ago to see if a significant departmental initiative has had an observable impact on students' course experiences.

The one-tailed t-test can be used to determine whether the mean score for all courses taught by particular instructor is higher than the mean score for a comparator group. The test can be modified to use other important points within the distributions such as specific percentiles, for example whether an instructor's mean score is higher than the 80th percentile in the comparator group. This indicates that, on average, the instructor's scores are among the top 20% of all scores in the norm group. These tests are useful to establish whether a candidate for renewal, tenure or promotion meets a minimum threshold, or to support a claim of teaching excellence in a teaching award dossier.

The Student's t-test is based on the assumption that the values fit a normal distribution, and strictly speaking it should not be used if one of the collections does not meet that requirement. However, in practice the t-test may be applied as long as both collections show a 'mound-shaped' distribution (Mendenhall & Reinmuth, 1989). The practitioner should at lease make sure the two distributions are mound-shaped and somewhat symmetrical before applying a t-test. If one of the distributions shows too skewed a shape it would be highly inappropriate to use and interpret the t-test result since the assumptions of the test are severely violated, not just bent.

While comparing one collection of scores to another, it is crucial to always remember that the summary SRI scores were originally derived from ordinal data, not interval data, and therefore the values and their differences are not as precise as they seem. The results cannot be used for hard 'scientific proof', but rather for indications whether differences likely exist. They cannot be used to definitively and unarguably state that an instructor has exceeded a threshold, but rather for guidance regarding which cases easily pass the threshold and which cases near the boundary need more careful examination.

Appendix B
SRI Course Report

Student Rating of Instruction for: Course 1 Instructor A 2013 Fall

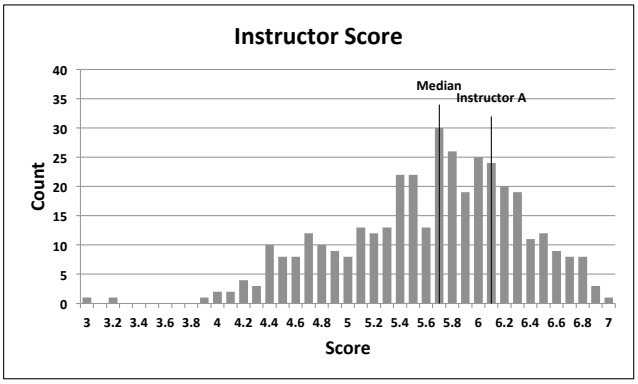
Instructor Score	Course Score	Number Enrolled	Responses	Course Required	Course Not Req'd
6.1	5.4	41	26	25	1

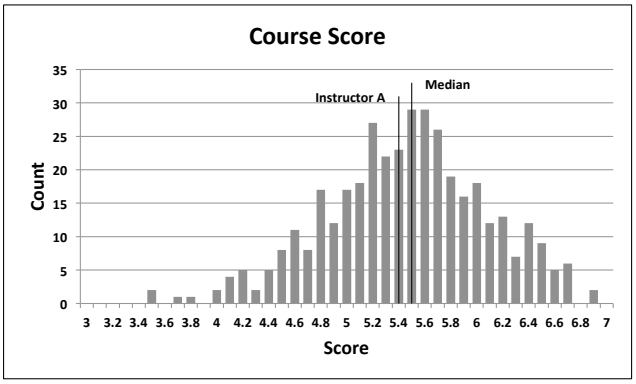
Level of enthusiasm for taking this course	Low	Medium	High
At the time of initial registration	0%	50%	50%
At the conclusion of the course	12%	44%	44%

Considering your experience with this course would you recommend it to other students?			
Yes	85%	No	15%

	Mean	1 Extremely Poor	2 Very Poor	3 Poor	4 Adequate	5 Good	6 Very Good	7 Out-Standing	n/a
Instructor (12 questions)	**6.1**			0%	4%	19%	**37%**	**39%**	
Overall Effectiveness (1 question)	6					27%	**46%**	27%	
Access & Rapport (5 questions)	6.5					7%	**33%**	**60%**	
Organization (2 questions)	5.2				20%	**46%**	**26%**	8%	
Presentation (4 questions)	6			1%	2%	20%	**47%**	**30%**	
Course (12 questions)	**5.4**		1%	3%	11%	**34%**	**38%**	12%	1%
Overall Value (1 question)	5.6				8%	**42%**	**29%**	21%	
Assessment Methods (3 questions)	5.5			3%	10%	**37%**	**35%**	15%	1%
Delivery (3 questions)	5.5			1%	11%	**36%**	**42%**	10%	1%
Organization (3 questions)	4.9		5%	9%	21%	**28%**	**28%**	8%	3%
Workload (2 questions)	5.8					**35%**	**52%**	13%	

How do my scores for this course compare to the rest of my department over the past ten years?





Your Instructor Score is in the 70th-90th percentile group.
This means that you were rated higher on the instructor-related questions than 70% of the other courses taught in your department over the past ten years. Very good!

Your Course Score is in the 30th-70th percentile group.
This means that your rating on the course-related questions was in the middle 40% of all courses taught in your department over the past ten years.

Appendix B
SRI Course Report

Student Rating of Instruction for: Course 1 Instructor A 2013 Fall

Instructor related questions:		Mean	1 Extremely Poor	2 Very Poor	3 Poor	4 Adequate	5 Good	6 Very Good	7 Out-Standing	n/a
1. presented material in an organized, well-planned manner		5.5				8%	**42%**	**38%**	12%	
2. was approachable for additional help		6.7					4%	**27%**	**69%**	
3. was accessible to students for individual consultation (in office hours, after class, open-door, by e-mail,phone)		6.6					4%	**31%**	**65%**	
4. The overall effectiveness of the instructor was		6					27%	**46%**	27%	
5. used instructional time well		4.9				**33%**	**50%**	13%	4%	
6. explained content clearly with appropriate use of examples		5.8				4%	**32%**	**40%**	24%	
7. was a clear and effective speaker		6.1					19%	**54%**	**27%**	
8. communicated enthusiasm and interest in the course material		6.3					12%	**46%**	**42%**	
9. stimulated your interest in the subject and motivated your learning		5.9			4%	4%	19%	**46%**	**27%**	
10. attended to students' questions and answered them clearly and effectively		6.5					12%	**24%**	**64%**	
11. was open to students' comments and suggestions		6.5					4%	**38%**	**58%**	
12. was sensitive to students' difficulties		6.3					12%	**42%**	**46%**	

Course related questions:		Mean	1 Extremely Poor	2 Very Poor	3 Poor	4 Adequate	5 Good	6 Very Good	7 Out-Standing	n/a
1. How effective was the course outline in communicating goals and requirements of the course?		5.4				20%	**28%**	**40%**	12%	4%
2. How reasonable was the level of difficulty of the course material?		5.7					**35%**	**58%**	8%	
3. How reasonable was the volume of the work required in the course?		5.8					**35%**	**46%**	19%	
4. The value of the overall learning experience was		5.6				8%	**42%**	**29%**	21%	
5. How consistently did the stated course goals match what was being taught in the course?		5.3			4%	8%	**48%**	**32%**	8%	4%
6. How appropriate was the course format for the subject matter?		5.5			4%	12%	**24%**	**52%**	8%	
7. How well did the methods of evaluation (e.g., papers, assignments, tests, etc.) reflect the subject matter?		5.7				8%	**29%**	**46%**	17%	
8. How fair was the grading of student work?		5.8					**41%**	**41%**	18%	5%
9. How timely was the grading of student work?		3.9		16%	**24%**	**36%**	8%	12%	4%	
10. How helpful were comments and feedback on student work?		5.1			8%	20%	**40%**	20%	12%	
11. How well did the instructional materials (readings, audio-visual materials, etc) facilitate your learning?		5.5				13%	**39%**	**35%**	13%	4%
12. How well did the instructional activities (lectures, labs, tutorials, practia, field trips, etc) facilitate your learning?		5.5				8%	**44%**	**40%**	8%	

References

Abrami, P.C. 2001. Improving judgments about teaching effectiveness using teacher rating forms. Chapter 4 in *New Directions for Institutional Research*, *109*, 59-87. Jossey-Bass.

Alderman, L., & Melanie, L. (2012). REFRAME: A new approach to evaluation in higher education. *Studies in Learning, Evaluation, Innovation & Development, 9*(1), 33-41. from http://sleid.cqu.edu.au

Alderman, L. & Melanie, L. (2013) Kaleidoscope: Reframing learning and teaching evaluation through a shared stakeholder lens. In *Australasian Higher Education Evaluation Forum (AHEEF) 2013*, 28-30 October 2013, Launceston, TAS.

Algozzine, B., Beattie, J., Bray, M., Flowers, C., Gretes, J., Howley, L., Mohanty, G., & Spooner, F. (2004). Student Evaluation of College Teaching: A Practice in Search of Principles. *College Teaching*. *52*(4), 134.

Apodaca, P., & Grad, H. (2005). The dimensionality of student ratings of teaching: Integration of uni- and multidimensional models. *Studies in Higher Education, 30*, 737-748.

Astin, A. (1970). The methodology of research on college impact (Part 1). *Sociology of Education, 43*(3), 223-254.

Beran, T., & Rokosh, J. (2009). Instructors' perspectives on the utility of student ratings of instruction. *Instructional Science, 37*(2), 171-184.

Beran, T., Violato, C., Kline, D., & Frideres, J. (2005). The utility of student ratings of instruction for students, faculty, and administrators: A "consequential validity" study. *The Canadian Journal of Higher Education, 35*(2), 49-70.

Berk, R. A. (2009). Using the 360° multisource feedback model to evaluate teaching and professionalism. *Medical Teacher*, *31*(12), 1073-1080.

Berk, R.A. (2013). *Top 10 Flashpoints in Student Ratings and the Evaluation of Teaching.* Stylus Publishing, Sterling VA.

Berk, R. A. (2013). Face-to-Face versus Online Course Evaluations: A" Consumer's Guide" to Seven Strategies. *Journal of Online Learning & Teaching,* 9(1), 140-148.

Cashin, W. (1995*). Student ratings of teaching: The research revisited*. IDEA Paper No. 32. Retrieved January 18, 2014 from http://ideaedu.org/sites/default/files/Idea_Paper_32.pdf

Cashin, W. & Downey, R. (1992). Using global student rating items for summative evaluation. *Journal of Educational Psychology, 8*4(4), 563-572.

Clayson, D. (2009). Student evaluations of teaching: Are they related to what students learn? A meta-analysis and review of the literature. *Journal of Marketing Education*. *31*(1), 16-30.

d'Apollonia, S., & Abrami, P. (1997). Navigating student ratings of instruction. *American Psychologist, 52*, 1198-1208.

Dillman, D.A. 2000. *Mail and internet surveys: The tailored design method*. Brisbane: Wiley.

Galbraith, C., Merrill, G., & Kline, D. (2012). Are student evaluations of teaching effectiveness valid for measuring student learning outcomes in business related classes? A neural network and Bayesian analyses. *Research in Higher Education, 53*(3), 353-374.

Glass, G.V., P.D. Peckham and J.R. Sanders. 1972. Consequences of failure to meet assumptions underlying the analyses of variance and covariance. *Review of Educational Research*, *42*, 237-288.

Goyder, J. (1987). *The silent minority: Nonrespondents on sample surveys*. Camridge: Polity Press.

Gravestock, P. and Gregor-Greenleaf, E. (2008) *Student Course Evaluations: Research, Models and Trends*. Toronto: Higher Education Quality Council of Ontario.

Hamilton, B., & Graniero, P. (2012). Disruptive cartography in academic development. *International Journal of Academic Development 17*(3), 243-258.

Hativa, N. (2013a). *Student ratings of instruction: Recognizing effective teaching*. USA: Oron Publications.

Hativa, N. (2013b). *Student Ratings of Instruction: A practical approach to designing, operating, and reporting.* Charleston, SC: CreateSpace Independent Publishing Platform.

Hativa, N., & Raviv, A. (1993) . Using a single score for summative teacher evaluation by students. *Research in Higher Education 34*(5), 626-646.

Hallinger P. (2010). Using faculty evaluation to improve teaching quality: A longitudinal case study of higher education in Southeast Asia. *Educational Assessment Evaluation and Accountability*; 22:253–274.

Joughin, G., & Boujos, K. (2014). *A report on third-party SRI management systems*. Technical Report. University of Windsor: Windsor, Ontario.

Joughin, G., & Winer, L. (2014). *Response to Teaching Evaluation Toolkit preliminary reporting*. Panel discussion presented March 5, 2014 in Toronto, Ontario.

Kelly, M. (2012). *Student evaluations of teaching effectiveness: Considerations for Ontario Universities*. COU Academic Colleagues Discussion Paper.

Langbein, L. (2008). Management by results: Student evaluation of faculty teaching and the mis-measurement of performance. *Economics of Education Review, 27*(4), 417-428.

Lang, J. W. B., & Kersting, M. (2007). Regular feedback from student ratings of instruction: Do college teachers improve their ratings in the long run? *Instructional Science, 35*(3), 187-205.

Likert, R. 1932. A technique for the measurement of attitudes. *Archives of Psychology*, 140, 1-55.

Ludlow, L. H. (1996). Instructor evaluation ratings: A longitudinal analysis. *Journal of Personnel Evaluation in Education*, *10*, 83-92.

Ludlow, L. H., (2005). A longitudinal approach to understanding course evaluations. *Practical Assessment, Research & Evaluation, 10*(1). Retrieved from http://pareonline.net/pdf/v10n1.pdf

Marsh, H. W. (2007). Students' evaluations of university teaching: A multidimensional perspective. In R. P. Perry & J C. Smart (Eds.), *The scholarship of teaching and learning in higher education: An evidence-based perspective* (pp. 319-384). New York: Springer.

Marsh, H. (1982). SEEQ: A reliable, valid and useful instrument for collecting students' evaluations of university teaching. *British Journal of Educational Psychology, 52*(1), 77-95.

Marsh, H. W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research, 11(3),* 253-388.

Marsh H. (2007). Students' Evaluations of University Teaching: Dimensionality, Reliability, Validity, Potential Biases and Usefulness. In R.P. Perry and J.C. Smart (eds.), *The Scholarship of Teaching and Learning in Higher Education: An Evidence-Based Perspective*, pp. 319–383.

Marsh, H., & Bailey, M. (1993). Multidimensional students' evaluation of teaching effectiveness: a profile analysis. *Journal of Higher Education 64*(1), 1-18

Mason, P. M., Steagall, J. W., & Fabritius, M. M. (1995). Student evaluations of faculty: A new procedure for using aggregate measures of performance. *Economics of Education Review, 14(4),* 403-416.

McBean, E., & Lennox, W. (1987). Measurement of quality teaching and course by a single question versus a weighted set. *European Journal of Engineering, 12*(4), 329-335.

Mendenhall, W. and J.E. Reinmuth. 1989. *Statistics for Management and Economics*. 6th rev. ed. Brooks/Cole.

Neilsen, H., Moos, R., & Lee, E. (1978). Response bias in follow-up studies of college students. *Research in Higher Education, 9*, 97-113.

Nulty, D.D. 2008. The adequacy of response rates to online and paper surveys: what can be done? *Assessment & Evaluation in Higher Education*, 33(3), 301-314.

Office of the Auditor General of Ontario (2012). *2012 Annual report of the office of the auditor general of Ontario: Section 3.12 University Undergraduate Teaching Quality*. Ministry of Training, Colleges and Universities.

Pallett, W. (2006). Uses and abuses of student ratings. In Peter Seldin (ed.) Evaluating faculty performance: A practical guide to assessing teaching, research, and service, 50-65.

Patrick, C. L. (2011). Student evaluations of teaching: effects of the big five personality traits, grades and the validity hypothesis. *Assessment & Evaluation in Higher Education. 36*(2), 239–249

Ramsden, P. (1991). A performance indicator of teaching quality in higher education: The course experience questionnaire. *Studies in Higher Education, 16(2),* 129-150.

Richardson, J.T.E. 2005. Instruments for obtaining student feedback: a review of the literature. *Assessment & Evaluation in Higher Education*, 30(4), 387-415.

Ryan, K. (1997). *Making student ratings comprehensible to faculty: A review of alternative reporting approaches*. Paper presented at the 78th American Educational Research Association. Chicago.

Salmon, P., Deasy, T.,& Garrigan, B. (2004). *What escapes the Net? A statistical comparison of responses from paper and web surveys.* Paper presented at the 2004 Evaluation Forum: Communicating Evaluation Outcomes: Issues and Approaches. Melbourne, Australia, 24-25 November.

Scriven, M. (1995). Student ratings offer useful input to teacher evaluations. *ERIC/AE Digest, ED398240.*

Slade, S., & Prinsloo, P. (2013). Learning analytics: ethical issues and dilemmas. *American Behavioral Scientist, 57*(10), 1509–1528. Retrieved from http://oro.open.ac.uk/36594/2/ECE12B6B.pdf

Slocombe, T., Miller, D., & Hite, N. (2011). A survey of student perspectives toward faculty evaluations. *American Journal of Business Education, 4*(7), 51-57.

StatsCan. 2005. *Research Data Centres (RDCs) Guide for Researchers under Agreement with Statistics Canada*. Statistics Canada. Available online at: http://www.statcan.gc.ca/rdc-cdr/pdf/researcher-rechercheur-guide-eng.pdf (accessed June 15, 2014).

Stevens, S.S. 1946. On the theory of scales of measurement. *Science*, 103(2684), 677-680.

Sullivan, S. L., Polnick, B., Nickson, L., Maninger, R., & Butler, J. Y. (2013). Student evaluation of teaching: The inequity of faculty scores in online versus face-to-face courses. *School Leadership*, *8*(2), 52-61. Retrieved from http://www.tcpea.org/cms/lib07/TX01923126/Centricity/Domain/191/summer13.pdf

Surgenor P. (2013). Obstacles and opportunities: addressing the growing pains of summative student evaluation of teaching. *Assessment & Evaluation in Higher Education*, 38(3), 363–376

Theall, M. (2002). Student Ratings: Myths vs. Research Evidence. *https://studentratings.byu.edu/info/faculty/myths.asp.* Accessed April 22, 2014

Theall, M., & Franklin, J. (2000). Creating responsive student ratings systems to improve evaluation practice. In K. E. Ryan (Ed.), *Evaluating teaching in higher education: A vision for the future: New Directions for Teaching and Learning, 83* (pp. 95-107). San Francisco: Jossey-Bass.

Theall, M. & Franklin, J. (2001). Looking for bias in all the wrong places – A search for truth or a witch hunt in student ratings of instruction? *In The student ratings debate: Are they valid? How can we best use them?* Theall, P., Abrami, L. and Lisa Mets (Eds.) *New Directions in Educational Research, 109*. Jossey-Bass: San Francisco.

Tufte, E. 2006. *Beautiful evidence*. Cheshire (CT): Graphics Press

Üstünlüoğlu, E. & Güngör-Culha, D. (2012). Investigating student evaluation of teachers by using latent class analysis: A case study at a tertiary level. *International Journal of Education, 4(3),*147-159.

Villascusa, Franklin & Aleamoni (1997) p20 Don't need it, cited from Hativa.

Watkins, D.A*., &* Hattie, J. *(*1985*).* A longitudinal study of the approach to learning of Australian tertiary students*. Human Learning, 4, 127-142.*

Watt, S., Simpson, C., McKillop, C., & Nunn, V. (2002). Electronic course surveys: Does automating feedback and reporting give better results? *Assessment & Evaluation in Higher Education 27*(4), 325-337.

Winer, L., di Genova, L, Vungoc, P., & Talsma, S. I*nterpreting end-of-course evaluation results.* Montreal: Teaching and Learning Services, McGill University

Wininger, S., & Birkholz, P. (2013). Sources of instructional feedback, job satisfaction, and basic psychological needs. *Innovative Higher Education*, *38*(2), 159-170.