Diagnostic Score Reporting for a Dental Hygiene Structured Clinical Assessment

by

Alix Clarke

A thesis submitted as partial fulfillment of the requirements for the degree of

Master of Science

Medical Sciences—Dental Hygiene

University of Alberta

Abstract

Background: Structured clinical assessments (SCAs) are an essential part of health professional education as they capture important information on not only what a student knows, but what they can do. However, this useful information is rarely translated into quality feedback that students can use to reflect upon and improve their clinical performance. Feedback is considered a fundamental component of both learning and professional development, and educators are calling for more and better feedback across the health disciplines. However, issues such as time limitations and test confidentiality make feedback provision for SCAs challenging. Diagnostic score reporting (DSR) presents a possible framework for providing all students with more detailed feedback on their SCA performances. DSR summarizes test performance by the underlying domains of knowledge, skills and/or abilities the test intends to measure, and includes resources for making individual improvements within those domains. DSR does not require the actual test items to be revealed to the students, and can be administered efficiently through online means. As such, DSR has some advantageous as potential feedback mechanism for SCAs. To date, DSR has largely been applied only within the context of large-scale assessment, particularly in primary and secondary education. Additionally, very little research has been conducted on the feedback's measurable impact on student outcomes. *Objectives*: To develop a general framework for applying DSR within SCAs; to develop and validate a course-specific diagnostic score report for a dental hygiene SCA; and to evaluate the effect of DSR on student reflection and performance. *Methods*: A literature-based adapted DSR framework was developed to guide the process of DSR for SCAs. This general framework was then applied towards a dental hygiene history taking SCA at the University of Alberta. This process involved identifying

the diagnostic domains of the assessment and linking competencies, test items, and learning resources to those domains. In order to evaluate the effect of DSR on student outcomes, a mock-SCA was developed where half the students were randomly assigned to receive DSR, while the other half received only one overall numerical grade following the assessment. All students were then asked to reflect upon their mock-SCA performance, and later completed their regularly scheduled year-end history taking SCA. The results were collected and analyzed to look for differences between groups on reflection quality, content, and year-end SCA results. *Results*: Four skills-based domains were identified as necessary to complete the dental hygiene history taking SCA: effective communication, client-centered care, eliciting essential information, and interpreting findings. No differences in reflection quality were found, while reflection content significantly differed by the experimental groups. The DSR group was significantly more likely to report needing to improve on interpreting findings ($p = .007$), while the control group focused on eliciting information ($p = .04$). Overall, students tended to perform quite well on eliciting information ($M = 92.11\%$, $SD = 9.63\%$), but poorly on interpreting findings ($M = 42.11\%$, $SD = 17.56\%$). The DSR group did not show significant improvements in their year-end SCA results. *Conclusions*: DSR appeared to result in improved identification of history taking skills that required improvement, however this improved self-assessment did not translate into improved performance. DSR presents a promising start for improving the feedback students receive following their SCAs, however further enhancements are required. Suggestions for improving the feedback to facilitate behaviour change include: improving the learning resources provided to the students within the report, adding video feedback (self and exemplars), providing a means for students to response to their feedback, and increasing the individualization/personalization of the reports.

Preface

This thesis represents original work by Alix Clarke. The research project, of which this thesis is a part, received ethics approval from the University of Alberta Research Ethics Board, "Evaluation of Online Diagnostic Score Reporting on Student Reflection and Clinical Performance for a Dental Hygiene Clinical Skills Assessment", Pro00062297, approved June 2016 (amendment approved February 2017).

Acknowledgements

First and foremost I would like to thank my supervisors, Dr. Minn Yoon and Dr. Hollis Lai, for their excellent mentorship and unwavering support. I would also like to thank Professor Alexandra Sheppard for her clinical expertise and Cody Surgin for his technical expertise, without which this project would not have been possible.

I would like to acknowledge CIHR's Canadian Graduate Masters Scholarship and the School of Dentistry's Educational Scholarship Research Fund for providing the financial support that allowed me to conduct my research.

Thank you to my supportive husband for encouraging me to accomplish my goals and for the strength he gives me every day.

Finally, I would like to thank Nadia Kobagi for being my confidante from day one, and for keeping me laughing throughout this journey.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

DSR: Diagnostic Score Reporting

ECSAS: Electronic Clinical Skills Assessment System

GPA: Grade Point Average

K-12: Kindergarten to grade twelve

KSA: Knowledge, skill, and ability

OSCA: Objective Structured Clinical Assessment

OSCE: Objective Structured Clinical Examination

SCA: Structured Clinical Assessment

SP: Standardized Patient

Chapter 1: Introduction

Background

Students need to receive feedback throughout their education in order for effective learning to take place.[1] By receiving specific information about how their performance relates to program expectations, students can better understand the subject material and make future improvements.[2,3] Several reviews indicate that, on average, feedback does lead to improved performance.[4-6] The education of health professionals, in particular, highlights feedback as a cornerstone to both learning and professional development.[7,8] Without feedback, students are destined to repeat their mistakes, and may carry over false or incomplete understandings into their professional practice.[7] However, feedback is frequently reported as inadequate within the health disciplines.[9,10] Methods for increasing and improving feedback within health education is required to improve learning outcomes.

Assessments present important opportunities for providing detailed feedback to each student, as they typically capture a tremendous amount of information on student abilities at key stages of learning.[10] In health education, competence (i.e. the specific knowledge, skills, abilities, and even values a person must possess for daily professional practice)[11,12] is typically measured through carefully constructed structured clinical assessments (SCAs). SCAs are performance assessments designed to measure not just what a student *knows* (such as in a written examination), but what a student actually *shows* they can *do*.[11,13,14] SCAs take on a variety of forms—the most common of which is the Objective Structured Clinical Examination (OSCE)—all with the goal of determining clinical competency in a fair, objective, yet realistic manner.[15] Specifically, SCAs involve carefully designed and standardized performance-based scenarios graded using impartial methods. These scenarios reflect the major learning objectives of the

course, to ensure student competence before they progress in their education/careers.[16,17] Course materials, professional competencies, expert opinions, and relevant literature should be gathered to aid in the determination of the vital content for each SCA.[17] Grading checklists are then created containing the key tasks/criteria the student must accomplish within each encounter (i.e. objective measures of competence), to establish marking consistency and accuracy.[16] Actors are frequently trained to portray patients in these assessments to standardize the experience across all students, while still mimicking actual clinical practice.[18] In sum, substantial effort is put into SCAs to reliably capture and assess student clinical competence.

Although detailed information on student competence is collected through SCAs, it is rarely translated back to the student in the form of feedback.[10] Such assessments are usually summative in nature, occurring at the end of a learning period, where feedback is not typically provided.[10,12] These summative assessments are contrasted in the literature by formative assessments, whose primary purpose is to help students learn, and for which feedback is an essential component.[12,19] However, this formative and summative dichotomy (i.e. assessments for learning versus evaluation purposes only) is being challenged in more recent literature, as educators realize all assessments provide an important opportunity for learning.[8,20,21] Assessments directly dictate how and what students learn, with the stakes of an assessment impacting how much time and effort a student will invest in any subject area.[22] Research also supports that a majority of students will interact with feedback following a summative SCA.[21] Therefore, feedback for summative SCAs has the potential to be quite impactful due to the apparent importance these assessments hold to the students. Archer[8] describes the need for a "culture of feedback" where students consistently receive information about their performance throughout their health education. He emphasizes that external feedback is key to developing

self-monitoring skills, which fosters an understanding of one's own capabilities to respond to any given healthcare scenario. Providing feedback for SCAs is one important step towards this culture shift, although little research has been conducted quantifying the impact of that feedback.

Diagnostic score reporting (DSR) represents one framework for providing meaningful feedback to students following an assessment.[23,24] DSR describes student performance on tests by domains (specific areas of knowledge or skill), highlighting strengths and weaknesses and providing learning resources that students can engage with for further development. These reports have the potential to reinforce positive behaviours and identify learning deficits so that individual improvements can be made.[25,26] The application of DSR has been largely confined to nationwide testing, particularly in primary and secondary education, and much of the literature has focused only on reporting features, such as usability and interpretability.[23-27] Further investigation into applying DSR within higher education, and its resulting effect on student-level outcomes is still needed.

The dental hygiene history taking SCA at the University of Alberta is an example of an SCA where students could benefit from the application of DSR. This SCA involves developing rapport with a client and questioning them on their health and dental history. Students must establish a dialogue, identify and follow-up on any significant findings, and ultimately determine if any necessary modifications, or even contraindications, to care exist. This comprehensive assessment combines much of the knowledge and skills taught throughout the dental hygiene program regarding interpersonal communication and risk assessment. Currently, feedback following this assessment is minimal, largely restricted to numerical grades. Incorporating DSR into this SCA would provide all students with more detailed information on their performance. However, there is very little literature documenting how the DSR framework, developed for

large-scale assessments, can be applied to smaller scale testing scenarios such as SCA performance assessments. Furthermore, the impact of using DSR as feedback following an SCA on student-level outcomes, such as reflective capacity and performance, is unknown. Introducing DSR within the dental hygiene history taking SCA would not only have potential for improving the learning of students, but would also provide information for effectively applying and understanding the effects of DSR in higher education.

Study Purpose

The purpose of this study is to apply DSR within the context of an SCA, and to evaluate the effect of this feedback method on student learning. More specifically, the research questions are:

1) To what extent can the framework of DSR be applied to provide all students with valid and high-quality feedback following an SCA?
2) How will online DSR impact student-level outcomes such as reflective capabilities and clinical performance?

In order to address these research questions, the framework for DSR will be adapted and summarized into a practical guide for providing high-quality feedback following an SCA. Applying this guide, a course-specific diagnostic score report for a dental hygiene SCA will be developed and validated. Finally, this course specific report will be piloted and evaluated using an experimental method and a mixed-methods analysis to determine the relative impact of DSR on student outcomes, including reflection and performance.

Thesis Outline

This thesis is organized into five chapters: the introduction, background literature, research methods, study results, and discussion of major findings. This introduction provides a brief overview of the subject matter, and the context and significance of the research. Following this outline, a definition of key terms is provided to aid the reader. The literature review will involve a thorough investigation into SCAs, feedback, and DSR, and how these three components have interacted in the literature to date. In the methods chapter, the first two sections reflect the first research question: organizing the information revealed from the literature review into a practical framework for applying DSR within SCAs, followed by the context for applying this framework towards a dental hygiene SCA. The third section relates to the second research question, describing the methodology for a student-level outcomes evaluation. The results chapter will similarly mirror the research questions, describing the resulting application of the general DSR framework within a dental hygiene SCA, and the effects of DSR on dental hygiene students' reflection and performance. The discussion will include a more detailed exploration of the results and their implications, situating them in relationship to other literature. The discussion will also include suggestions for improving DSR based on the study outcomes as well as recommendations for future research in this area.

Definitions of Key Terms

There are several key terms which will be used throughout this thesis. A detailed explanation and exploration of these terms will ensue in the following chapter. However, a brief operational definition of these terms is provided here for the reader's clarity and reference.

Assessment: Any measure of student performance (in comparison to expectations), including, but not limited to, examinations, evaluations, and tests. These terms may be used interchangeably within this document, despite small semantic differences.

Blueprinting: The process of linking test items to the objectives of the test to ensure accuracy in testing content.

Client: The preferred term to patient in the dental hygiene literature. Client is therefore used when referring to the dental hygiene profession, while patient is used when referring to other health professions such as medicine. Despite the different favored professional vocabulary, patient and client are largely interchangeable within dental hygiene.

Competence: Having and applying the appropriate knowledge, skills, abilities, and values for professional practice, typically reflecting the standards of practice dictated by governing bodies. ***Clinical*** competence may also be used when specifically referring to the health professions and their performance in clinical settings.

Diagnostic Score Reporting (DSR): A method of summarizing student performance on an assessment by specific domains so that students can easily identify strengths and weaknesses. The goal is to provide information that will be useable by the student to facilitate individual improvements.

Domain: The key overarching areas of knowledge, skill, and/or ability a test is trying to measure.

Feedback: Providing information to a student on how their performance relates to expectations or a standard, with the goal of improving student outcomes.

Formative Assessment: An evaluated learning activity specifically designed to improve student outcomes through provision of feedback.

Objectives: The knowledge and skills a student should acquire within a course. Learning objectives are similar to competencies but reflect course specific content rather than professional goals.

Outcomes: The result of an event or activity. ***Student*** outcomes are the result of a learning activity (e.g. receiving feedback) that directly impacts student knowledge or skills—such as reflective capacity or clinical performance.

Reflection: Actively thinking about one's experiences. Includes both *what* they are thinking about (i.e. reflection content), and the process of *how* they reflect upon their experiences (i.e. reflection quality).

Self-Monitoring: Being aware of one's own thoughts, feelings, and abilities at any moment in time.

Self-Regulated Learning: When students become active participants in their own learning. Includes well developed meta-cognitive processes (including self-awareness and other higher-ordered thinking processes) and internal motivations.

Standardized Patients (SPs): Actors who have been trained to portray a certain illness and medical history.

Structured Clinical Assessment (SCA): An umbrella term referring to all formal evaluations of clinical performance that involve a rigid grading structure (e.g. grading checklists), carefully chosen content (e.g. constructed scenarios), and take place in a real-world

environment (e.g. in the clinic versus the classroom, using SPs). The most common SCA is the Objective Structured Clinical Examination (OSCE).

Summative Assessment: An evaluation that summarizes student competence following a learning period (e.g. year-end finals or licencing examinations). Historically, feedback is not typically provided for these assessments.

Validity: The evidence we have to support the claims we are making. In test theory, validity is a very complex topic, involving multiple sources of evidence related to the testing content, structure, and consequences.

Chapter 2: Literature Review

This review begins with the history of SCAs, including their benefits, limitations, and current areas of technological advancement. Next, there will be a detailed exploration of the current state of feedback in higher education, its impact on learning and development, as well as characteristics that make feedback high-quality. The state of feedback for SCAs specifically will then be described, including current issues in providing quality feedback within these assessments. The remainder of the review will focus on DSR, elaborating on how DSR can be validly applied within SCAs. The review will end with a summary of key elements that outline the rationale for proceeding with the research conducted for this thesis.

History of Structured Clinical Assessments

Clinical assessments are a necessary part of health education to ensure students not only have knowledge, but are able to apply this knowledge competently.[11] Historically, clinical expertise in medical education was measured by written examinations and the assessment of students during interactions with real patients.[15] In the mid-to-late nineties it became clear that current methods of assessment were in adequate.[11,15] Written examinations, such as traditional multiple choice tests, lacked practical validity in their determination of skills-based competence.[15] Furthermore, evaluating students on patient-specific interactions was too variable, with chance playing a major role on the kinds of patients the students would see.[11,15] It became clear to the medical education community that more standardized methods for accurately evaluating students on key clinical objectives were necessary. The result was a push for *structured* clinical assessments (SCAs) that would allow evaluators more control over the scenarios being assessed, while still mirroring actual clinical practice.[11,15,16] Several methods of

providing realistic, yet organized and increasingly fair, assessments, throughout health education, have now been developed.

A primary condition of an SCA is that it accurately assess important clinical competencies.[16] During development of an SCA, care must be taken to determine the key scenarios that should be assessed at each stage of learning. Historically, SCA objectives were defined by expert opinions and previous literature, which outlined the main situations and problems medical students would encounter in their clinical practice.[16,17] In more recent years, professionally defined competencies have also come to play a major role in determining SCA content.[28] In addition to scenario content, these competencies and objectives aid in development of a structured grading rubric, where objective measures of student achievement within each scenario are pre-established. This process of linking objectives to assessment content and grading is referred to as blueprinting.[17,29] Blueprinting creates a link showing how student performance on an SCA accurately reflects their clinical skills and competence.

Another component at the forefront of SCA development is the use of standardized patients (SPs).[18] SPs are typically actors who have been trained to portray a certain illness and medical history, allowing for improved consistency in student-patient interactions across students.[12,14,16,18] These clinical simulations with SPs allow educators to ensure that all students have an opportunity to interact with a "patient" demonstrating key symptoms and specific needs, as required for assessment of certain learning objectives. Utilizing SPs also provides students an opportunity to test their proficiency in certain clinical skills before applying them to actual patients.[18] SPs play a key role in the majority of SCAs.

To date, the most commonly used SCA is the Objective Structured Clinical Examination (OSCE).[15] A typical OSCE involves multiple stations, each with a structured scenario, portrayed

by an SP, and graded by qualified examiners who use a predetermined checklist of yes/no (standard met/standard not met) criteria.[14-17] Having multiple stations means a student will see different patients and be graded by different examiners, increasing the fairness of high-stakes assessments (e.g. if some SPs or examiners are better trained than others these effects will average out over the OSCE).[14,15] Designing scenarios around specific objectives guarantees that each student is being assessed for whether they have developed competency on several important domains of learning.[14-16] The grading checklists provide examiners with objectively measurable tasks a student must complete, increasing marking consistency between different examiners.[14-16] These checklists may also contain global rating scales (i.e. qualities marked on a scale rather than yes/no), allowing for better evaluation of more fluid skills such as communication and patient-centered care, which may be marked by the examiners or the SPs themselves.[30-32] OSCEs have undergone substantial reliability and validity testing in a variety of different fields and levels of medical training.[28,33-37] They are now largely considered a cornerstone of clinical assessment in medical education.

The strengths of the OSCE has extended its popularity for use in health disciplines other than medicine.[38-51] However, other SCAs have also been developed, modifying the OSCE process as needed for different educational purposes. For example, one long scenario, with a single patient interaction (in contrast to using several shorter stations) is often used in nursing to evaluate a more holistic and comprehensive patient-centered interaction, referred to as an Objective Structured Clinical Assessment (OSCA).[39,40] Similarly, a single SP assessment with a focus on interpersonal communication is used in dental hygiene to examine students' competence as an oral health promotor.[52] Some SCAs, such as in dentistry, provide stations that focus more on determining technical proficiencies, such as reading x-rays or handling

instruments, as opposed to emphasizing patient interactions.[46] Technical simulations and SP use can be supplemented using realistic mannequins or even virtual reality, continuing to expand the realm of possibilities for SCAs.[53] SCAs have been readily used not only because they are able to measure competence with increased objectivity and fairness, but also because they have been found to improve student-level outcomes, such as confidence, capacity for self-assessment, and clinical abilities.[46,51,52] Moreover, the assessment results can also aid course coordinators in identifying overarching student weaknesses, which may lead to curricular modifications, improving the quality of education.[52,54] These benefits indicate that the use of a variety of SCAs will likely continue to grow in the education of healthcare providers.

In recent years, the integration of technology into SCAs has been a focus for improving educational and organizational outcomes. Concerns over the large administrative burden of SCAs—in particular the environmental/paper costs, and the man-hours needed to prepare, print, copy, disseminate, label, collect, and scan those papers—has led to the development of systems for electronically administering these assessments.[49,55-57] Typically, electronic administrations involve computerized grading (e.g. via iPads or laptops), an online platform for examination development, data storage, and analytics, and a student interface for accessing their results. Research indicates these systems successfully decrease the time needed to implement an SCA, as well as providing additional benefits such as timely feedback for students and efficient exam-related data analysis (such as identifying the most difficult stations or the most frequently missed checklist items).[49,56] Unique and improved applications of technology in health education is an important area requiring continued study and advancement.

Despite all the advantages of SCAs, there are still some areas of concern and criticism. It is well established that assessments impact what and how students learn.[22,58-60] For example,

research has demonstrated that medical students' motivation to study in any one subject area is directly related to how much that subject impacts their final grades.[22,59] As such, critics fear the emphasis on assessing competence has negated the promotion of clinical excellence, with students studying and practicing skills only to the extent that they achieve the minimum level of competence needed to receive a passing grade (as opposed to trying to master skills).[13,21] Others worry about the categorical nature of grading checklists, which may cause students to memorize a series of graded steps, rather than evaluating each patient in a holistic manner, and discourage students from varied/unique critical thinking abilities.[13,17] Despite these concerns, the benefits of SCAs in their objectivity, fairness, and assurance of achievement of competence, make them essential for clinical education. Addressing these concerns then becomes a balance between improving holistic thinking and promoting excellence, while retaining the structure of the assessments themselves. One mechanism which could allow for the retention of SCAs as is, while still encouraging development of skills beyond basic clinical proficiency, is feedback. The feedback students receive following SCAs is an opportunity for the promotion of skill development that does not directly impact the nature of the assessment.

Feedback in Higher and Clinical Education

Feedback is considered an essential pedagogical tool in higher education.[1,9] While assessments evaluate student performance (primarily through provision of grades), feedback is the optional opportunity to provide students with more detailed information about their performance, in an effort to improve future performances.[2,3,7,9] Kolb's[61] experiential learning model is highly applicable to hands-on clinical education, and describes learning as a cycle that begins with concrete experiences. These experiences lead to reflection and observation, followed by analysis and conceptualization, and finally integration of new information into modification

of their own behaviours in active experimentation—resulting in new experiences to reflect upon.[61] Situating this cycle within the context of clinical education, a student might perform a task in clinic, notice their performance needs improvement, reflect upon why and how, draw in information from other sources (e.g. text books, instructors), formulate an understanding of the issue and what they might do differently (i.e. the cognitive aspects of learning take place), and eventually implement the changes into their next clinical performance (i.e. behaviour is altered as a result of learning). The concept of learning as a process is largely accepted by the educational community, with feedback believed to play an integral role within this cycle.[1,62-64] Torbert[64] describes feedback as a critical factor predicting why some people fall into repetitive behaviour patterns, while others use their experiences to make improvements. He explains how accurate feedback is sometimes needed to indicate when learning is necessary, providing a prompt for the cycle to begin. In other words, feedback can help students notice when learning needs to take place by highlighting important experiences requiring reflection, or indicating when a behaviour or performance has deviated from an expected standard.[19] Sadler[1] explains how feedback can also be used to accelerate learning, guiding students through the process, rather than waiting for students to navigate the learning cycle on their own. For example, a student may reflect upon an experience, aware that their performance was inadequate, but without knowing how exactly to improve. Feedback may provide the student with pertinent details for making improvements, allowing them to move through the stages of reflection, conceptualization, and behaviour change more quickly. Bangert-Drowns et al.[6] describes the cognitive impact of feedback where, in addition to leading to behaviour change, the information received through feedback stimulates improved metacognitive processes that lead to purposeful learning and self-regulation (i.e. the student's motivation to take an active role in their own

learning). As such, feedback has the potential to develop lifelong learning skills, as students improve their ability to navigate the experiential learning cycle more accurately and effectively on their own. Feedback is therefore not only important for learning within the context of academic institutions, but rather a necessary step towards developing competent and independent professionals.[7-9] In regards to clinical education specifically, Ende[7] explains that "without feedback, mistakes go uncorrected, good performance is not reinforced, and clinical competence is achieved empirically or not at all." (p778) This quote describes the need for feedback not only for remediation of behaviours, but also to reaffirm positive acts, and to ensure that competence is internalized for lifelong use. The theoretical value of feedback for higher learning is well established in the literature.

In addition to the theoretical importance of feedback, evidence-based research further supports the use of feedback for behaviour change and improving performance/achievement (with achievement being a measurable outcome of performance).[4-6,65] A meta-analysis on general feedback interventions found an overall small positive effect size for improved performance.[65] Another meta-analysis regarding feedback provided within testing situations consistent with institutional learning, found that feedback tended to have a similarly small positive, but significant, effect on achievement.[6] This research also demonstrated that the context and form of the feedback mediated the relative effect, leading the authors to conclude that under ideal conditions feedback's effects could be rather large.[6] A systematic review on physician's performance in clinical settings demonstrated that about three-quarters of feedback interventions yielded significant improvements.[4] Finally, a review on audit and feedback within the health professions also resulted in small, but potentially meaningful improvements to physician performance, with the baseline level of proficiency a key indicator of the relative effects of the

feedback.[5] Furthermore, in addition to the potential for improved performance, feedback may also lead to more accurate self-assessment,[20] and can be used to increase confidence.[21,66] Students also consistently show appreciation for feedback, and frustration when feedback is vague or missing.[29,43,66] These findings illustrate the continued efforts on the research of, and sustained need for, feedback in clinical education.

With the theoretical and practical implications established regarding feedback's effect on learning and its potential to improve performance, the operational provision of feedback within clinical education will now be discussed. A definition of feedback for clinical education is "specific information about the comparison of the trainee's observed performance and a standard, given with the intent to improve the trainee's performance."[2(p193)] Key components of this definition are the *intent*, *directionality* from an observer to trainee, and that the feedback content must reflect some *set criteria* (i.e. a standard). While upholding this definition, feedback can be provided in a variety of different manners.[3,8,9,67] Feedback may involve direct interpersonal interactions, which may be primarily verbal, or a combination of showing and telling, such as when a clinical instructor demonstrates to a student how they could improve a technical skill.[6,8] Feedback can also be provided through written mediums returned to students following an assignment or assessment, including online/computer-based materials.[4,8,49,67,68] Verbal feedback is often considered superior to written feedback for its immediacy, and for allowing a two-way interaction between the student and the instructor (i.e. the student can follow-up and ask questions).[68] However, newer research indicates that written forms of feedback may be preferred by students and in some cases more effective than verbal forms, as it presents information in a more objective manner (less threatening), does not interrupt students in the middle of a task (less distracting), and allows students to digest the information at their own

pace, including reviewing it on multiple occasions (more accessible).[67-69] Both types of feedback appear to be equally beneficial when used appropriately.[4] The literature also describes feedback by its extensiveness, differentiating between brief opportunistic encounters and more lengthy planned feedback interactions, with the latter referred to as formal feedback.[9] Both forms of feedback are considered important for learning, but formal feedback tends to be more readily neglected in clinical education.[9,10] In general, whatever the form or context of feedback, the purpose remains consistent: to improve student-level outcomes.

A majority of the feedback literature has also focused on the strength or quality of feedback.[1,2,7,8,67,69,70] Despite the potential impact of feedback on learning and performance, the literature also repeatedly reports instances where feedback does not successfully alter behaviour.[1,4,65,67] This trend highlights the importance of providing not just any feedback, but high-quality feedback. As such, several guidelines have been developed to aid educators in structuring feedback provision so that it will be more effective. Quality feedback is said to be *task-specific*, but also *generalizable* to future situations (i.e. a student must have the intention of performing the task again in order for there to be an incentive to using the feedback).[1,2,7-9,60,67,70] The further away feedback gets from a specific task and changeable behaviours, and the closer it moves towards personal characteristics, the weaker the feedback is considered.[2,7] Feedback should provide a clear link between the students' performance and course expectations, so students can make adjustments accordingly.[1,60] In addition, quality feedback should be provided in plain, non-judgemental language from a credible source.[1,4,5,7-9,60,67] Feedback is also considered more effective when it is expected by the students (i.e. they have been told ahead of time that feedback is going to be provided),[7] and when it avoids interrupting students currently engaged in a task (unless immediate intervention is required for safety reasons).[8,67] Feedback

should also be provided to students in a timely manner while the application of the feedback is still salient to the student.[1,7,8,60,71] It is believed that following these prescriptive guidelines for quality feedback will increase the potential for feedback to impact student performance.

The aforementioned quality feedback guidelines represent a majority of the literature on providing effective feedback to date. While they do provide valuable information, they may also present an overly simplistic view of feedback. The literature also depicts a multifaceted student-feedback interaction, showing that certain methodologies may be more valuable for certain types of learners in different situations[1,8,67]—illustrating the inherent complexities of providing ideal feedback. For example, delayed feedback may be more effective for high achievers but not for others, and emotional states can influence how feedback is interpreted.[8] Feedback must be utilized by the student with conscious and mindful effort, demonstrating that no matter how good the feedback may be, it is of little use if the student has no intent to use it.[6,64] This fact ties in several important psychological characteristics that may impact a student's use of feedback, including self-efficacy and self-motivation. Self-efficacy refers to a person's belief in their own abilities, and with regard to feedback, their ability to follow through on recommendations or suggestions.[8] Self-motivation refers to the perceived importance of the outcome an action will produce,[72] and thus the student's personal desire to implement (or not to implement) the feedback. Both self-efficacy and self-motivation are critical aspects of self-regulated learning.[72] Without fostering these additional attributes that drive student learning, feedback by itself may have little impact. The inherent complexity of feedback and its effects indicates that any new unique application of feedback requires proper evaluation before it can be deemed effective. Experts on feedback in professional education, Boud and Molloy,[3] have recently expressed that it is no longer appropriate to judge feedback only by its successful fulfillment of certain design and

implementation guidelines. Evaluative procedures are also required to reveal the student-level impact of the feedback. Ultimately, high-quality feedback should result in improved student outcomes.

Feedback and Reflection

In the experiential learning theory, reflection plays a critical role.[19,61] Without reflection, experiences themselves may have no direct impact on future behaviours.[19] In its most basic form, reflection is simply the active process of examining one's own experiences.[73] Reflection is considered key to higher learning as it allows a person to organize and connect different knowledges, experiences, and even attitudes, yielding improved critical thinking and problem solving capabilities.[9,73] Another aspect of reflection is the ability to evaluate or self-assess one's own performance, and identify learning needs—note that some authors make distinctions between self-reflection and self-assessment, while a majority consider them synonymous or overlapping concepts.[8,71,73] Reflection is thus another important component for improving student-learning outcomes.

Feedback has been shown to have an interesting interaction with reflection.[20,71,74] Several experts believe that while feedback can help correct/perfect technical skills, it is reflection that produces long-term internalized behaviour change leading to well-rounded independent professionals.[9,29,71,75,76] The literature indicates that quality feedback can encourage reflection, and even result in improved reflective capacities, such as a more accurate self-assessment of performance.[8,71,77] Reciprocally, structured reflections can encourage students to better engage with written feedback, increasing its the potential for improving performance by making sure the students attend to pertinent suggestions for behaviour modification.[20,71,74] Relating these findings to the experiential learning cycle, feedback can impact learning both before reflection (i.e.

feedback guiding reflection) and after reflection (i.e. reflection guiding feedback use). Thus there appears to be an important symbiotic relationship between feedback and reflection.

The current literature on reflection in medical education describes its clinical application. Reflection within this context is not only about examining an experience, but also about examining one's self, and one's own thoughts.[19,76] The goal of reflection in medical education is to process information from a one's experiences and ultimately act on that information to become a more responsible and capable professional.[19,76] Reflection is believed to have three main purposes: to improve learning (as described by the experiential learning theory), to improve interpersonal relationships with patients (by becoming aware of biases or attitudes which may affect a relationship), and to facilitate reflection-in-action (where a health professional can quickly make sense of, and act on, complex and unfamiliar information seen in their practice).[19] For these purposes, several methods of teaching and encouraging reflection have been developed. One method is guided reflection, where a mentor facilitates reflective thinking upon pivotal experiences.[9,19] Similarly, direct feedback provision is often considered essential to direct a student's attention to key events, so that the reflective processes can begin.[19] Another modern approach is the use of journaling and portfolios, where students create written works analyzing major events, or identifying learning needs and describing how to solve the deficits.[19] These written documents have the added benefit of allowing educators to assess student's reflective abilities. Finding opportunities to encourage reflection and improve reflective capacities remains an important part of clinical education.

As described in the previous section, feedback's effect on student-level outcomes must be evaluated before the quality of any feedback method can be definitively established.[3] The most straightforward measure of feedback's effectiveness, is through evaluations of student

performance before and after feedback provision. However, research also indicates that feedback may need to be provided consistently and over the long-term before its impact on performance can be truly ascertained.[3,4] Therefore, it would be prudent to include additional outcome measures to any evaluation project, in order to determine if the feedback is encouraging students in the right direction. Reflection provides an additional measure for evaluating the effect of feedback on student learning, since feedback can directly impact the reflection process.[71] Reflection represents both an intermediate step towards improved performance (as students first need to accurately identify a deficit before they can address it), but also a unique skill in and of itself that will carry over into their professional careers.[19] Written reflections following a key experience (such as a high-stakes assessment) can provide insight into how students are thinking about their performance. A student's written reflection can then be evaluated for both its content and its quality.[75] Content refers to *what* the student is reflecting upon, for example the specific parts of an experience they choose to focus on, and their ability to accurately identify strengths and weaknesses.[73,75] Quality refers to *how* they are reflecting, and is typically considered on a spectrum from simply descriptive to a deeper critical analysis.[19,71,75,78,79] For example, written reflections that consist of statements merely describing what happened during an experience would demonstrate low levels of reflection.[19,71,79] As the student begins to provide context to the descriptive statements, by explaining how their attitudes or emotions affected the experience, or by integrating external knowledge components into the experience, the student shows signs of higher levels of reflection.[19,71,79] At the highest level, students are able to indicate how their experiences will shape future actions.[19] Therefore, reflection is an important step towards improved performance and independent life-long learning.

History of Feedback in Structured Clinical Assessments

As described in the introduction, the summative nature of SCAs means the provision of feedback for these assessments has been minimal to date.[12] However, with the pedagogical push for feedback for all assessments,[8,20,21] and the understanding that SCAs comprehensively capture important information on clinical and professional abilities,[10] some educators have begun to incorporate feedback opportunities within their SCAs.[10,21,80-84] The most traditional method of feedback provision for SCAs is the immediate observer form, whereby the examiner verbally provides praise or informs the student how they could have improved their performance.[29,80-83] This feedback commonly takes place during the examination itself (in between stations), but may also be part of a debrief following completion of the examination. Another common feedback method is to provide students with their checklist results, so they can identify the unsuccessfully completed tasks.[80] Examiners might also provide written comments regarding student performance on each station for students to review at a later time.[10,84] Other possible forms of feedback have been explored, but are not as commonly used. Video feedback can be provided where students view exemplar performances or self-assess their own performance.[20,80] With the utilization of technology to provide feedback following an SCA increasing, additional approaches including use of audio recordings made by examiners in between stations, or online reports summarizing student performance across the examination.[10,21] There is some evidence to support the effectiveness of SCA feedback as a means of improving clinical performance, at least short term.[20,81] Students also consistently report appreciating the feedback received during/following an SCA.[10,21,43,82-84] Therefore, continuing to provide feedback for SCAs is an important endeavor.

When evaluating the quality of the feedback typically applied within an SCA, several issues can be identified. Providing only verbal feedback within the confines of the assessment itself has several limitations. Prescriptive guidelines indicate that feedback should not interrupt student's completion of a task, which may include completion of the SCA itself.[67] Furthermore, feedback provided during stressful situations, where emotions may be high, can actually have a negative impact on students.[8,68] Hollingsworth and colleagues[82] also demonstrated through their research that verbal comments received within an SCA were equivalent in amount across all types of students. These findings indicate that examiners may be hesitant to overload poorly performing students with too much negative feedback during this small time frame, and therefore not all issues will be addressed.[82] Their study also showed that positive comments tended to be global in nature (e.g. "very well done") without specific accomplishments being reinforced. Consequently, providing only verbal feedback within the examination itself is not ideal. In order to avoid some of the issues with mid-examination feedback, some educators provide a verbal debrief with groups of students following completion of the SCA.[80] However, providing feedback in group settings is often ineffective as students may rationalize that the feedback is meant for others and not for themselves.[68] As such, providing detailed written feedback following an SCA may be beneficial.[68,69] Currently, when feedback *is* provided at the end of an SCA, it typically involves written notes by station examiners, which is likely provided in a hurried and incomplete manner as the time between stations is limited to only a few minutes.[10] Furthermore, no in-depth or cross-station analysis is conducted, so overarching student strengths and weaknesses are not identified. This issue also applies to using the checklist items themselves as a form of feedback, which has also been criticized for its potential to promote rigid and memorized thinking patterns.[13] There is still an unfulfilled need to find opportunities to provide

formal quality feedback in clinical education.[9] When formal feedback is provided following an SCA, it is typically restricted to those who have failed the assessment.[10,20] As such, students who are adequate, but perhaps not excellent, and students who are struggling, but managing to pass, receive little structured information on their clinical performance. Efforts need to be made to guarantee *all* students are receiving quality feedback following an SCA.

There are two overarching issues that often preclude the provision of quality feedback following an SCA. The first is timing. Providing feedback itself can be a time consuming endeavour requiring extensive faculty commitment for preparation and provision.[8,20] Moreover, it is important to provide feedback to students in a timely manner.[1,7,8,60,71] For an SCA, the time it takes to determine and approve individual results, and then compile and release feedback to the students, means that written feedback following an SCA is often too late to be impactful.[84] Notably, electronic administrations of examinations have begun to address this issue by decreasing administrative time demands.[49,55,56] The second major issue is test security and item confidentiality. Development of a reliable and valid SCA is labour intensive, and therefore the same (or similar) examinations are used over and over, throughout the academic year, or across years.[16,29,43] The outcomes of these assessments are frequently high-stakes (i.e. student competence, and ability to progress within the program, are being determined).[20] As such, there is a reasonable security concern that providing students with certain types of feedback, such as the assessment checklist itself, could result in contamination of future SCA results and thus the validity of subsequent decision-making.[85,86] Therefore, novel approaches to providing feedback following SCAs are needed.

Diagnostic Score Reporting

Even with a comprehensive understanding of what quality feedback should look like (e.g. timely, specific, concise), a framework for how to consistently provide this type of feedback to all students following an SCA is still required. One such framework is diagnostic score reporting (DSR), which is a method of breaking down test results into learning domains so students can receive more specific information on their performance. Before delving into the process of DSR, the literature on score reporting in general will first be described. The initial focus on score reporting emerged from a need to effectively report the outcomes of large-scale assessment taken in kindergarten to grade 12 (K-12) to relevant stakeholders such as parents, educators, and test-takers.[23,25,27,87] These score reports are typically single page summaries of assessment results, and much of the literature has focused on their usability and interpretability.[23,25-27] Suggestions include, a need for clear and concise language, proper definitions of terms, esthetically pleasing designs (e.g. use of colour, headings, and white space), and graphical representations of data, in addition to narratives, to facilitate an accurate understanding of the major findings.[23-27] A quality report will highlight both positives and negatives,[24,27] and provide cohort comparisons to help stakeholders understand the students' relative standings.[23,26] When designing score reports, it is also essential to consider the unique needs and abilities of the intended users of the report.[23,25-27,87,88] A score report for different groups of people would need to emphasize different pieces of information (e.g. statistics vs. narratives), and reports need to take into account the group's qualities such as their age, cohort, and relevant knowledge and expertise.[88] For example, providing standardized zed-scores may be beneficial for test developers or psychometricians, but may confuse a parent or student. Finally, the importance of timeliness in providing score reports is also regularly cited.[25-27] These suggestions illustrate that a majority of the guidelines for

providing quality feedback in higher education are compatible with the characteristics of effective score reporting.

DSR expands on the score reporting literature, with an emphasis on making improvements at the individual level.[23,24,89] DSR provides test-takers with information on global performance, performance by domains, relative standings to professional standards and/or peers, and specific suggestions for improving performance. Domains reflect the specific areas of knowledge, skill, and ability (KSA) the examination intends to capture, so that student strengths and weaknesses can be readily identified.[23-26,88] These diagnostic domains (sometimes referred to as sub-scores) are defined by the relevant standards/competencies guiding the examination's purpose and development (which will be discussed in detail under the test specifications section). Domain scores are determined from a subset of test items that reflect unique KSAs, and provide additional information to the student beyond a single summary test score.[25] For example, consider a high school mathematics examination in which a student performed somewhat poorly (e.g. 65%). Perhaps the student did well on questions requiring trigonometry and calculus skills, but struggled in the area of probability and statistics. Without this detailed information, the student may erroneously conclude that (s)he is "bad at math", when in fact the student possesses an inclination for certain types of math skills. These certain mathematical skills would aid the student in an engineering career—but perhaps the student should avoid a career as a statistician. Alternatively, diagnostic domains may be based on hierarchical cognitive abilities, such as basic knowledge, application of that knowledge in simple contexts, and application of that knowledge for critical problem solving.[24] This structure allows students to ascertain what level of mastery they have achieved in a specific knowledge area, and where they need to focus their efforts to advance to the next level.[24] The choice of diagnostic domains as subject-based, hierarchical, or

even skills-based, must be carefully and purposefully determined depending on the intent and the content of the examination. Regardless of the domain format, the goal is to provide more specific details of examination results so that students can better understand their strengths and weaknesses.

The other chief component of DSR is to provide clear and concise information on how to improve within these domains.[23,25] Simply informing students of their strengths and weaknesses may not be enough to encourage change at the individual level, especially if the student does not know where to access resources for making improvements. These resources can be as simple as referring the student to an appropriate educator or textbook, or more detailed, such as directly providing test answers and supplementary information on the score report itself.[24,27] Overall, DSR includes a summary score, scores by domains, relative standings, and suggestions for improvements, all while maintaining general quality characteristics of effective score reporting.

The use of web-based score reports are beginning to increase in popularity, although they have not been as widely implemented or studied as paper-based reports.[23,25-27] Several advantages of online DSR have been speculated. First and foremost, they facilitate increased ease and timeliness in reporting results.[26] The amount of information that can be provided to stakeholders through online means is increased, and theoretically limitless[23,25,88]—although providing too much information may detract from key findings, and thus may also serve as a potential disadvantage.[23,25,88] More interactive was of displaying data are also possible, such as using online games and quizzes, which may encourage increased student interaction with their online feedback.[25,88] Finally, additional learning resources and references can be provided and accessed through web-links,[23,27] providing easier access to information that can assist students in

improving performance. Despite these potential benefits, the impact of web-based DSR (and, in fact, the impact of DSR in general) on student-level outcomes has been largely unresearched.

Lastly, it is important to acknowledge that DSR development is an ongoing and iterative process.[25-27,87,88] Reports require continuous monitoring, and can be modified and improved upon over time. As a first step, pilot reports need to be validated prior to full scale implementation.[25,26,88,89] Their usability and effectiveness should then continue to be evaluated across administrations.[25,26,88,89] An important part of the evaluation process as described in the DSR literature, is to solicit the opinions of the report users themselves.[27,87,88] Several measures must be taken to ensure the objectives of DSR are being met.

Test-Specifications and Diagnostic Domains

Establishing the diagnostic domains that are to be reported is fundamental to DSR. The literature on test development provides a guide for how domains can be identified for large-scale assessments. These assessments are designed (or should be designed) using detailed test specifications, which outlines *what* is to be assessed, *how* it will be assessed, and *why* it should be assessed.[86,90] Throughout the stages of test development, an assessment blueprint is eventually generated that shows links between the purpose of the examination, the KSAs being assessed, the test items measuring those domains, the claims that will be made from the test scores, and supporting evidence validating those links.[90,91] Evidence-centered design describes a layered approach to this blueprint development.[91] First, information must be collected to identify the KSAs the assessment must cover. Information can be obtained from a variety of sources including expert opinions and experiences, real-world observations, analysis of any previously collected data, course materials, and/or professional standards.[91] Next, this information needs to be organized into a logical narrative, grouping the KSAs into key overarching domains, that

describes how student performance on each domain leads to accurate inferences about student competence. Consideration of counter-arguments is also encouraged (e.g. we can assume A, unless B), to identify potential sources of error confounding the testing claims.[91] Once the domains of the test are established, the specific elements of the examination itself can be determined, such as the most appropriate test form (e.g. multiple choice versus essay), the number of test items needed to make conclusions about each domain, the content of those test items, and the logistics of how the assessment should be administered (e.g. time, location). Finally, the actual test items are developed (reflecting the information collected in step one) and mapped appropriately to the existing domains to yield the final assessment blueprint. Examinations can be piloted, and the results collected and analyzed to help confirm the validity of the examination, and to make further improvements. The evidenced-centered design method creates a map demonstrating how test results validly measure student KSAs.[91] Since the assessment blueprint contains the diagnostic domains and the test items that determine performance on each domain, this information can then be reported to test-takers through DSR.

DSR was specifically designed for large-scale national assessments, where the process of test development is rigorous, meticulous, and based on very detailed test specifications.[90] For smaller-scale examinations (such as many SCAs in higher education), the predetermined test specifications may be more limited. Typically, the fundamental content of most health professional curricula is established to some extent in regard to learning objectives, content areas, compulsory assessments, and more.[58,92-94] With an ideological shift towards competency-based clinical education, a number of core competencies are now being defined by professional and regulatory bodies to align curricular development across educational institutions.[95-101] The process of blueprinting is also encouraged throughout the health professional

literature.[13,28,29,35,42,58] However, even when an assessment blueprint has been established, the comprehensiveness and adequacy of this blueprint for DSR will vary.[35] For effective DSR there is an important trade-off between domain specificity, and having enough test items representing each domain reliably.[25,89] If domains in a blueprint are measured by many test items, the domains might be very statistically reliable, but they may also be broad and incapable of providing specific enough feedback to students to be useful. For example, if the assessment blueprint contained the domain knowledge, reporting scores on this domain might be appropriate for a single-subject examination, but may need to be broken into separate domains if the examination measured different types of knowledge (such as in final examinations that typically cover several independent topics). Alternatively, if the assessment blueprint shows only one or two test items mapped to a domain the information may be highly specific, but the domain scores would likely be unreliable (e.g. it would be inaccurate to say a student is a poor communicator when they got a single test item reflecting communication wrong on a test).[25,87,89] As such, blueprinted domains may need to be regrouped (or redeveloped entirely) into informative and valid domains. Such issues may also reflect inadequacy in the initial test-development process (e.g. where insufficient items were introduced to cover the breadth of each learning domain), and the test itself may require modification. Finally, despite a well-developed curriculum and established professional competencies, a majority of courses still rely on the course instructor alone to develop the assessments based on his/her judgement of how to best determine student achievement in that subject.[93] Thus, while a logical rationale for the assessment content was likely used, there will often be no formal blueprint created linking test items to original learning objectives. As such, for DSR to be implemented into smaller-scale SCAs, an adapted assessment blueprinting process, similar to the one described above, is necessary. Specifically, a blueprinting process is

needed that focuses on the goals of DSR and providing students with useful information on valid diagnostic domains.

Diagnostic Score Reporting and Structured Clinical Assessments

DSR presents a promising framework for providing feedback to students following an SCA, as the major barriers to quality feedback can be addressed. Structuring the feedback by diagnostic domains will provide an in-depth analysis of strengths and weaknesses to students, without needing to reveal the test items themselves (i.e. examination confidentiality is maintained). Having a predetermined structure for the score reports established prior to the examination will help facilitate timely feedback. Timeliness can be further addressed through online administration of these reports, which would theoretically allow students to receive feedback immediately following completion of an assessment. By improving efficiency, *all* students can receive diagnostic information—not just those targeted as requiring remediation—without drastically increasing the time-demands placed on instructors. DSR thus has several advantages as a feedback method for SCAs.

Elements of DSR have been slowly emerging in the SCA feedback literature.[21,43,84] Manogue, Brown, and Martin[42,43] describe their use of course objectives to create a blueprint of domains (including: knowledge, clinical reasoning, history-taking, and communication) to be assessed in their restorative dental formative SCA. Following the examination, students received additional feedback by these comprehensive domains such that strengths and weaknesses could be readily identified.[42] Survey results indicated the students felt the feedback helped them better understand their performance.[43] One of the issues in their feedback design was that whole stations (rather than test items) seemed to contribute to each domain score, and since each station measured several domains, the accuracy of their determinations are questionable. Alternatively,

Taylor and Green[84] utilized a different approach to domain development, by mapping each test item of their already established medical SCA to one of their nationally defined competencies.[96] This study is one of few that designed a randomized controlled trial in attempts to determine how their feedback would affect performance. However, their findings were not straightforward, and deemed educationally insignificant, which was largely attributed to a lack of specificity and a time lag in feedback provision.[84] Furthermore, it appears that neither of these studies included information for the students on how to improve their performance within the domains. Nor was there any mention in these articles regarding adherence to important reporting characteristics (e.g. domain descriptions, graphical representations of data) or esthetic considerations (e.g. color, white space, and headings). Thus both of these feedback methods did not appear to fully utilize either the framework of DSR or the guidelines for quality feedback.

Perhaps the best illustration of DSR in an SCA is the work done by Harrison and colleagues[21] at the Keele University School of Medicine. They designed and experimented with a new interactive feedback website for students following a summative medical SCA. Each station in their SCA was marked using the same nine rating scales based on the Generic Consultation Skills (GeCoS) Assessment Framework. This framework was previously developed and validated to help instructors teach and assess key nationally defined physician competencies.[96,102] Harrison et al.[21] were then able to use the nine GeCoS skills (e.g. clinical reasoning, building and maintaining the relationship, and organization) to provide students with information on each skill across the SCA stations, similar to providing domain scores in DSR. Their feedback website also presented scores by station/subject, with information presented in a variety of ways (pass/fail vs. overall), and graphical comparisons to expected standards and cohort averages. The website also had “next steps” pages which provided information on how to reflect upon and best utilize the

feedback (e.g. by setting SMART goals) to improve future performances.[21] In total, there were 130 different pages that each student could access through their feedback website.

The research conducted by Harrison et al.[21] focused on exploring student-interaction with the feedback. Accessing the website was not mandatory, and student logins, pages visited, and time spent on the website were tracked. Their results showed that 95% of the students interacted with the feedback to some extent, with almost half of the students returning to the site at least a second time. On average, 123 of the available 130 webpages were viewed by the students, at a pace of about six pages per minute. They also conducted a latent class analysis to determine student sub-groups based on usage patterns, which they defined as comprehensive users (50%), minimal users (27%), and selective users (23%). They concluded that most students engaged with online written feedback following a summative SCA and that half engaged thoroughly. Certain personality characteristics were also explored so see if they affected student use. They found that comprehensives users expressed a greater value of feedback in general, and that minimal users expressed a greater reliance on extrinsic motivation. Additionally, they noted interesting trends regarding student performance on the SCA and engagement with the feedback, with students who performed the best and worst visiting the most web pages. A worrying engagement pattern emerged where students who were *close* to failing utilized the feedback the least. Since students who *did* fail received additional sources of feedback during mandatory remediation procedures, the previous finding led the authors to conclude that those who arguably needed to review their feedback the most did so the least.[21] These findings may directly relate to the concept of self-regulated learning, where some students, typically lower achievers, are simply not motivated to seek out information that could help them improve their learning and achievement.[72] Nevertheless, the Harrison et al.[21] study was based only on quantitative usage

patterns, and it is difficult to make accurate conclusions about how the students actually interpreted, reflected upon, and utilized their feedback for the purpose of making improvements.

It is also important to note that DSR is not intended to replace other feedback methods, but can be used to supplement them. Harrison et al.[21] required students who failed their SCA to undergo remediation procedures. Taylor and Green[84] had some students (depending on their feedback group) receive both their domain scores (i.e. scores by skill) and written comments from examiners. The purpose of using DSR in clinical education is to make sure all students are provided some amount of quality feedback, and to enhance/reinforce the feedback some students already receive. DSR provides a stable written source of feedback students can reference and refer back to when needed, and therefore has added value to current effective feedback methods.

The research described above shows a promising start for providing feedback in SCAs through elements of DSR. While Harrison et al.[21] did not directly refer to their feedback website as a diagnostic score report, it most closely reflects the tenants of DSR by providing information by domains, showing comparisons to standards and peers, utilizing visual aids such as graphs and colour, and informing students about ways they could use their feedback to make improvements. On the other hand, the domain pages were somewhat limited, lacking in narrative information, domain descriptions, and specific advice on how to improve within each domain. It would also be prudent to consider whether including 130 pages on the website meant more useful information was being provided to students, or whether this might be a case of overloading students and/or distracting them with unnecessary information.[23] Taylor and Green[84] also illustrated the importance of keeping in mind the general guidelines of quality feedback in higher education, such as timeliness and specificity, in order for feedback to be impactful at the performance-level. All three of the previous studies described the importance of linking

examination content and their feedback to course objectives to improve validity,[21,42,84] with two of the studies utilizing national competencies.[21,84] However, a common methodology for developing diagnostic domains for SCAs is not established. Harrison and colleagues[21] were easily able to convert their generic rating scales into diagnostic domains, but most SCAs are assessed using carefully determined checklist items unique to each examination, requiring creation of a similarly unique assessment blueprint to establish diagnostic domains. Brown and colleagues[42,43] used their learning objectives to drive their feedback development, whereas Taylor and Green[84] focused on using the established test items to drive competency selections. Providing a practical general framework for determining competency based diagnostic domains for DSR would encourage other SCA administrators to adopt DSR within their assessments, without falling into potential shortfalls such as invalid domain development or incomplete score reporting. Finally, with a limited number of studies available, there is still a general lack of evidence-based information on how DSR in SCAs impacts student-level outcomes. As such, more research on applying DSR within SCAs is required.

Validity and Score Reporting

Throughout the diverse research described in this review, there has been a common refrain of the need for, and the importance of, validity. In psychological and educational test theory, validity refers to whether the test is actually measuring what it intends to measure.[103] Historically, validity was reduced to single sources of evidence depending on the theoretical framework used.[104] In criterion-validity it was necessary to establish a relationship between a measure and its outcome, for example a university admissions test score should correlate with success at university.[104] Content-validity required expert opinions to establish or verify the appropriateness of test content.[104] Many researchers, even to this day, have focused on validating

tests and psychometric instruments using only statistical measures of reliability and internal consistency.[105] Today, an argument-based approach to validity is considered the gold standard, with logical discourse and multiple sources of evidence necessary to validate the interpretation of test scores (rather than the tests themselves).[103,106] Theoretical frameworks, appropriate test content, statistical measures of internal structure, and establishing relationships to outcome variables are now all considered important sources of evidence toward a single validity argument.[105] The more sources of evidence, the stronger the validity claim.[106] The consequences of the test scores are also important to consider, with tests that yield higher consequences (such as medical licencing examinations and their impact on public safety) requiring stronger evidence in order to be considered valid.[106,107] Validation is thus an ongoing process, and as more and more evidence becomes available, the claims either become stronger, or the tests (and maybe even the theoretical frameworks) are appropriately revised.

Several methods of providing validity evidence for effective score reporting and diagnostic domain selection are available (several of which have already been described above but will now be combined and explored in greater detail). Reporting content can be validated by using a literature-based framework to establish an evidence-based design.[90] Domain selection can be guided through use of expert and student opinions, and pre-validated clinical competencies, to ensure intended learning outcomes are reflected in the feedback provided.[25,84,88,107] The Delphi technique, an iterative survey based approach to reaching agreement between multiple experts, can also be used to further validate the results quantitatively.[108] In this technique, several experts independently evaluate the results of a qualitative analysis by using a numerical rating scale, and the findings are then adapted until the average/median ratings of the experts is acceptably high. The benefit of this technique is that it

provides a certain level of confidentiality to the experts, prevents persons with the loudest opinions from controlling the decision making process, and provides an objective numerical grade to the appropriateness and reliability of qualitative content.[108] Additional quantitative/statistical techniques can also be used to confirm domain validity, such as measures of internal consistency and correlations between domains—if diagnostic domains are truly measuring different skills (as opposed to one overall skill) then the results should *not* be highly correlated.[109] A key validity source for effective score reporting will be the outcomes of an evaluation study to provide evidence for the ultimate claim that the reports can facilitate student-level improvements.[106,107] The results of surveys on student opinions of DSR can also help to validate the response process of the report—which relates to how students think while using the feedback—and verify usability of the reports from the user perspective.[88,105] Therefore, all the information provided throughout this thesis is ultimately working towards a validity argument for applying DSR in SCAs. It must be reiterated that the goal is not to specifically validate DSR, as validation is an ongoing process, but rather to provide sources of evidence, both for and against, so that DSR can be iteratively improved to achieve its ultimate goal of aiding student learning.

Summary

SCAs are an essential component of health professional education, as they assess not only if a student has knowledge, but whether or not they know how to apply that knowledge competently. These examinations are intense, thorough, and capture a tremendous amount of information on student ability, and yet this information is rarely provided to the student as feedback. Quality feedback has been linked to improved reflective capabilities and clinical performance, and is considered a cornerstone for developing students into responsible, capable

professionals. Nevertheless, time limitations and test security make feedback provision difficult for SCAs. Online DSR presents a mechanism for facilitating quality feedback to all students in a timely manner following an SCA. DSR is a method of describing student performance by the underlying domains the test intends to measure (rather than by the individual test items themselves). While DSR has been described rigorously in the K-12 literature, there is a lack of information on how to apply this mechanism in higher clinical education. A general framework for applying DSR into SCAs is required. Furthermore, the majority of research on DSR (and feedback in SCAs in general) has focused on student opinions of, or interaction with, the feedback. Experimental studies determining the effect of DSR on student-level outcomes are needed.

Chapter 3: Methods

The purpose of this study is to describe a general framework for applying DSR within SCAs and to determine the effect of DSR on student-level outcomes. Specifically, the research questions to be addressed are:

1) To what extent can the framework of DSR be applied to provide all students with valid and high-quality feedback following an SCA?
2) How will online DSR impact student-level outcomes such as reflective capabilities and clinical performance?

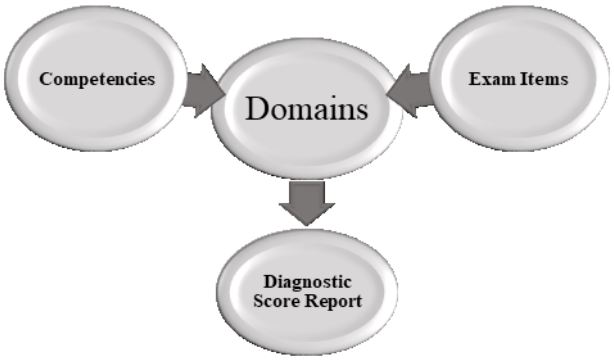
The dental hygiene history taking SCA at the University of Alberta provides the context for a course-specific application of DSR. Therefore, the specific contributions of this research are: to provide a general framework for using DSR to provide quality feedback in SCAs; to develop and validate a course-specific diagnostic score report for a dental hygiene SCA using the general framework; and to pilot online DSR in a dental hygiene SCA to assess DSR's effect on reflection and performance.

This methods chapter will be divided into sections detailing the process for each of the three major contributions described above. The first section of this chapter will draw on the extensive literature review, to provide a general process for applying the DSR framework into SCAs. The second section will detail the dental hygiene context for applying that framework. The third section will describe the experimental procedure used to evaluate the effect of DSR on student-learning. Finally, a section on the methodology for measuring two secondary outcomes will also be provided, addressing information revealed by the evaluation study that is of scholarly interest, but does not directly reflect the original research questions.

Framework for Diagnostic Score Reporting in Structured Clinical Assessments

The basic framework for DSR in SCAs can be seen in Figure 1. The main structure involves linking competencies and examination items to the diagnostic domains that the SCA intends to measure. The diagnostic score report is then generated from the student's score within each domain (i.e. the number of test items within that domain the student successfully completes) and the specific output each developer chooses to show the students (e.g. narratives, graphs, ways to improve, and so on). Based on previous work within the DSR or SCA literature, there are several overarching recommendations that may be extrapolated to help guide this process. First, the output should be student-centric, with domains and content established to provide the most meaningful information to the students. Second, DSR developers are encouraged to utilize their expert knowledge and experience to ensure that DSR is meeting the unique objectives of their SCA. Finally, evidence-based literature and validation strategies should be sought out to guide development of authentic and useful reports.

Figure 1. Framework for Diagnostic Score Reporting

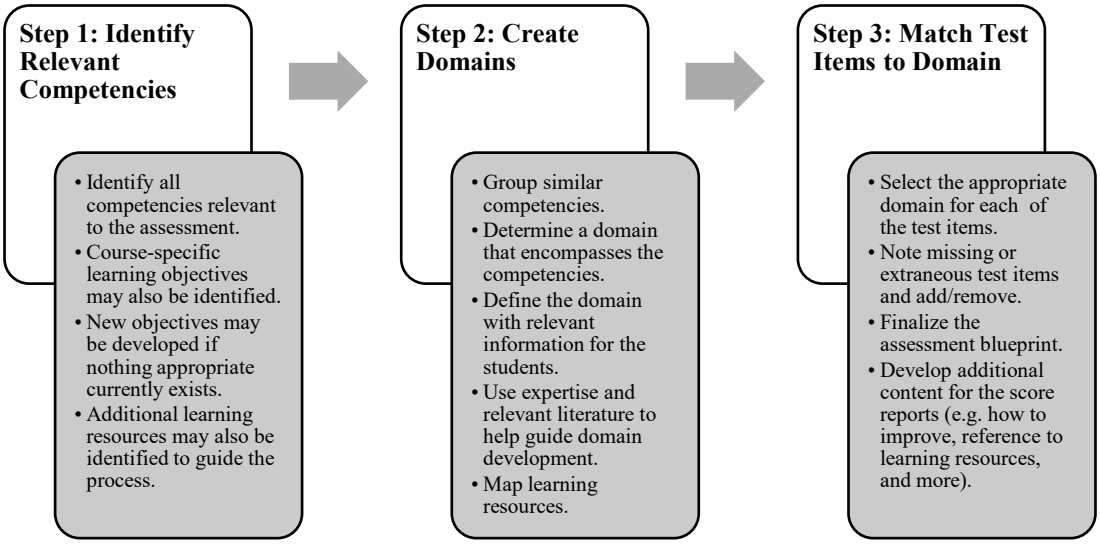


Assessment Blueprinting: Two specific methods of establishing diagnostic domains for assessment blueprinting within an SCA are available: top-down or bottom-up. The top-down

process is illustrated in Figure 2. This approach is based on the evidence-centered design method,[91] but has been specifically adapted for DSR, as opposed to test development (i.e. it assumes the test has already been developed). All competencies (and/or learning objectives) that are relevant to the examination should be identified. If detailed competencies are not available, then assessment-specific learning objectives will need to be developed. Development of these objectives should include a subject-based literature review, expert opinions, and validation strategies such as the Delphi technique for expert consensus—for this purpose, additional resources and methods that are out of the scope of this document may also be required. The identified competencies should then be grouped together based on similarity or shared characteristics. This grouping process can be considered a basic form of iterative thematic analysis.[110] Developers should read through the competencies several times until overarching themes emerge. Each competency can then be coded for its most relevant theme and organized accordingly. It is important to note these competencies could be rearranged in different ways. In contrast to traditional grounded theory thematic analysis, where researchers are supposed to be guided only from the data,[110] developers are encouraged to use their expertise, and consider the student perspective and form/goal of the assessment to decide on the most appropriate groupings (e.g. by subject, skill, or cognitive hierarchy). Finally, the overarching theme of these groupings will need to be defined, establishing the key diagnostic domains. One method of ensuring that appropriate domains are established is to enlist the help of the students themselves, and/or other persons with related expertise to confirm the adequacy of the development process and outcomes. Recruiting additional stakeholders reduces the personal biases any one developer may inadvertently bring.[110] Once domains are established, each test item needs to be mapped to the most appropriate domain. There should be several test items representing a single domain with

more items increasing the domain's reliability.[89] This process can also be verified by other experts to ensure its accuracy. The mapping of competencies and test items to diagnostic domains represents the main framework to which additional and individualized elements can be linked for DSR.

Figure 2. Assessment Blueprinting Process for Diagnostic Score Reporting



The top-down process described above has the benefit of identifying if there are missing test items or insufficient items to accurately measure the domains the SCA is trying to capture. For example, communication skills are frequently cited as an important competency for the vast majority of patient-clinician interactions, but actual test items to assess communication are infrequent.[29] The developer may recognize a need for additional test items to improve the comprehensiveness of the SCA. This method can also identify test items which do not appropriately match the current assessment objectives. The developer will then need to determine if the remaining test items represent important competencies that have not been properly identified, or if the items are extraneous, presenting sources of error rather than

information when determining the students' scores. If the latter, these items should be removed from the assessment to improve the accuracy of the scoring determinations.

The alternative blueprinting process is a bottom-up design. In this case, the test items themselves drive the domain selections. Specifically, test items can be grouped based on similarity, and the overarching theme of those items used to establish domains. Relevant competencies can then be mapped to those domains as sources of evidence. The test items could also be mapped individually to competencies and then those competencies grouped and used to establish domains, continuing with the top-down process described above. This test-driven process has advantages in efficiency and ease of domain development, however it lacks the increased validity the top-down process provides to the assessment as a whole—and thus the feedback that will be provided to the students. Therefore, the top-down process (or a combination of approaches) is strongly recommended.

Report Creation: Once the assessment blueprint is established, the specific information the developer wants to provide to the students in their report will need to be prepared. While the potential information that can be included in DSR is theoretically limitless, a minimum number of fundamental components should be included. The purpose of the SCA overall, and the domains individually, should be clearly defined in the reports. Information on how to improve within each domain should also be included, such as specific details relating to the test items themselves (depending on the degree of confidentiality required), references to textbooks and other learning resources, and/or ways to interpret and use the feedback to improve performance. Collecting SCA-relevant learning materials at the start of the blueprinting process, and directly linking each resource to a diagnostic domain, will aid in the decision of what information for making improvements can be provided to the students. Where appropriate, developers should

include graphs, colour, headings and white space to increase the general usability of the reports. Indicating relative standings to expected achievement levels and cohort comparison are also encouraged so students can better gauge their level of performance. Ultimately, the final content and design is at the discretion of the developer, but developers are encouraged to use an evidenced-based rationale for each piece of information they provide within their score reports.

Development of a Diagnostic Score Report for Dental Hygiene

The Dental Hygiene SCA: The dental hygiene program at the University of Alberta currently requires one year of prerequisite courses prior to admission, followed by two years of dental hygiene specific instruction (junior and senior years), with an optional fourth year for degree completion. The senior dental hygiene year is thus the final year of instruction before a majority of students will enter into professional practice. During the midpoint of this final year, the senior dental hygiene students must take a history taking SCA, which assesses the development of their client-clinician interaction skills. This year-end dental hygiene history taking SCA was chosen to pilot DSR. This SCA is a comprehensive single client assessment, where students must use their interpersonal communication skills to establish a rapport, conduct a full health and dental history, and identify any risk factors contraindicating or requiring modifications to dental hygiene therapy.[111] The client is portrayed by a trained SP, who receives detailed information prior to the SCA about the client's demographics (e.g. age, employment), medical health and background, oral health status and beliefs, and even personality cues. These SPs are also prompted to ask certain questions, to guide the conversation and ensure an equal opportunity for students to demonstrate their knowledge and communication skills. This SCA is evaluated using a grading checklist with a list of observable items describing key messages students are required to communicate, and other expectations of the interaction with the client—

marked as completed or not completed (yes/no). There are also several rating scales to be completed following the SCA, some by the examiner and some by the SP themselves, to assess more global skills demonstrated throughout the interaction (e.g. organization and communication). The senior dental hygiene SCA is considered high-stakes in that history taking skills are applicable to every appointment, and improper completion could lead to serious complications, such as a medical emergency.[111,112] The SCA also summarizes a student's progress in the program, with students who perform poorly receiving one-on-one consultation with the course coordinator. The assessment is summative in nature (taking place at the end of the year/semester), but these is still one term of clinical practice remaining where students could apply their SCA results and feedback towards continued learning within an academic institution.

The senior dental hygiene history taking SCA requires and assesses a variety of essential practitioner skills, however, the amount of feedback that the students receive, outside of those who perform very poorly, is minimal. Currently, there is a guided group debrief following each round of the SCA, where four to five students discuss their experiences and findings with each other and the clinical course coordinator. Students later receive scores on their overall checklist results, an average of the examiner's rating scales, and an average of the SP's rating scales. Students also receive their mark on a written reflection they are expected to provide within days of their SCA experience, which is graded from zero to five. These final numerical grades provide little meaningful feedback, with no explanation of why students received their scores, how those scores relate to their abilities, and where their scores stand relative to their peers and the program objectives. One of the major issues identified for providing feedback in this SCA is the confidentiality of the test items, as a similar assessment is used each year. DSR thus provided a useful option for giving additional feedback to all dental hygiene students following their SCA.

Assessment Blueprinting: DSR for the senior dental hygiene SCA was conducted using the general framework described above. The Canadian dental hygiene entry-to-practice competencies were chosen to guide the assessment blueprinting process.[95] These competencies were meticulously developed using a review of the literature and input from twenty-two dental hygiene specialists from across Canada.[113] The goal of these competencies is to help unify the national education standards of a rapidly evolving profession. The University of Alberta dental hygiene program has already adopted these competencies across their curriculum, in the hopes of providing the students with an increased understanding of their professional identity. Eight different overarching roles and responsibilities of a dental hygienist have been identified, with over 100 specific competencies describing those roles.[95] This extensive list of competencies provided an excellent tool for guiding the assessment blueprinting process during DSR development.

Two researchers with dental hygiene expertise, and experience with the history taking SCA, independently conducted an assessment blueprinting process using the Canadian dental hygiene competencies.[95] One researcher (a former student who previously took the same SCA) conducted the top-down process as described in Figure 2, and the other (the course coordinator) conducted a bottom up process. The two researchers then reconciled the blueprint to see if, a) the same competencies were chosen by both researchers, and b) whether the test items were mapped to domains that included their most relevant competency. It was decided that multiple competencies could be reflected by one test item, and thus the competencies that were chosen by the first researcher and not the second were retained or eliminated through discussion and consensus. The coverage of the test items representing each domain was also reviewed to see if missing or unnecessary test items were present.

An additional procedure for validating the blueprint was undertaken using a modified Delphi approach to expert consensus.[108] Four dental hygiene clinical instructors with SCA familiarity were recruited to complete a blueprint validation questionnaire. These experts were independently asked to rate all test items on their appropriate level of fit within the assigned diagnostic domain based on a five-point scale from zero to four (no fit to excellent fit). An example of the validation scale used can be seen in Appendix A. Additional space was available for reviewers to make any general comments about the skills-based domains. Any item that received a comment, a score of two or less from any examiner, or a median score of three or under, was reviewed. Changes were made to the blueprint where appropriate.

Report Creation: The DSR output was determined using the prescriptive guidelines described in the general framework above. A third researcher reviewed the final report for an objective perspective on clarity and interpretability. The DSR output for the students included all the information from the blueprint (excluding the test items), course expectation and cohort comparisons, and information on how to improve. Careful consideration was taken regarding aesthetics, such as use of headings, white space, and colour.

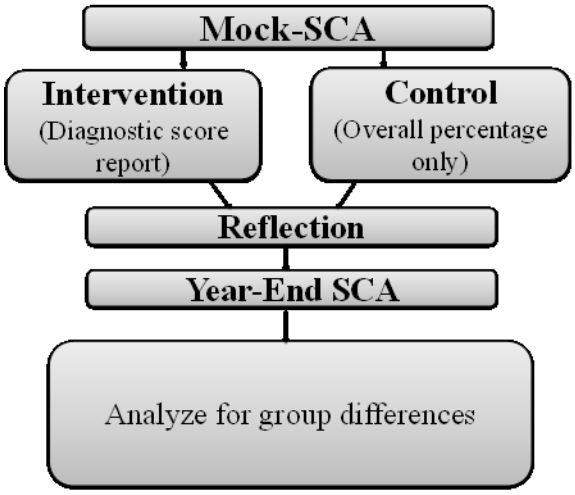
To facilitate increased timeliness in feedback provision, DSR was provided to students through online means. The University of Alberta has an established electronic clinical skills assessment system (ECSAS) for electronic administration of SCAs. SCA grading checklists are incorporated into an online platform so that SCA examiners can access these checklists and grade students through use of iPads. Once the examiners submit a graded checklist, this assessment data is uploaded to the online system for storage and analytics. The students interface with the system through a separate portal, where they login to view their results. This student interface was expanded to include the DSR output. The assessment blueprint provided a formula

for calculating domain scores instantly following submission of checklist results. By incorporating DSR into the ECSAS, reports could be immediately generated for all students following completion of the SCA.

Evaluation of Diagnostic Score Reporting on Student-Level Outcomes

Study Design: To evaluate the effect of DSR on student-level outcomes, a strict experimental design was necessary. A mock-SCA was designed so that DSR could be randomly assigned to half a dental hygiene student cohort, creating an intervention and control group. The intervention group received the DSR output with their individual results (which will be described in the results section), while the control group received only one overall percentage of their SCA performance. After receiving their results, all students were asked to reflect upon their mock-SCA performance, and later completed their regularly scheduled year-end SCA, so the relative effect of DSR on reflection and performance could be ascertained. Figure 3 illustrates this experimental process. A follow-up survey for the students was also designed and implemented to help validate DSR and the study results. Ethics approval for this research was obtained from the University of Alberta Research Ethics Office (Pro00062297).

Figure 3. Experimental Design for Evaluation of Student-Level Outcomes



Mock-SCA Development: The mock-SCA was developed to cover the same diagnostic domains as the year-end SCA, but with different content. The client's demographics, health history, and oral health perceptions/concerns were thus unique, but the skills required to conduct the complete history taking were the same. Two experts developed the content using their dental hygiene knowledge and experience, and additional course resources such as the University of Alberta dental hygiene clinic manual and the Darby and Walsh[114] dental hygiene theory and practice textbook. This process involved developing: instructions to the SP that described in detail the history of the client and how they should interact with the student, the grading checklist for evaluating student performance, instructions to the examiners for grading and facilitating the assessment, and instructions to the students for preparing for and taking the SCA. The SP instructions can be found in Appendix B (note the majority of the Appendices use the term OSCE in place of SCA as a more familiar term). Where applicable, the content and structure mirrored the actual year-end SCA to improve the practical validity of the mock-SCA and thus generalizability of the results. The goal was to assess the impact of DSR in a

typical/natural assessment environment. The dental hygiene clinic coordinator reviewed all mock-SCA documents to confirm their appropriateness for the student population.

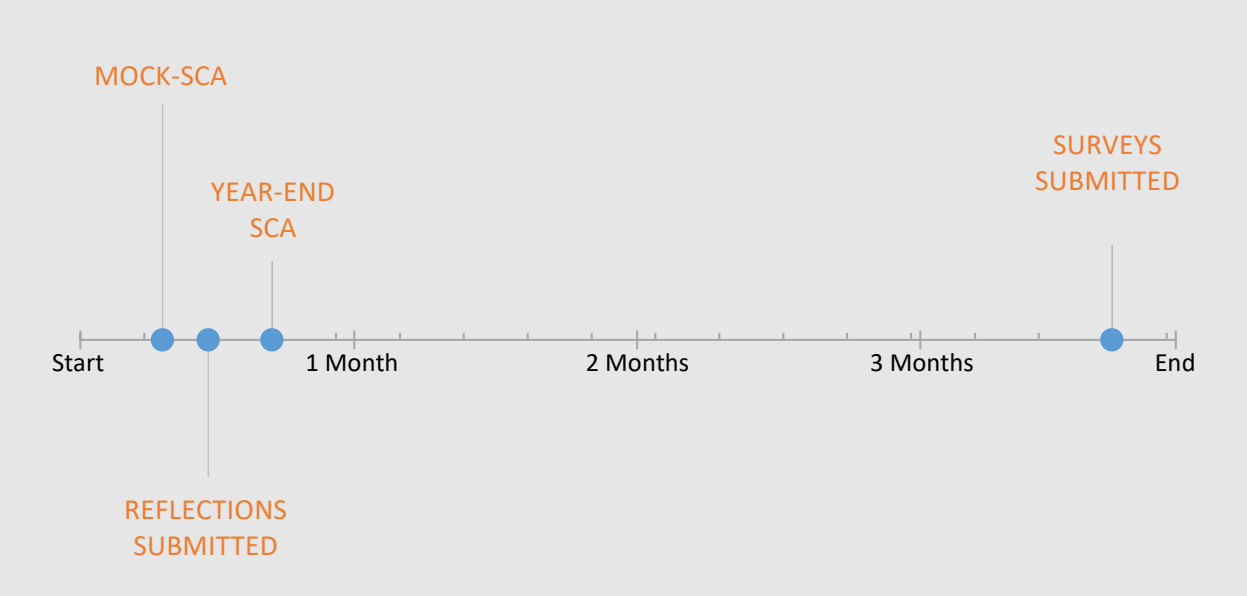
The mock-SCA test items were developed to cover the already established diagnostic domains in a similar fashion (e.g. in style and number) to the year-end SCA test items. In total 26 checklist items, and 5 global rating scales were developed. A researcher who was not involved in item development verified their appropriateness within each domain. The reliability of these links was further validated using the same modified Delphi technique and criteria described above for the assessment blueprinting of the year-end SCA (similar to Appendix A but with mock-SCA test items instead of competencies). Three test items received a median score below three, and one more item received a score below two by at least one examiner. After review, one test item was removed from the examination (resulting in 25 checklist items), one was reworded, and one was moved to a more appropriate domain. The result was a validated assessment blueprint for the mock-SCA, with test items mapped to diagnostic domains. The DSR output for the students in the intervention group was identical to the output designed for the year-end SCA, except that domain scores were generated using the mock-SCA blueprint.

Participants: The senior dental hygiene class (graduating 2017) was identified as the sample for the DSR evaluation study. This sample represented the entire population of senior dental hygiene students enrolled at the University of Alberta during the time of the experiment. They were the first group of students to experience DSR for a history taking SCA, which has since become a regular component of the assessment. As part of a research project, the mock-SCA was a non-graded assessment, and students were not required to participate. Twenty minutes of class time was allocated for a presentation to the students on the research project, participant involvement (i.e. participating in a mock-SCA, submitting a reflection, and access to

their year-end SCA grades), the benefits of participation (e.g. an opportunity to practice for their high-stakes SCA in a low-stakes environment), the risks of participation (e.g. assessment related stress), and how confidentiality would be maintained. In an effort to improve confidentiality and reduce undue pressure to participate, the clinical course coordinator (with whom the students were very familiar) was kept at an arms-length from the actual study procedure, assisting in content development and validation, but unaware of which students agreed to participate or individual student results. The students were explicitly made aware that half of the participants would be randomly assigned to receive only a numerical percentage of their mock-SCA results, while the other half would receive a new form of online feedback. However, students were blinded to group assignment until they received their results, and precise details of the feedback were not provided to minimize student pre-conceptions, bias, and group contamination. Students were given an opportunity to ask questions, and then information letters and consent forms were distributed, see Appendix C. Students were asked to read the information letter at their own convenience, email the researchers if they had any remaining question or concerns, and return a completed consent form to the researcher if they chose to participate.

Experimental Procedure: The overarching timeline for each element of the evaluation study can be found in Figure 4.

Figure 4. Overview Timeline for Evaluation Project (Days)



The mock-SCA. The date of the mock-SCA was selected for two reasons. First, this date did not interfere with any other course-related obligations (such as the two-week distance clinical rotation students attend at different times throughout the semester), ensuring that all students were provided an equal opportunity to participate. Second, it positioned the mock-SCA within two-weeks of their year-end SCA, close enough so that any differences in the treatment and control group scores could be more readily attributed to the feedback intervention. The examiners of the mock-SCA were regularly scheduled clinical instructors, who were told about the purpose of the research in an information session prior to the mock-SCA, but were blinded to group allocation. Student confidentiality was explained to the instructors, and it was emphasized how performance in the mock-SCA should have no bearing on their future clinical experiences or standing in the course. Examiners were also told not to provide any additional feedback (i.e. verbal or written comments) within/following the mock-SCA so that DSR was the only feedback received. A researcher was onsite during administration of the mock-SCA to ensure adherence to

the research protocol. The day after the mock-SCA, half the students received DSR, and the other half received one overall percentage summarizing their performance.

The reflections. Five days following the assessment the students were prompted to reflect upon their mock-SCA performance. This date gave students a few days to review their results and/or feedback, while still being relatively soon after the experience so it would still be salient. Specifically, students were asked "what do you think you did well during the mock-SCA?" and "what do you think you could have improved upon in your mock-SCA performance?" The goal of these prompts was to encourage students to think about their strength and weaknesses, without providing too much guidance which might lead students to reflect in a manner different than their initial reactions (i.e. to facilitate authentic reflections). Students were provided fifteen minutes of class time to write their reflections in order to achieve a high response rate, and to control for the amount of time spent writing the reflections. By keeping the relative lengths of the reflections the same, the reflections differed predominantly by their quality and content.

The year-end SCA. The regularly scheduled year-end SCA took place twelve days following the mock-SCA and the results were collected for analysis. Following the year-end SCA all students received DSR.

The follow-up survey. In order to further validate the study results and DSR in general, a follow-up survey was developed to determine student opinions of, and interactions with, the feedback. This survey is presented in Appendix D. The first part of the survey was designed to determine if there was contamination between groups (i.e. to see if the intervention group shared their DSR output with students in the control group), and to determine whether the feedback was used after the mock-SCA to prepare for the year-end SCA. The final part of the survey was to assess all students' general perceptions and value of the feedback received following their year-

end SCA. The survey was given to the students several months after the year-end SCA in order to allow for preliminary data analysis and adjustment of survey questions, as needed, based on the initial findings. The survey was administered online through Google Forms, and students were provided ten minutes of class time to complete the survey to encourage a high response rate. An additional email was sent requesting completion of the survey by students who were not present during this designated class time (e.g. students who were sick or on clinical rotation). Students were provided a separate information sheet at the start of the survey, and were required to provide electronic consent to participate in the survey portion of the study. It was made clear to the students that completion of the survey was voluntary, but that the results could be used to improve the feedback received after future SCAs.

Data Coding and Analysis: Data analysis was conducted using STATA 14.[115]

Analysis of SCA results. Results of the SCAs were screened prior to analysis using descriptive statistics and a visual inspection of box plots and histograms to identify outliers and examine the distributions of the data. SCA results were summarized as percentages using means and standard deviations. Normality and equality of variances tests were conducted to identify violations to assumptions prior to parametric analyses. Differences in mock-SCA results between the control and intervention groups were analyzed using independent t-tests. The year-end SCA results were analyzed for group differences using linear regression controlling for the mock-SCA results, as is considered best practice in test-retest designs.[116] In cases of extreme violations of normality, equivalent non-parametric tests were used in place of parametric ones. In cases of violations of homogeneity of variance, pooled variances were used to calculate the test statistics.

Analysis of reflection quality. In order to analyze to the quality of the reflections, a grading rubric was developed, guided by the University of Alberta Heath Sciences Education and

Research Commons (HSERC) Interprofessional Reflection Guide[79] and a literature review of tools to assess reflections. The initial rubric described quality by three levels: descriptive, analytical, and practical implications. Through an iterative process, the rubric was adjusted and modified until it allowed two trained raters to evaluate the quality of the reflections reliably. Approximately 15% of the (anonymous) reflections were randomly selected for coding by both researchers. Based on the current rubric, researchers independently appraised the reflections, and then through a process of comparison, discussion, and reconciliation the rubric was adjusted for clarity and usability, and the process undertaken again with different reflections, until an acceptable level of inter-rater reliability was achieved. Due to the nature of the reflection prompts, it was difficult to categorize responses using the initial three-point scale, especially with the two questions yielding different types and tenses (i.e. past and present) of responses. It became apparent that two unique rubrics were required for evaluating each question using a dichotomized scale of low and high-quality reflections, see Appendix E. Each unique comment (delineated by a new a sentence, unless adjacent sentences reflected the same topic) within each question was graded either 0 (low level reflective statement) or 1 (high level reflective statement), by the two researchers. The two researchers then individually coded all the reflections based on this finalized rubric. Following the individual evaluations the exact inter-rater agreement was 84%, with a Cohen's Kappa of 0.64, indicating substantial agreement.[117] The remaining 16% of items were reviewed and coded as 0 or 1 through discussion and consensus. The reflection quality for each question, and in total, was then summarized using means and standard deviations. Preliminary screening of the histograms indicated potentially non-normal data distributions for the questions individually, however, the Shapiro-Wilk test of normality revealed no significant deviations from normal, $p > .05$. Due to the robustness of t-

tests to violations of normality, it was deemed appropriate to use independent t-tests to look for differences between the average reflection quality between the two groups within and across the reflection prompts.

Analysis of reflection content. The content of the reflections was analyzed through the process of content analysis.[110] The units of analysis were the diagnostic domains, with the domain descriptions, competencies, and test items within the assessment blueprint guiding the rater's domain selection. The basic content scoring rubric can be seen in Appendix F. Similar to the quality analysis, two researchers independently coded the reflective statements using the scoring rubric. Notably, the raters were allowed to code a comment as representing more than one domain where applicable. An "other" category was also provided to capture the number of reflective statements that did not fit within the diagnostic domains. Initial evaluations yielded an exact inter-rater agreement of 70%. With the large number of possible categories (sixteen in total when considering each selection of multiple domains as a unique category) the coefficient of agreement by chance was very low (0.18), resulting in another high Cohen's Kappa of 0.64. The remaining 30% of domain selections showing disagreement were reviewed and coded through discussion and consensus. The reflection content was summarized using frequencies and percentages of reflective statements within each diagnostic domain. The number of reflective statements within each domain was then analyzed using Poisson regression to look for differences between the two groups. A goodness of fit test was used to check for violations of the equidispersion assumption, a strict assumption that must be met when using a Poisson distribution.

Analysis of survey data. The follow-up survey data was analyzed primarily through descriptive statistics. A ten-point scale was used to assess student perceptions of the feedback,

instead of a traditional five-point Likert scale, as a method of forcing respondents to give non-neutral responses. Responses were then summarized as frequencies and percentages of those who disagreed and agreed with each statement, simplifying interpretation. The actual ten-point scale was still available for use in additional ordinal analyses investigating the relative strength of agreement/disagreement. Specifically, Spearman's correlations were used to see if survey responses were associated with other variables from the study such as year-end SCA performance, reflection quality, and program grade point average (GPA).

Secondary Outcomes

Piloting DSR for all students following the year-end dental hygiene SCA captured numerical data that could be used to provide additional statistical evidence validating DSR. While the goal of this study was to focus on the student-level outcomes, rather than the assessment's psychometric properties, this information can provide useful information for future researchers in this field and towards the validity argument. Cronbach's alpha was used as a measure of internal consistency for the SCA results.[105,106] Pearson's correlations were used to provide additional evidence towards the presence of the diagnostic domains.[109]

The design of the evaluation study also allowed for the effects of the mock-SCA itself (i.e. practice) to be investigated in addition to the primary objectives. Overall student improvements between the mock and year-end SCA provided insight into the effects a non-graded practice-SCA had on student performance and achievement. An additional measure of the relative effects of the mock-SCA was to compare the results of the previous cohort of dental hygiene students who completed the history taking SCA in 2015 (without receiving a mock-SCA) to the cohort involved in this study in 2016 (who received a mock-SCA). The goal of this secondary outcome was to explore the benefits of including a mock-SCA as part of the course

syllabus, where the findings would assist dental hygiene educators and curriculum developers. The within-groups changes between the mock and year-end SCA were analyzed using paired t-tests. An independent t-test was used to look for overall differences on year-end SCA scores between the 2015 and 2016 cohorts. Fisher's exact tests were used—as the preferred test of proportions for small sample sizes—to detect differences in the frequency of pass/fail rates between the two cohorts.

Chapter 4: Results

In this chapter, the study results are described reflecting the order of the methods section. The first section details the outcomes of applying the general DSR framework within the dental hygiene SCA, including a description of the assessment blueprint, the validation processes that were undertaken, and the output provided to the students. The second section reports the findings of the evaluation study. The results of the secondary outcomes are then reported. Finally, a condensed summary of the major findings of this research is provided.

Development of a Diagnostic Score Report for Dental Hygiene

The Assessment Blueprint: During the blueprinting process, a need for skills-based domains was identified. The Canadian dental hygiene entry-to-practice competencies that were chosen to guide the assessment blueprinting process are already arranged into specific roles (e.g. professional, clinical therapist)[95] that could theoretically be used as diagnostic domains. However, utilizing a student-centric approach to domain selection, it was determined that the competencies could be arranged differently to provide more specific information to the student for this particular assessment. Four competency-based skills necessary to complete the history taking SCA were revealed and labelled as: effective communication, client-centered care, eliciting essential information, and interpreting findings.

During review of the test item coverage of the domains, both effective communication, and, in particular, client-centered care were found to be underrepresented. As such, four additional test items were added to the grading checklist. The modified Delphi approach to validation led to further edits of the initial blueprint (see Appendix A). For the competency ratings, only two items received a median score of three or less, and two additional items received a score below two by at least one reviewer. Three of the test items received a median

score of three or below, with an additional five items receiving a score below two by at least one rater. The blueprint was adjusted based on a review of those items and the reviewer comments. One of the competencies and two of the test items were moved to a more appropriate diagnostic domain. The definition of one of the domains was expanded so that the test items fit more appropriately within that domain. The final assessment blueprint can be seen in Table 1. The blueprint includes domain names, definitions, relevant competencies, and the number of test items within each domain, including one example test item—for reasons of confidentiality the actual test items are not provided, but the true assessment blueprint did include all test items.

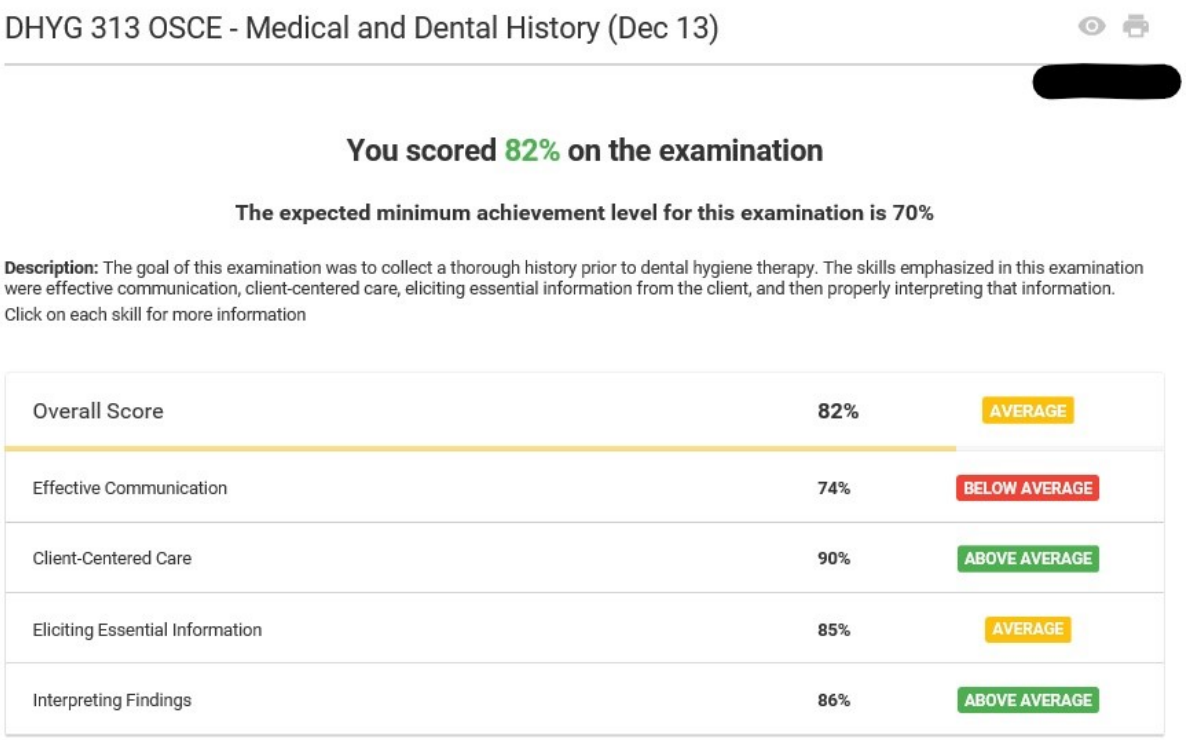
Table 1.

The Assessment Blueprint for a Dental Hygiene Structured Clinical Assessment

Domain	Definition	Competencies	Test Items
Effective Communication	This skill emphasizes *how* you are communicating, as opposed to *what* is being said. Effective communication strategies make the client feel safe and comfortable. It also means providing information in an appropriate manner the client can follow and understand.	• Use effective verbal, non-verbal, visual, written and electronic communication. • Demonstrate active listening and empathy to support client services. • Select communication approaches based on clients' characteristics, needs, and linguistic and health literacy level. • Facilitates confidentiality and informed decision-making in accordance with applicable legislation and code of ethics. • Convert findings in a manner relevant to clients using the principles of health literacy. • Manage time and other resources to enhance the quality of services provided. • Create an environment in which effective learning can take place.	Two checklist items and three rating scales Example: "Establishes rapport with client"
Client-Centered Care	This skill emphasizes how well you have incorporated the client in the care discussions and decisions. Client-centered care means respecting the needs, opinions, and autonomy of the client.	• Respect the autonomy of clients as full partners in decision-making. • Respect diversity in others to support culturally sensitive and safe services. • Design and implement services tailored to the unique needs of individuals. • Consider the views of clients about their values, health and decision-making. • Work with clients, to assess, diagnose, plan, implement and evaluate services for clients. • Prioritize clients' needs through a collaborative process with clients. • Negotiate mutually acceptable individual or program learning plans with clients. • Select educational interventions & develop educational materials to meet clients' learning needs.	Five test items and one rating scale Example: "Addresses the client's chief concern"
Eliciting Essential Information	This skill emphasizes what you are asking. Eliciting essential information from the client involves using appropriate prompts and follow-up questions to collect all necessary information prior to starting dental hygiene therapy. Comprehensive questioning is needed to identify status and risks of both oral and overall health.	• Collect accurate and complete data on the general, oral, and psychosocial health status of clients. • Elicit information about the clients' oral health knowledge, beliefs, attitudes and skills as part of the educational process. • Assess clients' need to learn specific information or skills to achieve, restore, and maintain oral health and promote overall wellbeing.	Ten checklist items Example: "Asks the client about current medications"
Interpreting Findings	This skill emphasizes how accurately you analyze the information revealed by the client. Proper interpretations include making appropriate modifications to the dental hygiene appointment, providing accurate recommendations, and identifying contraindications to care.	• Apply principles of risk reduction for client safety, health and wellbeing. • Evaluate clients' health and oral health status using determinants of health and risk assessment to make appropriate referral(s) to other health care professionals. • Apply theoretical frameworks to the analysis of information to support practice decisions. • Apply evidence-based decision-making to the analysis of information and current practices. • Apply the behavioural, biological and oral health sciences to dental hygiene practice decisions. • Identify clients for whom the initiation or continuation of treatment is contra-indicated based on the interpretation of health history and clinical data. • Identify clients at risk for medical emergencies. • Formulate a dental hygiene diagnosis using problem solving and decision-making skills to synthesize information. • Provide recommendations in regard to clients' ongoing care including referrals when indicated.	Eight checklist items Example: "Recognizes the need for a medical consult prior to dental hygiene treatment"

The Diagnostic Score Report: When a student logged in online to view their results, the first page detailed their overall score on the SCA and their scores on the four diagnostic domains, see Figure 5. Their overall score was compared to the course passing grade, set at 70%, which is also the benchmark used by the course coordinator to determine whether a one-on-one consultation with the student is required. To provide students with an understanding of their relative achievements, percentage results were also accompanied by cohort comparisons that described their performance as below average, average, or above average. This determination was based on a fixed mathematical formula which placed the "average" category within plus or minus one-half a standard deviation around the mean score. The rationale was to evenly distribute the ratings so that about one-third of the students would fall within each category (assuming a normal distribution of results). These cohort comparisons presented an opportunity to include meaningful colored components within the report and thus were colour coded red, yellow, and green respectively. To personalize the reports, the student's name appeared in the top right-hand corner of the opening page (removed from Figure 5 for confidentiality). Students were prompted to click on each of the diagnostic domains to receive more detailed information about each skill and thus how their scores were determined and how they could improve.

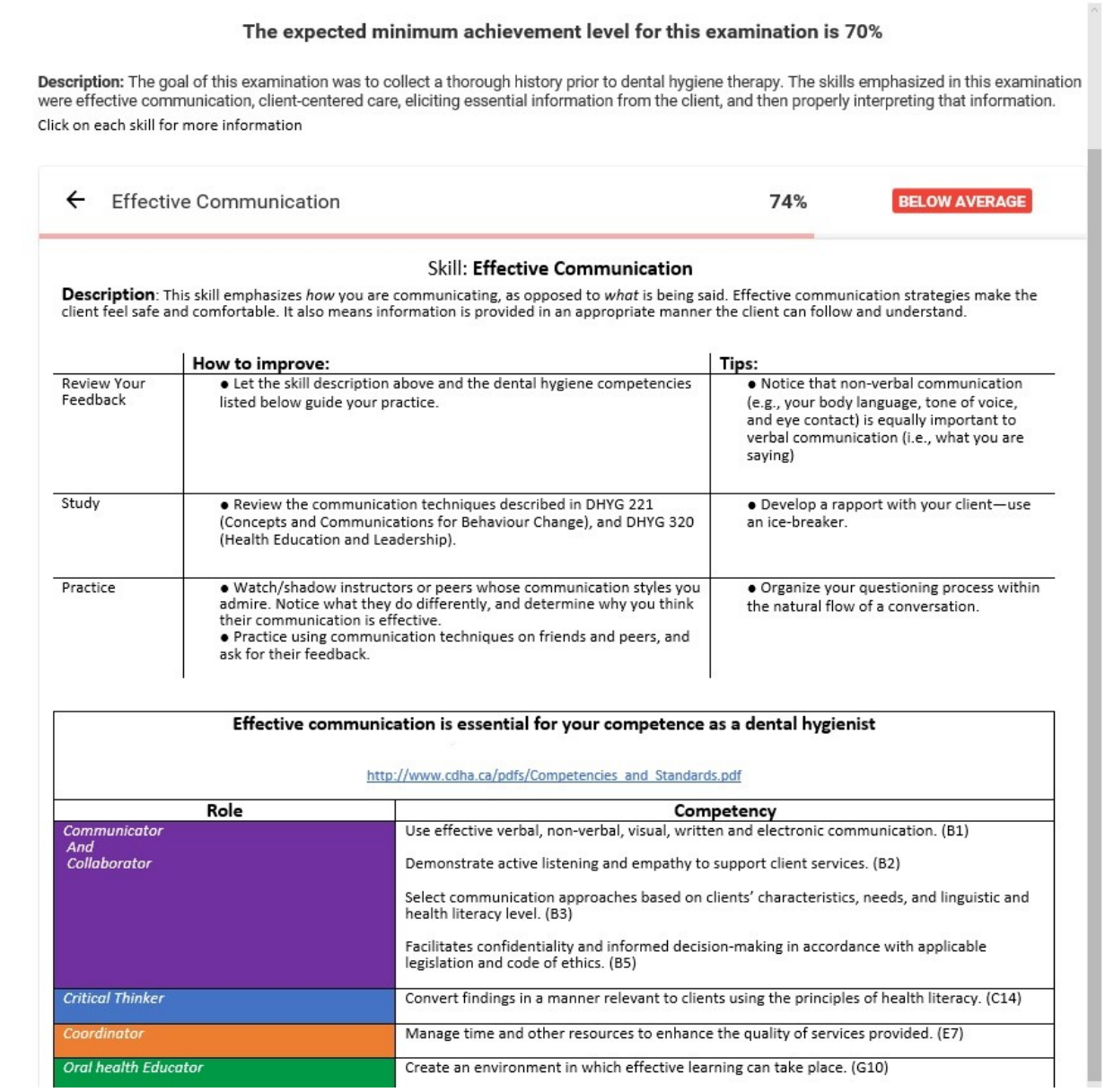
Figure 5. DSR Output Initial Page



Clicking on a domain displayed the diagnostic information for that skill including the definition, information on how to improve, and the relevant dental hygiene competencies. An example of domain-specific output, i.e. effective communication, can be seen in Figure 6. The improvement section was broken into three concise suggestions for reviewing the feedback, studying, and practicing—the goal was to provide useful information that could be easily read and absorbed without bombarding the students with too much text within the report itself. Specifically, students were encouraged to think critically about the skill descriptions and to use the dental hygiene competencies to better understand their professional roles and responsibilities. Study guidance provided references to texts and coursework they would need to review in order to improve within that skills-based domain. Practice advice was designed to encourage students to apply the feedback and skills outside of the classroom, into their clinical practice or other real-

world settings. Each improvement suggestion was accompanied by a “tip”, which focused the advice into more precise recommendations. In some cases these tips were specific to actual test items, and student uptake could directly improve their future SCA performances—although the student was not made aware of this fact, maintaining test confidentiality. In regard to aesthetics, clicking on each domain correspondingly changed the page heading, additional headings were bolded to draw focus, and white space was used to demarcate sections of the report. Information was also provided in a variety of ways including use of narratives, tables, and bullet points. A link to the Canadian Dental Hygienists Association webpage that lists the Canadian dental hygiene competencies was also provided,[95] and the specific roles of the dental hygienist were colour coordinated across each domain-specific report. The goal was to provide a visually attractive report that would encourage interaction by the students.

Figure 6. Diagnostic Score Reporting Output for a Diagnostic Domain



The expected minimum achievement level for this examination is 70%

Description: The goal of this examination was to collect a thorough history prior to dental hygiene therapy. The skills emphasized in this examination were effective communication, client-centered care, eliciting essential information from the client, and then properly interpreting that information.

Click on each skill for more information

← Effective Communication 74% BELOW AVERAGE

Skill: **Effective Communication**

Description: This skill emphasizes *how* you are communicating, as opposed to *what* is being said. Effective communication strategies make the client feel safe and comfortable. It also means information is provided in an appropriate manner the client can follow and understand.

	How to improve:	Tips:
Review Your Feedback	• Let the skill description above and the dental hygiene competencies listed below guide your practice.	• Notice that non-verbal communication (e.g., your body language, tone of voice, and eye contact) is equally important to verbal communication (i.e., what you are saying)
Study	• Review the communication techniques described in DHYG 221 (Concepts and Communications for Behaviour Change), and DHYG 320 (Health Education and Leadership).	• Develop a rapport with your client—use an ice-breaker.
Practice	• Watch/shadow instructors or peers whose communication styles you admire. Notice what they do differently, and determine why you think their communication is effective. • Practice using communication techniques on friends and peers, and ask for their feedback.	• Organize your questioning process within the natural flow of a conversation.

Effective communication is essential for your competence as a dental hygienist

http://www.cdha.ca/pdfs/Competencies_and_Standards.pdf

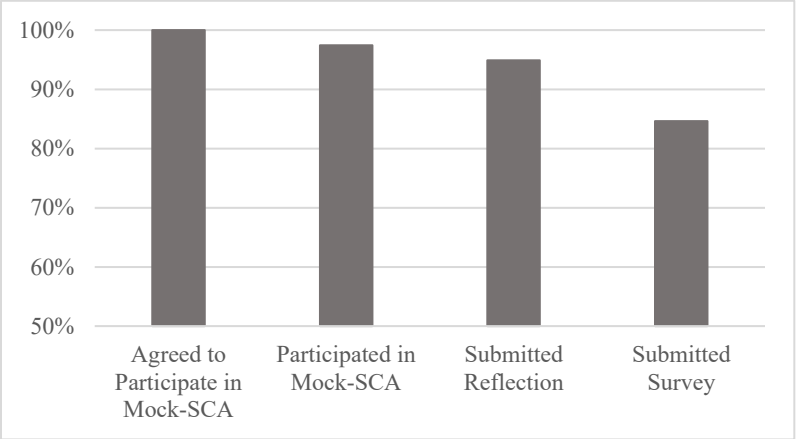
Role	Competency
Communicator And Collaborator	Use effective verbal, non-verbal, visual, written and electronic communication. (B1) Demonstrate active listening and empathy to support client services. (B2) Select communication approaches based on clients' characteristics, needs, and linguistic and health literacy level. (B3) Facilitates confidentiality and informed decision-making in accordance with applicable legislation and code of ethics. (B5)
Critical Thinker	Convert findings in a manner relevant to clients using the principles of health literacy. (C14)
Coordinator	Manage time and other resources to enhance the quality of services provided. (E7)
Oral health Educator	Create an environment in which effective learning can take place. (G10)

Evaluation of Diagnostic Score Reporting on Student Level-Outcomes

Thirty-nine students were in the dental hygiene graduating class of 2017. The vast majority of students were female (n = 38, 97%), with only one male student (3%). Student

participation rates for each stage of the evaluation project were very high, see Figure 7. One-hundred percent (n = 39) agreed to participate in the study, with 97% (n = 38) following through with participation in the mock-SCA, 95% (n = 37) submitting a written reflection on their mock-SCA performance, and 85% (n = 33) completing the follow-up survey. There is a noticeable trend of a steadily increasing attrition rate with every task requested of the students.

Figure 7. Student Participation Rates for Evaluation Study



The Mock-SCA: The results of the mock-SCA can be seen in Table 2. The average score on the examination was 75% (SD = 7.80). Ten students fell below the expected achievement level of 70%. Students scored best on the domain eliciting essential information, averaging 92% (SD = 9.63), with 19 students receiving a score of 100% (indicating a considerable ceiling effect). Students performed worst on the domain interpreting findings, averaging 42% (SD = 17.56), with a range of 0% to 71%. The assumption of normality was satisfied for all mock-SCA variables, except for the results under the domain eliciting essential information, W = .82, $z(38)$ = 4.03, $p < .001$. This assumption violation reflected the skew of the data and the ceiling effect, and as such, a non-parametric test (i.e. the Mann-Whitney U test) was chosen instead of an independent t-test to analyze the data. No significant differences were found between the control

group and the DSR group on mock-SCA totals, or for results by domain, $p > .05$. These findings indicate equality in SCA ability between the two experimental groups prior to feedback provision.

Table 2.

Mock-Structured Clinical Assessment Result: Mean % (SD)

	Combined ($n = 38$)	Control Group ($n = 20$)	DSR Group ($n = 18$)
Total	74.92 (7.80)	73.67 (7.69)	76.30 (7.89)
Effective Communication	77.05 (12.06)	75.28 (13.55)	79.01 (10.18)
Client-Centered Care	76.32 (10.13)	76 (10.01)	76.67 (10.54)
Eliciting Essential Information	92.11 (9.63)	91.67 (9.45)	92.59 (10.08)
Interpreting Findings	42.11 (17.56)	37.86 (10.65)	46.83 (22.35)

The Effect of DSR on Reflection: The average quality of the mock-SCA performance reflections can be seen in Table 3. Overall, the students showed somewhat low levels of reflection, with a total average score of 0.40 ($SD = .31$), which falls below the midpoint of the quality scoring rubric ranging from zero to one. The quality rating for identification of strengths was also below the midpoint, with an average score of 0.44 ($SD = .44$). The ratings for identification of areas requiring improvement were even lower, with an average of 0.36 ($SD = .44$). The independent t-tests revealed no significant differences between the DSR and the control group on reflection quality overall and within each question, $p > .05$.

Table 3.

Quality of Mock-Structured Clinical Assessment Reflection: Mean (SD)

	Combined ($n = 37$)	Control ($n = 20$)	DSR ($n = 17$)
Strengths	.44 (.44)	.5 (.43)	.37 (.45)
Improvements	.36 (.44)	.38 (.46)	.34 (.42)
Total	.40 (.31)	.44 (.34)	.36 (.27)

The content analysis of the reflections can be seen in Table 4. The reflection content fit well within the four diagnostic domains (n = 153, 93%), with only 7% (n = 7) of the comments allocated to the "other" category. Furthermore, there was no significant difference in the number of "other" comments between the two groups, p = .53, indicating that the four diagnostic domains encompassed the majority of student reflection content even without being prompted to think within those fours domains through the provision of DSR. Overall, the majority of comments focused on eliciting information (n = 54, 33%) and communication (n = 47, 29%), with fewer reflective statements on client-centered care (n = 30, 18%) and interpreting findings (n = 22, 13%). Due to the inherent differences in the two reflective prompts, results were analyzed for group differences separately by each question.

Table 4.

Reflection Content: Frequency (%)

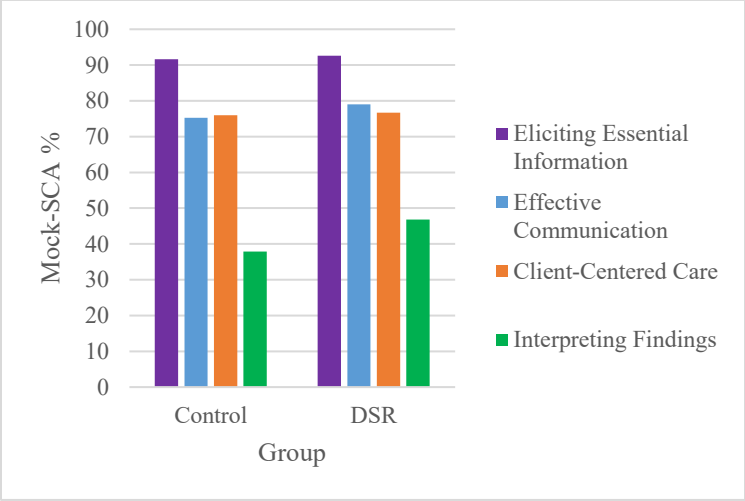
		Effective Communication	Client-Centered Care	Eliciting Essential Information	Interpreting Findings	Other
Strengths	Combined	24 (28%)	15 (17%)	35 (40%)	7 (8%)	6 (7%)
	Control	11 (22%)	8 (16%)	21 (40%)	6 (12%)	4 (8%)
	DSR	13 (35%)	7 (19%)	14 (38%)	1 (3%)	2 (5%)
Weaknesses	Combined	23 (30%)	15 (19%)	19 (25%)	15 (19%)	5 (6%)
	Control	15 (35%)	8 (19%)	15 (35%)	2 (5%)	3 (7%)
	DSR	8 (24%)	7 (21%)	4 (12%)	13 (38%)	2 (6%)
Total	Combined	47 (29%)	30 (18%)	54 (33%)	22 (13%)	11 (7%)
	Control	26 (28%)	16 (17%)	36 (39%)	8 (9%)	7 (8%)
	DSR	21 (30%)	14 (20%)	18 (25%)	14 (20%)	4 (6%)

*Percentages reflect row totals, and may not sum to 100% due to rounding.

Significant differences were detected regarding student identification of weaknesses and areas requiring improvement. For the domain interpreting findings, there was a difference in the

number of comments, $\beta = 2.03$, $z(37) = 2.68$, $p = .007$, with the DSR group providing significantly more comments than the control group. More specifically, taking the exponent of the beta coefficient gives us an incident rate ratio of 7.65, which indicates the DSR group had 7.65 times the average number of comments within the interpreting findings domain than the control group. In fact, the most common comments regarding weaknesses for the DSR group was within the domain interpreting findings ($n = 13$, 38%), while the control group reflected on interpreting findings the least ($n = 2$, 5%). There was also a difference in the number of comments for eliciting essential information, $\beta = -1.16$, $z(37) = -2.06$, $p = .04$, showing that the control group had significantly more comments within this domain. Specifically, an incident rate ratio of 0.31 can be calculated, indicating the DSR group had 0.31 times the average number of reflective statements regarding interpreting findings (i.e. 70% fewer). The descriptive statistics show that the control group reflected the most upon eliciting essential information ($n = 15$, 35%) as their area of weakness (tied with effective communication), while the DSR group reflected on eliciting essential information the least as a weakness ($n = 4$, 12%). Figure 8 shows that the students within both groups performed the best on eliciting essential information, averaging above 90%, and performed worst on the domain interpreting findings averaging below 50%. Therefore, it appears the students within the DSR group reflected more accurately upon their weaknesses.

Figure 8. Mock-Structured Clinical Assessment Percentages by Domain



In regard to identification of strengths, there were no significant differences between the groups, $p > .05$, but there were trends to further support the claim of more accurate self-assessments in the DSR group. The DSR group was more likely to reflect upon eliciting essential information as a strength ($n = 13$, 35%) compared to the control group ($n = 11$, 22%) and only one student in the DSR group reflected upon interpreting findings as a strength (3%) compared to six students in the control group (12%). As such, DSR seemed to result in more accurate identification of both strengths and weaknesses.

The Effect of DSR on Performance: Results of the year-end SCA can be seen in Table 5. All scores increased to some degree compared to the mock-SCA, with similar trends on domain scores. Students performed best on eliciting essential information ($M = 92.91$, $SD = 8.27$), with 17 students receiving a score of 100%, and worst on interpreting findings ($M = 48.36$, $SD = 25.36$), ranging from 0% to 100%. Overall differences between groups on the year-end SCA and within each domain were analyzed using regression, controlling for the mock-SCA scores. There were no significant differences between the DSR and control group on overall year-end SCA or

client-centered care scores, $p > .05$. There was a non-significant trend showing students in the DSR group performing better than the control group on the domain interpreting findings by an average of 10% ($SD = 51.35$) after controlling for the mock-SCA scores, $\beta = 10.19$, $t(35) = 1.21$, $p = .24$. There was a borderline significant difference on communication scores, $\beta = -6.58$, $t(35) = -2.04$, $p = .05$, which indicated that the DSR group actually performed poorer on communication than the control group ($M = -6.58$, $SD = 19.61$). The assumption of normality was violated again for the domain eliciting essential information, $W = .86$, $z(38) = 3.52$, $p < .001$, and so a non-parametric test of the differences between year-end and mock-SCA scores was used to look for differences between groups, which yielded no significant difference, $p > .05$. Overall, there was limited evidence to support that DSR improved student performance on their year-end SCA.

Table 5.

Year-End Structured Clinical Assessment Results: Mean % (SD)

	Combined ($n = 38$)	Control Group ($n = 20$)	DSR Group ($n = 18$)
Total	82.60 (7.05)	82.96 (6.60)	82.20 (7.69)
Effective Communication	85.32 (10.18)	88.16 (10.23)	82.16 (9.40)
Client-Centered Care	83.16 (10.16)	83.00 (10.81)	83.33 (9.70)
Eliciting Essential Information	92.91 (8.27)	92.31 (8.65)	93.59 (8.02)
Interpreting Findings	48.36 (25.36)	45.00 (26.72)	52.08 (25.36)

Survey Results and Use of Feedback: Of the 33 students (85%) who responded to the follow-up survey, 16 were in the control and 17 were in the DSR group. In regard to the DSR group, 100% ($n = 17$) of the students reported reviewing their feedback following the mock-SCA. Seventy-one percent ($n = 12$) reported reviewing the feedback briefly, while 29% ($n = 5$) reviewed it more thoroughly. Seventy-six percent ($n = 13$) also reported using the feedback to help them prepare for their year-end SCA either a little bit ($n = 10$, 59%) or a lot ($n = 3$, 18%). Therefore, the implementation of the intervention appeared to be successful, as the students in

the DSR group reported interacting to some extent with their feedback prior to the year-end SCA.

Alternatively, there appeared to be considerable contamination between groups, with 63% (n = 10) of students in the control group reporting that another student shared their DSR results with them. However, only four of those students reported using that feedback to help them prepare for the year-end SCA, potentially limiting the effect of the contamination on the actual results. It is expected that the contamination would have a larger impact on the year-end SCA results than the reflections results due to the minimal time to share the data between the mock-SCA and the reflections, and the increased likelihood that as the year-end SCA approached, the students would grow more eager to use/see their classmates' results to help them prepare. Notably, reordering groups to account for self-reported contamination did not result in any significant differences on year-end SCA scores (including eliminating the borderline significant group difference in communication scores). All the students in the control group who responded to the survey wished they had received more feedback following the mock-SCA (n = 16), indicating the perceived importance of feedback for student learning.

Student Perceptions of DSR: Student impressions of the DSR were consistently evenly split, with one group showing appreciation and the other group feeling more critical. Table 6 shows the summary of survey responses. Only 48% (n =16) of the students found the feedback generally helpful, and similarly 52% (n = 17) said the feedback helped them identify strengths and weaknesses, and would help them improve their performance in clinic. Fifty-eight percent (n = 19) of students said they understood how to use their feedback to make improvements. Correlations between the questions were all very strong and highly significant ($p < .001$) ranging from 0.76 to 0.89, indicating that positive responses to one question were related to positive

responses to the others. These findings suggest that rather than the feedback itself lacking in one area (i.e. not helpful, not applicable to performance, or difficult to use by the students), the value of DSR was directly related to differences in personal qualities between two groups of students who generally appreciated or did not appreciate the feedback. To further investigate this supposition, correlations were conducted between survey responses and variables that may have affected perceptions of the feedback. Dental hygiene program GPA was self-reported at the start of the survey averaging 3.45 (SD = .40), and ranging from 2 to 4. GPA did not significantly correlate to survey responses, $p > .05$. Neither were there significant correlations between year-end SCA results or reflection quality and the survey responses, $p > .05$. These findings indicate other variables not measured within this study are more strongly related to student perceptions of DSR.

Table 6.

Student Perceptions of Diagnostic Score Reporting

Question	No	Yes	Mean(SD)	Median
The feedback was helpful	52% (17)	48% (16)	5.33(2.48)	5
The feedback helped me identify strengths and weaknesses	48% (16)	52% (17)	5.27(2.95)	6
The feedback will help me improve my performance in clinic	48% (16)	52% (17)	5.27(2.97)	6
I understand how to use my feedback to improve	42% (14)	58% (19)	5.82(2.57)	6

*Based on a 10-point scale, No/Yes reflects ≤ 5 or ≥ 6 respectively

Eleven (33%) of the students who responded to the survey provided an optional written comment. Of those comments the majority (n = 7, 64%) explained that providing more specific feedback through DSR would better meet their learning needs. Four of the comments (36%) showed support for DSR, for example one student wrote “I found it very helpful and it made me

realize the areas I was succeeding and [needing] more work on." Equally, four (36%) of the comments expressed negative feelings of DSR, for example one student exclaimed "I felt that I didn't understand what I did wrong and couldn't use this information to improve for the [SCA]." (Note there is some overlap between supportive/negative comments and wanting increased specificity, and two comments were not actually directed towards DSR). These comments suggest improvements to the specificity of DSR are still needed, but again, relative perceptions towards DSR vary from student to student.

Notably, while 33 students replied to the follow-up survey regarding general perceptions of DSR, electronic records show that only 30/38 (79%) actually logged in to the online system to review their results following the year-end SCA. Alternatively, 100% (n = 38) logged in to see their feedback or overall score following the mock-SCA. These findings indicate that more students will engage with feedback following a formative SCA (i.e. the mock-SCA) compared to a summative SCA (i.e. the year-end SCA), however, the large majority will still engage to some extent.

Secondary Outcomes

Statistical Reliability of Diagnostic Domains: The Pearson's correlations for the mock and year-end SCA domain scores can be seen in Table 7. The strength of the correlations for the mock-SCA ranged between 0.04 and 0.59, with only one correlation achieving statistical significance. There was a significant positive correlation between scores on client-centered care and effective communication, $r = .59$, $p < .001$. For the year-end SCA, strength of correlations ranged between 0.07 and 0.40, with only one correlation achieving significance. This time there was a significant positive correlation between the scores on client-centered care and interpreting findings, $r = .40$, $p = .01$, however this relationship was only low-moderate in its strength. The

remaining correlations were weak and insignificant. These low correlations provide further evidence supporting the presence of these underlying domains within the dental hygiene history taking examination, and suggest that reporting scores within these domains provides unique information to the student. The exception is the domain of client-centered care, which may be confounded with other domains. However, with only a low-moderate correlation, its inclusion in the history taking SCA would likely still provide some additional information to the students.

Table 7.

Correlations between Domains (N = 38)

Mock-SCA

Year-End SCA

	Effective Communication	Client-Centered Care	Eliciting Essential Information	Interpreting Findings
Effective Communication	___	.59*	.26	-.07
Client-Centered Care	.28	___	.16	.04
Eliciting Essential Information	.12	-.10	___	.04
Interpreting Findings	-.11	.40*	-.07	___

* $p < .05$

Internal consistency was measured using Cronbach's alpha. For the year-end SCA, Cronbach's alpha was 0.63 indicating moderate internal consistency, however the domain scores ranged from 0.36-0.47 indicating unacceptable levels. However, this determination should be used with caution. The mock-SCA achieved a Cronbach's alpha of 0.74, using similar test items to the year-end SCA, and this substantial difference in internal consistency can be attributed to the decreased variance within the test items for the year-end SCA as students improved their performance. Similarly, the domain effective communication obtained a Cronbach's alpha of

0.70 for the mock-SCA using almost identical test-items as the year-end SCA which scored only 0.47. The general improvement of the students across all domains, and the ceiling effects of many of the test items led to reduced score variability and thus lower measures of internal consistency.[118] Due to some of the inherent limitations, Cronbach's alpha has been included in this study for completeness, but only as a secondary outcome. For the purposes of this research, the mock and year-end SCA showed adequate overall internal consistency for their use in experimentation.

Practice and Performance: Overall, students significantly improved performance from the mock to year-end SCA by an average of 8% (SD = 8.28), $t(37) = 5.72$, $p < .001$. Looking at domain scores, students significantly improved on communication (M = 8.27, SD = 14.69), $t(37) = 3.47$, $p = .001$, and client-centered care (M = 6.84, SD = 10.65), $t(37) = 3.96$, $p < .001$. Students did not significantly improve on eliciting essential information or interpreting findings, $p > .05$. The large ceiling effect within the domain eliciting essential information could explain why students did not show overall improvements, as there was little room left to improve. For interpreting findings, there was a non-significant trend of improvement by 6% (SD = 33.44), $p = .26$, and the maximum score achieved within this domain increased from 71% to 100%. However, the lack of statistically significant improvements could relate to differences in the content between the mock and year-end SCA (i.e. the findings students had to interpret were very different, even though they both required similar skills such as identifying contraindications to care). Therefore, it could be inferred that a practice SCA improves some skills through simple repetition (such as organization and communicating effectively), but will not necessarily improve skills that may require further study (such as interpreting findings).

Comparing the 2016 year-end SCA results to the 2015 results revealed only a non-significant trend of the 2016 class scoring higher on the SCA by 3% (SD = 17.42), $t(73.88)$ = 1.33, p = .19. There was, however, a significant difference between the number of students who fell below the expected achievement level of 70% for the SCA, with 19% (n = 8) of the students in 2015, compared to only 3% (n =1) of students in 2016 falling below expectations, p = .03. These findings suggest that a practice SCA may help more students achieve the minimum level of competence, but does not encourage students who have already met the marker for competence to achieve beyond that level of ability. It is important to acknowledge that several confounders such as time-effects and the personal/academic qualities of these two cohorts were not controlled for in this comparison, and therefore inductive claims are limited. However, the overall results for the effects of practice on performance suggest another educational tool is likely needed to promote students towards additional study and the goal of excellence.

Summary of Major Findings

The general framework for DSR was successfully implemented for a dental hygiene history taking SCA. Four diagnostic domains were identified as: effective communication, client-centered care, eliciting essential information, and interpreting findings. An assessment blueprint was established linking test items to professional competencies and their respective domains, which was validated through expert consensus. Additional statistical evidence supporting the presence of these skill-based domains include the low correlations between domain scores for the mock and year-end SCAs. In regard to student-level outcomes, DSR appeared to have an effect on student's ability to accurately reflect upon their performance, and self-assess their strengths and weaknesses. Specifically, the DSR group reflected significantly more on interpreting findings as an area of weakness, on which students did perform poorly,

while the control group reflected significantly more on eliciting essential information as a weakness, on which students actually performed very well. However, this accuracy in determining areas that required improvement did not directly translate into improved performance. Regarding student perceptions, half of the students found this feedback generally useful, applicable to their clinical practice, and easy to use, while the other half were critical of DSR. The most common suggestion to enhance DSR was to provide more specific information on how to improve performance. As such, DSR presents a promising start for improving student-level outcomes, however modifications and additions are still required if DSR is to be effectively utilized by all students to improve performance.

Chapter 5: Discussion

SCAs are a common assessment method for medical and allied health education, measuring whether major competencies are being met by students in a standardized, fair, yet realistic manner.[11,15,40] These assessments capture detailed information on several important clinical skills, however the opportunity to provide this information to students as feedback tends to be neglected.[10] Feedback has been described as a key pedagogical tool for facilitating reflection, learning, and overall professional development.[6,7,9] As such, there has been a push for increased and improved feedback provision in health professional education.[8,9] DSR presents a promising framework for providing feedback to all students following an SCA. DSR describes test results by performance on the underlying domains of learning the test intends to measure, and includes resources for making individual-level improvements.[23-26] While DSR has potential as a feedback mechanism for SCAs, it has largely been described and studied in the context of large-scale national assessments.[23,25,27] Furthermore, studies on DSR have focused on report interpretability and usability, [23,25,27] with little research on student-level impact. Applying DSR within SCAs, and evaluating DSRs effect on student learning are needed areas of investigation.

As such, the purpose of this research project was to establish and apply a general framework for providing DSR within the context of an SCA, and to evaluate the effect of DSR on student-learning outcomes. This discussion chapter is organized into four sections. It begins with a summary of the findings as they address the two research questions including their implications in the larger context. Next, the limitations of the research study are discussed. The following section on future directions describes how DSR could be improved based on the study findings and includes recommendations for future research within this topic area. Finally, the conclusion will summarize the major contributions of this research.

Implications of Results

Research Question 1: *To what extent can the framework of DSR be applied to provide all students with valid and high-quality feedback following an SCA?*

Score report validity. The adapted framework for applying DSR within SCAs was developed following an extensive literature review, and was used to create a course-specific score report for a dental hygiene year-end history taking SCA. The Canadian dental hygiene competencies were chosen to guide domain development.[95] A competency-based model for domain selection helps to validate that the assessment measures important practitioner skills, and ensures that students receive information not only on their performance for this one assessment, but towards development of those larger overarching essential skills.[101] Four skills-based domains necessary for successful completion of the SCA were identified as: 1) effective communication; 2) client-centered care; 3) eliciting essential information; and 4) interpreting findings. An assessment blueprint was established that showed the links between these domains and their operational definitions, relevant competencies, and test items. This blueprint provided evidence for how student performance on the SCA accurately represented student competence.[17] The modified Delphi technique for incorporating multiple expert opinions led to important iterative changes that further increased the accuracy of the blueprint. A review of the test item coverage of the domains also encouraged improvement of the blueprint, and the assessment itself, where new items were developed to more accurately assess the skills the test intended to capture. The rigorous assessment blueprinting process undertaken presents a main source of evidence for the validity of the DSR structure.

After piloting DSR, statistical information was obtained to further validate the assessment blueprint. It has been suggested that domain scores provide useful information above and beyond

what a single overall score provides when: the sub-scores are reliable, the test has low-reliability, the sub-scores are distinct from each other.[87] The correlations conducted between the different domain scores revealed generally small insignificant relationships, which supports reporting by the distinct domains. Cronbach's alpha revealed only moderate reliability of the SCA overall, but inadequate levels of reliability for the domains. According to the statistical suggestions for sub-score utility, the dental hygiene score report met two of the three criteria. In regard to the poor domain internal consistency, the issue may be one of test development rather than domain selection. Many test developers suggest items with minimal variance are of no use in an assessment as they are poorly discriminating.[119] Alternatively, a competency-based model of test development would suggest that any item that reflects achievement of a professional competency is worth measuring, even if reliability is compromised to some extent.[101] The statistical findings of this study showed that the domain eliciting essential information obtained the lowest Cronbach's alpha for both the mock and year-end SCA, which can be attributed to the fact that students scored extraordinarily high on this part of the test (averaging over 90%), with limited total variance. The content of this portion of the SCA directly reflects the required questions a dental hygienist must ask at every appointment in order to gather an accurate picture of health prior to treatment. Therefore, removal or restructuring of these questions for the purpose of the SCA would not be appropriate. Another option for improving internal consistency would be to increase the number of test items within each domain.[89] Increasing the length of the checklist would lead to longer scenarios, and may increase the likelihood of examiner mistakes, such as missing a student's completion of an item. In contrast to increasing the number of checklist items, domains could be merged together. However, this option would provide even less specific information to the student, when they are already asking for more. Therefore, addressing the

issue of low internal consistency is a challenging one that may not always serve the interests of a student-centric feedback framework. Despite the domains' poor internal consistency, there is other evidence to support the use of skills-based diagnostic domains in the year-end SCA. The blueprint was designed using a literature-based methodology, further validated through expert opinions and some supporting statistical evidence. Overall, the qualitative and quantitative strategies undertaken suggest the dental hygiene diagnostic score report is based on valid diagnostic domains that provide utility to the student.

Report quality. In addition to developing a valid assessment blueprint, certain guidelines for providing quality feedback and for effective score reporting were also adhered to in the development and administration of the dental hygiene score report. For example, the report was provided to students with information on how to improve their performance using clear and non-judgmental language, appropriate for the audience level, in timely and aesthetically pleasing manner.[1,7,8,23,25-27,69,88] However, the feedback survey provided to the students following their year-end SCA indicated that only half of the students thought that DSR was helpful, allowed them to better identify strengths and weaknesses, and could be used to improve their performance in clinic. Perhaps DSRs biggest weakness was its limited specificity. It has been well established that quality feedback must be task-specific.[1,2,7-9,60,67,70] While providing students with information on their performance by diagnostic domains is more specific than providing vague overall scores,[25] it still may not have been specific enough to meet the students' needs. The balance between specificity and important characteristics of DSR such as test confidentiality and domain reliability is a difficult one. However, there are still several options for making the reports more specific, such as increasing the personalization and individualization, and by providing more detailed learning resources to the student (see section on future directions).

Incorporating such components into DSR may help improve student perceptions of the feedback's quality.

Several sources of evidence for (and some against) DSR's quality and validity have been revealed through its development and administrative process. However, perhaps the most important source of evidence towards both claims is DSR's actual effect on student-level outcomes. This key evaluative element is described in the section addressing the second research question.

Institutional benefits of DSR. Several difficulties providing feedback following an SCA from the administrator's perspective have been described in the literature, and DSR has the potential to solve many of these issues. Item confidentiality is often a major concern for high-stakes SCAs, and so feedback is often restricted following these assessments.[85] DSR presents a feedback framework that can reveal as little (or as much) information on the actual test items as required to maintain test security.[25] The piloting of DSR in the dental hygiene program demonstrated how score reports could be generated without revealing any of the actual test items, while still providing details and resources that could be useful to the students.

Another issue providing quality feedback for SCAs are time limitations, both in administrative hours and in effective timely provision to students. The act of preparing and giving feedback can be a time consuming endeavor,[8,20] and thus it is typically restricted to students who need to undergo remediation.[10,20] This study showed how DSR requires only a moderate initial time investment prior to the assessment, and can then be used to provide feedback to *all* students following (multiple) SCAs. By using an iterative approach to DSR development, faculty and staff can begin by developing a basic DSR assessment blueprint, which can then be modified, and expanded upon, year after year, to improve and maximize the

feedback—thus continuing to require only short time investments across administrations. Online score reporting further increases the efficiency of DSR,[23,25-27] and with the progress of electronic SCA administration techniques,[49,55-57] online DSR is becoming easier to facilitate. This study demonstrated how reports could be provided to students within 24 hours (and theoretically instantly) following completion of their assessment. It should be noted, however, that DSR could also be distributed through non-automated processes such as manually-populated electronic forms (e.g. using Microsoft Excel and Word) or even paper-based formats, and that lack of an online administration mechanism does not preclude its use. While the latter means of administration certainly increase the administrative time demands to some extent, the majority of the work for DSR is done before the assessment and thus reports can still be returned to students in a relatively timely manner. Research suggests the perceived effectiveness of timely versus instant feedback is relatively equivalent.[120] Therefore, online DSR is not the only—although it might be the most ideal—administration method. Overall, DSR presents one additional way of increasing the amount of valid and quality feedback students receive within their health education. It is hoped that the comprehensive yet simplified framework for implementing DSR within SCAs produced as part of this thesis will encourage DSRs uptake by SCA developers/administrators across the health disciplines.

Research Question 2: *How will online DSR impact student-level outcomes such as reflective capabilities and clinical performance*?

Improved reflection. Through a process of experimentation, the effect of DSR on student reflection was investigated. The results indicate that DSR had no impact on the quality of student's reflections, but did effect the reflection content. Specifically, students were significantly more likely to accurately identify areas of weakness if they were provided DSR

following a mock-SCA. There was also trending evidence that DSR helped students more accurately identify their strengths. As such, it was concluded that DSR helped students better self-assess their history taking strengths and weaknesses. Reflection is considered a key component of learning.[19,61,73] Through the process of experiencing, reflecting, thinking (conceptualizing), and acting,[61] each SCA provides an opportunity to learn and improve future clinical performances. Improved self-assessment aids in this learning cycle by helping students to notice when their performance needs improvement (encouraging learning to begin),[19,64] and by potentially speeding up the process by helping students to focus their attention on key issues (eliminating potential uncertainty or misdirection).[1] Correctly identifying problem areas is also one important step towards reflection-in-action, a life-long skill essential for healthcare professionals.[19] Steps should be taken throughout higher education to encourage accurate reflection, in an effort to internalize these abilities for carryover into professional practice.[121] Thus DSR appears to be encouraging student development in a positive direction.

Improving performance. Ultimately, the goal of both feedback and reflection is to improve performance.[2,3,9,19] Despite the encouraging finding of improved self-assessment of strengths and weaknesses, this study did not find that this knowledge translated into improved performance. Dental hygiene students who received DSR following a mock-SCA showed no significant improvements, relative to the control group, on a similarly designed year-end SCA less than two weeks later. These findings suggest a breakdown of the experiential learning cycle between the stages of reflection and action. There are several potential reasons for why DSR did not result in improved clinical performance. These reasons relate to the feedback process, the feedback content, and the students.

In regard to the process, this study used only a short timeframe for investigating the effects of DSR. It is also possible that DSR may have an impact when used consistently over the long-term. The review by Veloski et al.[4] on feedback and physician performance showed that studies which found a significant impact of feedback tended to be of longer duration than studies that did not find significant effects. Similarly, the Ivers et al.[5] review on audit and feedback also found feedback was more effective if it was provided more than once. Thus, further investigation into the effect of frequency and duration on DSR's impact is still needed.

This study chose to use reflection as a measure of the effect of DSR (i.e. to see if feedback improved reflection). Alternatively, some researchers describe how the reciprocal process may also be beneficial: where feedback is made more effective by actively encouraging students to reflect upon that feedback.[71,74] Quinton and Smallbone[71] devised a structured method of encouraging reflection on feedback using certain literature-based prompts. Their prompts included asking students about their feelings towards the feedback, and only after this emotional venting process, asking students to think about and make a plan for applying their feedback. Other research supports the need for providing students with a way to deal with the emotional impact of feedback.[8,68] Jackson and Marks[74] investigated the effect of mandatory reflection on feedback, and found there was a significant improvement in overall grades after introducing this reflective component. Incorporating the effects of structured reflection on DSR may better reveal the full potential of DSR as a means to improve student performance.

Finally, in regards to process, there is some emphasis in the literature that feedback needs to be a two-way process.[3,8,68] While some research has indicated that written and verbal feedback can be equally effective,[4] with many students preferring written feedback,[67,69] written feedback (such as a score report) has also been chastised as a one-way process that does not give students

an opportunity to respond.[68] Therefore, it could be argued that students were unable to translate their feedback into improved performance because they were unable to follow-up on the feedback for clarification and further details. Incorporating a process where students can respond to the information provided by DSR could improve its efficacy.

Another reason that DSR may not have improved performance could be related to the feedback itself. It has already been mentioned above that DSR may not have been specific enough, which may have mitigated its impact on student performance. Efforts to improve the task-specific nature of DSR will likely have a positive effect on student learning. In addition, there are some reporting characteristics of DSR that should also be considered. The research conducted by Trout and Hyde[27] on effective score reporting suggested that, while colour was an appreciated aesthetic characteristic, the colour red was perceived as distressing. For the reports piloted within the dental hygiene program, a below average score was colour-coded red. While the intent was to draw focus to this domain as an area requiring attention, it may have had the opposite consequence, evoking an emotional response that actually paralyzed students from taking action.[8,67] In fact, one student received a score report where all domain scores were below average, and undoubtedly this entirely red report could have been quite alarming. As such, the students who may have benefited the most from the information provided within each diagnostic domain, may instead have chosen to ignore the feedback to uphold their self-esteem.[8,67] A similar issue is the use of norm-referenced comparisons themselves. The DSR literature encourages cohort comparisons so relative performance can be easily assessed,[23,25] and the Veloski et al.[4] review suggested that the use of cohort versus standards comparisons did not affect feedback's impact. However, there is other literature that is less supportive of this feature.[3,67,121] Shute[67] describes how norm-referenced comparisons can make poorly performing students feel less

capable than their peers, and thus powerless to improve. Shute[67] suggests using self-referenced comparisons, where a student is judged only by their own relative capabilities (e.g. highlighting which domains that one student performed best and worst on). Others suggest using competency or criterion-referenced comparisons (e.g. how well a student did on each domain relative to program expectations).[3] It is possible that certain reporting characteristics chosen for the dental hygiene score report may have unintentionally mitigated some of effectiveness of DSR.

Lastly, no matter how well a feedback source is designed, its impact will be directly influenced by the student's intent to use (or not to use) that feedback. Research indicates that students' perceptions and value of assessments directly impacts their performance on those assessments.[122] Similarly, the perceived value of an assessment will likely impact their interaction with their results (i.e. their score reports).[88] Additionally, students' general value of feedback will also affect their engagement with score reports.[21] While these three concepts may seem rather obvious, they directly impact the effectiveness of DSR on student achievement, and are important to consider. It is essential that students are guided not only to pass their SCAs, but to understand the assessment's purpose and importance. Likewise, the intent and value of the feedback provided should also be emphasized. If students understand that DSR is being provided for their own personal benefit and growth, and not just as another supplementary evaluation method, they may be more encouraged interact with their feedback. Instilling these fundamental beliefs may improve the impact of DSR on student outcomes.

Finally, there are psychological characteristics that may have played a role in DSR's student impact. In order for improvements to occur, students need to self-regulate, i.e. they need to become accountable for their own learning both cognitively and behaviourally.[72] Self-regulated learning is directly related to motivational processes such as confidence, self-esteem,

self-efficacy, and intrinsic interest.[72,121,123] Many of the reflective statements provided by the dental hygiene students indicated a lack of confidence in their SCA/history taking abilities. Therefore, self-efficacy may have been an issue. Moreover, the provision of grades can interfere with a student's motivations and self-esteem, distracting the student from the actual content of their feedback.[60,121] Comparison of the mock-SCA and year-end SCA results, showed that the students who improved the most were those who performed below the expected achievement level of 70% (averaging a 16% increase compared to a 5% increase). Although ceiling effects and feedback quality may have played a role in this finding, it may also suggest an external motivation to pass the assessment, rather than an internal motivation to improve their skills. Further research in this area is indicated to substantiate such a claim. Self-regulated learning is an essential skill for health professionals, which must be both taught to and demanded of students.[124] However, self-regulation is also a complex multifactorial system relating to students' beliefs and experiences, in conjunction with the learning environment.[124] Strategies for improving self-regulated learning extend beyond the context of applying DSR in SCAs, but are still important to discuss in some generalities. Certain strategies include goal setting to improve self-efficacy,[72] practicing self-monitoring (i.e. being aware of one's own thoughts and behaviours),[72,124] focusing on smaller details when self-assessing,[124,125] and promoting self-directed learning activities in educational currciula.[124] Tactics specific to the influence of grades on motivation include: providing feedback without the provision of grades, having students review their feedback prior to the receipt of grades, or allowing students opportunities to improve their grades after receiving their feedback.[121] Finally, a general suggestion particularly relevant to the research conducted within this thesis, is to provide quality feedback that "encourages positive motivational beliefs and self-esteem."[121(p205)] Self-regulation is a growing

area of research, showing interesting interactions with feedback and the learning process. Future research on DSR may need to incorporate measures of self-regulated learning strategies and psychological characteristics to understand exactly how this multifaceted concept may affect DSR's impact on student outcomes.

Study Limitations

There are some limitations to the evaluation study conducted as part of this thesis project that may limit the findings. This study focused on the dental hygiene student population, and thus results may not be generalizable to other allied-health professions. Very little research has been conducted on DSR in SCAs and student-level outcomes in general, making it difficult to compare the study findings to results with different samples. Similar studies to this one should be conducted using different student populations and different SCAs to confirm and elaborate on the findings. Additionally, the available sample for this study was rather small, with only 39 students comprising the only senior dental hygiene class in the province. The small sample size may not have presented enough power to understand the full significance of DSR on learning. Investigations with larger sample sizes would provide more power and allow for more elaborate statistical analyses to be conducted. There were also some limitations to the study design. The reflection quality was difficult to assess due to variations in tense and form, requiring the original scoring rubric to be modified multiple times to achieve acceptable inter-rater reliability. For example, some students wrote in full sentences, while others wrote in point form, restricting interpretations of the intent and context of many of the statements. Therefore, this study may not have captured the true quality of the student's internal reflections. Replication of this study using more carefully structured reflection prompts and writing instructions, or even other qualitative techniques such as structured interviews, may help to validate the finding of equivalently low

levels of reflections across both experimental groups. Furthermore, the follow-up survey revealed contamination between the intervention and the control groups, where the score reports were shared between classmates—although this also indicates a positive interest in the feedback. While students in the contaminated control group were not provided their personal results by domain, they were still provided with the details on how to improve within each domain. As such, the contamination may have impacted the relative effect of DSR on performance. Conducting research within the context of academic learning is often difficult, as research methods, samples, and timelines are restricted by the program and its curriculum. However, despite these limitations this study was conducted with rigour to the extent feasible, and does provide insight into an important and neglected area of research.

Future Directions

Next Steps: DSR should be an iterative process,[25-27,87,88] and the findings from this study can inform future improvements. First, a few modifications to the current dental hygiene score report may be necessary. Due to the potentially negative impact of norm-referenced comparisons,[67] a switch to standards-referenced comparison may be more appropriate. By providing students with their scores by standards, students may be more motivated to improve, as the steps remains stationary (in contrast to cohort comparisons, where if everyone improves a person may still remain below average) and thus reaching the next level may appear more achievable. In order to accurately present these comparisons, a standard setting exercise would need to take place for the SCA and its domains. Standard setting is an organized and structured approach to determining the cut-scores for certain levels of achievement on an assessment.[126] For example, instead of presenting scores as below average, average, above average, specific cut-scores could be determined for below standard, standard, above standard. Several approaches to

determining these cut-score exist, but a common method is to have qualified experts predict the probability that a minimally competent student would get each test-item correct (known as the Angoff method).[126] The average of those probabilities then becomes the cut-score for achieving the minimum standard (typically referred to as the pass/fail mark). A similar approach could be undertaken to determine a cut score for exceptionality, or however many cut-scores are deemed appropriate for the assessment. On a similar topic, the use of red colouring for the "below" rating may be seen as threatening,[27] and perhaps should be changed to something less emotionally provoking such as orange. Managing emotional responses to feedback is an important endeavor, and these small changes may prove beneficial.

The dental hygiene score report may also require modification to the diagnostic domain client-centered care. For the mock-SCA, client-centered care significantly and strongly correlated with effective communication, and for the year-end SCA was significantly and moderately correlated to interpreting findings. Conversely, the rest of the domains showed very small and insignificant correlations. These low correlations validate presenting scores to students by separate domains, illustrating that these scores do in fact represent different skills, and thus provide additional information to the test-taker.[87,109] Therefore, the significant correlations found for client-centered care may indicate that this domain is not entirely unique to the other domains, and does not provide as much information. It may be prudent to redistribute the test items within client-centered care, as appropriate, to the other domains (e.g. for the year-end SCA a majority would likely fit under interpreting findings based on the moderate correlation). This redistribution of test items has the added benefit of increasing the number of items within each domain, which would likely help to increase each domains' reliability. Alternatively, since client-centered care was revealed as an important competency-based skill during the blueprinting

process, determining new and more unique test items to represent this domain may also prove beneficial. With only a moderate correlation with one domain for the year-end SCA, leaving client-centered care as is for future DSR would still provide *some* additional information to the students. However, improving the statistical properties of the assessment blueprint will also improve the validity argument, and should be a long-term goal.

The score report could also be expanded upon in several ways. These suggestions are also generally applicable to other educators looking to implement and maximize DSR. The literature indicates the need for a two-way feedback process, where students can respond to the feedback and follow-up for clarification and assistance.[3,8,68] While the typical suggestion for this two-way process is to provide feedback face-to-face,[68] the improvements in technology and the advancement of online DSR means that written feedback does not necessarily preclude a two-way process. Adding online discussion boards and messaging systems would allow students to follow-up on their score reports and ask for further guidance from their instructors. Discussion could also be encouraged amongst the students, where peers could share insights and study strategies. In fact, encouraging peers to practice assessing and providing feedback with each other is considered a useful strategy in helping students better understand the intent and processes of using feedback in general.[1] Furthermore, research on feedback seeking behaviours indicates that students are more likely to ask for feedback from computerized sources than directly from a person.[127] Since self-esteem can impact whether students will seek out feedback,[128] online messaging has the added benefit of allowing for anonymity. Students will be able to request additional feedback and clarification, without fearing how this request will make them appear to their instructor or peers. However, there are still some instances where a face-to-face follow-up may be most appropriate. For example, Kluger and Adler[127] found that students

with low self-esteem performed better when provided with computer-based feedback, while students with high self-esteem performed better after person-mediated feedback. As such, following DSR, instructors should also provide students with an optional opportunity to meet and discuss their results in person. Providing multiple avenues for a two-way feedback process following an SCA may aid in the impact of DSR on performance.

Increasing the specificity of DSR is another area requiring improvement. There are several different ways this goal may be achieved. The most obvious and simplistic option is to provide the students with the actual test items and their corresponding results. As mentioned, for a high-stakes SCA, where test confidentiality is a chief concern, this option may not be feasible. However, even when actual test items cannot be revealed, providing *example* test items presents a feasible option that may still aid in the interpretation of results.[25,27] Another option is to improve the quality of, and access to, learning resources provided to students through online DSR. In this study, reports only mentioned the references students could use to improve performance. Providing actual links that would take students directly to important sections of textbooks or other online documents may prove beneficial.[23,27] Links to video-based feedback is an additional avenue for providing more detailed feedback to students. Videos could take the form of exemplars, where trained professionals demonstrate important interactive skills.[1,129] Students could also be recorded during their SCA, and their own video performance provided for self-review.[129,130] A meta-analysis of self-video feedback and professional communication skills, found a significantly positive effect.[130] Providing both exemplars and personal videos would allow students to compare and evaluate their performance against an ideal target. The textual components of DSR would provide the students with key information on what they should be focusing on during their video reviews. Example test items, links to online resources, and videos,

all present viable methods for increasing the specificity of the resources provided by DSR for making improvements.

Finally, an important step for DSR is to improve the degree of personalization and individualization.[23,26,27,67] These processes also relate to providing students with feedback that is more specific to their personal needs. To facilitate individualized score reports, different reporting components could be displayed based on different SCA checklist response patterns. For example, if a student did not complete the task "palpate the submental nodes", only then would the student receive a section on his or her score report providing information on that skill. Further personalization could be incorporated by generating the student's name directly into the improvement suggestions.[23] Through computer programming techniques, certain sentence components could be tagged to certain SCA outcomes, and sentences like the following could be instantly generated for personalized specific feedback using online DSR: "Jenny, it looks like you forgot to palpate the submental nodes, click here for guidelines on how to complete a full head and neck examination." This individualization could go even further by generating different reporting components based on different levels of achievement. Trout and Hyde[27] indicate low achievers might benefit most from information on how to interpret their scores and make improvements, while advanced students might benefit from additional learning activities beyond the scope of the assessment. Low achievement on an SCA might also reflect a perceived lack of importance of the assessment itself.[122] Therefore, score reports for students who perform poorly on an SCA could provide greater details on the assessment's purpose, and how it prepares them for professional practice. At its greatest level of individualization, students would undergo a complete psychological and learning style assessment to determine characteristics such as levels of intrinsic versus extrinsic motivation, self-esteem, self-efficacy, and so on, and score reports

would be altered to provide details in the manner most effective to each student. It is clear that online DSR has much potential in the area of individualization that should continue to be explored and studied. Improving the specificity and personalization of these reports will likely be a major factor towards encouraging student-level improvements.

Future Research: In addition to some of the general suggestions for future research (e.g. replication, different populations, larger sample sizes, and longer duration) other specific research needs have been identified throughout this discussion. Each of the different modifications and expansions described in the previous section would require comparative studies so that their relative impacts can be determined. For example, how does the use of cohort vs normative comparisons, or video vs. non-video learning resources, effect student outcomes? Comparing different processes of DSR administration would also be beneficial, for example, seeing if mandatory reflection on DSR improves its impact. Furthermore, adding qualitative components within these studies would provide more detailed insight into *why* exactly certain factors have such an effect. These findings would ultimately help to guide effective DSR.

Another important area of exploration within feedback provision is the complex person-feedback interaction. Within the dental hygiene student sample used for this research, half the students responded positively to DSR, while the other half did not. This perception did not appear to be related to GPA or SCA scores. Qualitative interviews or focus groups with these two populations might better explain this split perception, allowing DSR to be improved, or, perhaps, better personalized, depending on student characteristics. Incorporating measures of confounding variables and effect modifiers such as motivation, self-efficacy, confidence, and feedback and assessment perceptions into DSR related studies would help to better identify key personal influences on DSR's effectiveness. The dental hygiene student population in particular

could benefit from increased exploration within this area, as a majority of the current health education literature on feedback, and related concepts such as self-regulated learning or reflection-in-action, take place within the field of medical education. To maximize DSR's effect, there needs to be more personalization to meet the specific needs of different students. This goal cannot be accomplished until more detailed research is conducted defining those needs. As students begin to experience DSR more regularly in their curricula, more and more data can be collected. With larger datasets, certain data reduction techniques can be used (e.g. factor or cluster analysis) to look for different response patterns in SCA performance. These response patterns could indicate different groups of students with different performance deficits and unique learning needs. Online DSR could then be personalized to provide different information to students based on these scoring patterns. Increased individualization is the next big step in DSR, requiring more data collection and research.

Conclusion

Online DSR presents a promising framework for providing valid and timely feedback to all students following an SCA. This study demonstrated how online DSR following a dental hygiene SCA resulted in more accurate reflections, particularly in identifying skills that required improvement. However, this improved self-assessment did not translate into improved performance. Therefore, while the DSR system described in this research presents an encouraging starting point for providing quality feedback, further enhancements are required. Suggestions for improving the feedback in order to help facilitate behaviour change include: providing links to relevant references, incorporating video feedback, and developing more personalized/individualized reports. Further research is also required to better understand the psychological and motivational factors that influence DSRs impact on students.

References

1. Sadler DR. Beyond feedback: Developing student capability in complex appraisal. *Assessment and Evaluation in Higher Education*. 2010;35(5):535-50.
2. van de Ridder JMM, Stokking KM, McGaghie WC, ten Cate OTJ. What is feedback in clinical education? *Medical Education*. 2008;42(2):189-97.
3. Boud D, Molloy E. *What is the problem with feedback?* In: Boud D, Molloy E, editors. Feedback in higher and professional education: Understanding it and doing it well. New York, NY: Routledge; 2013. p. 1-10.
4. Veloski J, Boex JR, Grasberger MJ, Evans A, Wolfson DB. Systematic review of the literature on assessment, feedback and physicians' clinical performance: BEME Guide No. 7. *Medical teacher*. 2006;28(2):117-28.
5. Ivers N, Jamtvedt G, Flottorp S, Young JM, Odgaard-Jensen J, French SD, et al. Audit and feedback: effects on professional practice and healthcare outcomes. *The Cochrane Database of Systematic Reviews*. 2012(6).
6. Bangert-Drowns RL, Kulik C-LC, Kulik JA, Morgan M. The instructional effect of feedback in test-like events. *Review of Educational Research*. 1991(2):213-38.
7. Ende J. Feedback in clinical medical education. *The Journal of the American Medical Association*. 1983;250(6):777-81.
8. Archer JC. State of the science in health professional education: Effective feedback. *Medical education*. 2010;44(1):101-8.
9. Branch WT, Paranjape A. Feedback and reflection: Teaching methods for clinical settings. *Academic Medicine*. 2002;77(12):1185-8.

10. Harrison CJ, Molyneux AJ, Blackwell S, Wass VJ. How we give personalised audio feedback after summative OSCEs. *Medical teacher*. 2015;37(4):323-6.

11. Miller GE. The assessment of clinical skills/competence/performance. *Academic Medicine*. 1990;65(9):63-7.

12. Epstein RM. Assessment in medical education. *New England Journal of Medicine*. 2007;356(4):387-96.

13. Boursicot K, Etheridge L, Setna Z, Sturrock A, Ker J, Smee S, et al. Performance in assessment: Consensus statement and recommendations from the Ottawa conference. *Medical teacher*. 2011;33(5):370-83.

14. Harden R. What is an OSCE? *Medical teacher*. 1988;10(1):19-22.

15. Harden RM, Stevenson M, Downie WW, Wilson GM. Assessment of clinical competence using objective structured examination. *British Medical Journal*. 1975;1:447-51.

16. Harden RM, Gleeson FA. Assessment of clinical competence using an objective structured clinical examination (OSCE). *Medical Education*. 1979;13(1):39-54.

17. Newble D. Techniques for measuring clinical competence: objective structured clinical examinations. *Medical education*. 2004;38(2):199-203.

18. Barrows HS. An overview of the uses of standardized patients for teaching and evaluating clinical skills. AAMC. *Academic Medicine*. 1993;68(6):443-51.

19. Sandars J. The use of reflection in medical education: AMEE Guide No. 44. *Medical Teacher*. 2009;31(8):685-95.

20. White CB, Ross PT, Gruppen LD. Remediating students' failed OSCE performances at one school: The effects of self-assessment, reflection, and feedback. *Academic Medicine*. 2009;84(5):651-4.

21. Harrison CJ, Könings KD, Molyneux A, Schuwirth LW, Wass V, van der Vleuten CP. Web-based feedback after summative assessment: how do students engage? *Medical education*. 2013;47(7):734-44.

22. Wormald BW, Schoeman S, Somasunderam A, Penn M. Assessment drives learning: An unavoidable truth? *Anatomical Sciences Education*. 2009;2(5):199-204.

23. Goodman DP, Hambleton RK. Student test score reports and interpretive guides: Review of current practices and suggestions for future research. *Applied Measurement in Education*. 2004;17(2):145-220.

24. Roberts MR, Gierl MJ. Developing score reports for cognitive diagnostic assessments. *Educational Measurement: Issues and Practice*. 2010;29(3):25-38.

25. Zenisky AL, Hambleton RK. Developing test score reports that work: The process and best practices for effective communication. *Educational Measurement: Issues and Practice*. 2012;31(2):21-6.

26. Roberts MR, Gierl MJ, editors. Development of a framework for diagnostic score reporting. *Annual meeting of the American Educational Research Association,* San Diego, CA; 2009.

27. Trout DL, Hyde E, editors. Developing score reports for statewide assessments that are valued and used: Feedback from K-12 stakeholders. *Annual Meeting of the American Educational Research Association*, San Francisco, CA; 2006.

28. Pugh D, Hamstra SJ, Wood TJ, Humphrey-Murto S, Touchie C, Yudkowsky R, et al. A procedural skills OSCE: Assessing technical and non-technical skills of internal medicine residents. *Advances in Health Sciences Education*. 2015;20(1):85-100.

29. Larsen T, Jeppe-Jensen D. The introduction and perception of an OSCE with an element of self-and peer-assessment. *European Journal of Dental Education*. 2008;12(1):2-7.

30. Setyonugroho W, Kennedy KM, Kropmans TJ. Reliability and validity of OSCE checklists used to assess the communication skills of undergraduate medical students: A systematic review. *Patient education and counseling*. 2015;98(12):1482-91.

31. Sim JH, Abdul Aziz YF, Vijayanantha A, Mansor A, Vadivelu J, Hassan H. A Closer Look at Checklist Scoring and Global Rating for Four OSCE Stations: Do the Scores Correlate Well? *Education in Medicine Journal*. 2015;7(2).

32. Scheffer S, Muehlinghaus I, Froehmel A, Ortwein H. Assessing students' communication skills: Validation of a global rating. *Advances in health sciences education*. 2008;13(5):583-92.

33. Sloan DA, Donnelly MB, Schwartz RW, Strodel WE. The Objective Structured Clinical Examination. The new gold standard for evaluating postgraduate clinical performance. *Annals of surgery*. 1995;222(6):735-42.

34. Sloan DA, Donnelly MB, Schwartz RW, Felts JL, Blue AV, Strodel WE. The use of the objective structured clinical examination (OSCE) for evaluation and instruction in graduate medical education. *Journal of Surgical Research*. 1996;63(1):225-30.

35. Newble DI. Assessing clinical competence at the undergraduate level. *Medical Education*. 1992;26(6):503-11.

36. Patrício MF, Julião M, Fareleira F, Carneiro AV. Is the OSCE a feasible tool to assess competencies in undergraduate medical education? *Medical teacher*. 2013;35(6):503-14.

37. Barman A. Critiques on the objective structured clinical examination. *Annals Academy of Medicine Singapore*. 2005;34(8):478-82.

38. McWilliam P, Botwinski C. Developing a successful nursing objective structured clinical examination. *Journal of Nursing Education*. 2010;49(1):36-41.

39. Mitchell ML, Henderson A, Groves M, Dalton M, Nulty D. The objective structured clinical examination (OSCE): optimising its value in the undergraduate nursing curriculum. *Nurse Education Today*. 2009;29(4):398-404.

40. Rushforth HE. Objective structured clinical examination (OSCE): review of literature and implications for nursing education. *Nurse education today*. 2007;27(5):481-90.

41. Selim AA, Ramadan FH, El-Gueneidy MM, Gaafer MM. Using Objective Structured Clinical Examination (OSCE) in undergraduate psychiatric nursing education: Is it reliable and valid? *Nurse education today*. 2012;32:283-8.

42. Manogue M, Brown G. Developing and implementing an OSCE in dentistry. *European Journal of Dental Education*. 1998;2(2):51-7.

43. Brown G, Manogue M, Martin M. The validity and reliability of an OSCE in dentistry. *European Journal of Dental Education*. 1999;3(3):117-25.

44. Hanioka T, Ojima M, Kawaguchi Y, Hirata Y, Ogawa H, Hinode D, et al. Education on tobacco use interventions for undergraduate dental students. *Japanese Dental Science Review*. 2015;51(3):65-74.

45. Näpänkangas R, Karaharju-Suvanto T, Pyörälä E, Harila V, Ollila P, Lähdesmäki R, et al. Can the results of the OSCE predict the results of clinical assessment in dental education? *European Journal of Dental Education*. 2016;20(1):3-8.

46. Schoonheim-Klein M, Habets L, Aartman I, Van der Vleuten C, Hoogstraten J, Van der Velden U. Implementing an objective structured clinical examination (OSCE) in dental education: effects on students' learning strategies. *European Journal of Dental Education*. 2006;10(4):226-35.

47. The National Dental Examining Board of Canada (NDEB). *OSCE* 2017 [cited 2017. Available from: https://ndeb-bned.ca/en/accredited/osce-examination.]

48. Navickis MA, Bray KK, Overman PR, Emmons M, Hessel RF, Cowman SE. Examining clinical assessment practices in US dental hygiene programs. *Journal of dental education*. 2010;74(3):297-310.

49. Snodgrass SJ, Ashby SE, Rivett DA, Russell T. Implementation of an electronic Objective Structured Clinical Exam for assessing practical skills in pre-professional physiotherapy and occupational therapy programs: Examiner and course coordinator perspectives. *Australasian Journal of Educational Technology*. 2014;30(2):152-66.

50. Gorman SL, Lazaro R, Fairchild J, Kennedy B. Development and implementation of an objective structured clinical examination (OSCE) in neuromuscular physical therapy. *Journal of Physical Therapy Education*. 2010;24(3):62-8.

51. Farahat E, Rice G, Daher N, Heine N, Schneider L, Connell B. Objective Structured Clinical Examination (OSCE) improves perceived readiness for clinical placement in nutrition and dietetic students. *Journal Of Allied Health*. 2015;44(4):208-14.

52. Blue C. Objective structured clinical exams (OSCE): A basis for evaluating dental hygiene students' interpersonal communication skills. *Access*. 2006;20(7):27-31.

53. Scalese RJ, Obeso VT, Issenberg SB. Simulation technology for skills training and competency assessment in medical education. *Journal of General Internal Medicine*. 2008;23(1):46-9.

54. Zartman RR, McWhorter AG, Seale NS, Boone WJ. Using OSCE-based evaluation: Curricular impact over time. *Journal of Dental Education*. 2002;66(12):1323-30.

55. Snodgrass SJ, Ashby SE, Onyango L, Russell T, Rivett DA. Electronic practical skills assessments in the health professions: a review. *Internet Journal of Allied Health Sciences and Practice*. 2014;12(1):8.

56. Meskell P, Burke E, Kropmans TJ, Byrne E, Setyonugroho W, Kennedy KM. Back to the future: An online OSCE management information system for nursing OSCEs. *Nurse education today*. 2015;35(11):1091-6.

57. Hochlehnert A, Schultz J-H, Möltner A, Tımbıl S, Brass K, Jünger J. Electronic acquisition of OSCE performance using tablets. *GMS Zeitschrift für Medizinische Ausbildung*. 2015;32(4).

58. Wass V, Van der Vleuten C, Shatzer J, Jones R. Assessment of clinical competence. *The Lancet*. 2001;357(9260):945-9.

59. Newble DI, Jaeger K. The effect of assessments and examinations on the learning of medical students. *Medical Education*. 1983;17(3):165-71.

60. Rust C. The impact of assessment on student learning how can the research literature practically help to inform the development of departmental assessment strategies and

learner-centred assessment practices? *Active learning in higher education*. 2002;3(2):145-58.

61. Kolb D. *The process of experiential learning*. In: Experiential Learning: Experience as the Source of Learning and Development. 2nd ed. Upper Sadle River, New Jersey: Pearson Education; 2015. p.20-38.

62. Molloy E, Boud D. *Changing conceptions of feedback.* In: Boud D, Molloy E, editors. Feedbak in higher and professional education: Understanding it and going it well. New York, NY: Routledge; 2013. p.11-33.

63. Kolb A, Kolb D. Experiential learning theory: A dynamic, holistic approach to management learning. *Journal of Education and Development*. 2008;17(9):312-7.

64. Torbert WR. *The interplay of feedback, attention, and conciousness*. In: Learning from experience: Toward consciousness. New York, New York: Columbia university press; 1972. p. 7-36.

65. Kluger AN, DeNisi A. The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *American Psychological Association*. 1996;119(2)254-84.

66. Weaver MR. Do students value feedback? Student perceptions of tutors' written responses. *Assessment & Evaluation in Higher Education*. 2006;31(3):379-94.

67. Shute VJ. Focus on formative feedback. *Review of Educational Research*. 2008;78(1):153-89.

68. Jolly B, Boud D. *Written feedback: What is it good for and how can we do it well?* In: Boud D, Molloy E, editors. Feedback in higher and professional education: Understanding it and going it well. New York, NY: Routledge; 2013. p.104-24.

69. Ferguson P. Student perceptions of quality feedback in teacher education. *Assessment and Evaluation in Higher Education*. 2011;36(1):51-62.

70. Hewson MG, Little ML. Giving Feedback in Medical Education. *Journal of General Internal Medicine*. 1998;13(2):111-6.

71. Quinton S, Smallbone T. Feeding forward: using feedback to promote student reflection and learning–a teaching model. *Innovations in Education and Teaching International*. 2010;47(1):125-35.

72. Zimmerman BJ. Self-regulated learning and academic achievement: An overview. *Educational Psychologist*. 1990;25(1):3-17.

73. Boud D, Keogh R, Walker D. *What is reflection in learning?* In: Boud D, Keogh R, Walker D, editors. Reflection: Turning experience into learning. London: Kogan Page; 1985. p. 7-17.

74. Jackson M, Marks L. Improving the effectiveness of feedback by use of assessed reflections and withholding of grades. *Assessment and Evaluation in Higher Education*. 2016;41(4):532-47.

75. Fund Z, Court D, Kramarski B. Construction and application of an evaluative tool to assess reflection in teacher-training courses. *Assessment and Evaluation in Higher Education*. 2002;27(6):485-99.

76. Hargreaves K. Reflection in medical education. *Journal of University Teaching and Learning Practice*. 2016;13(2).

77. Clark S, Duggins A. *What research guides our beliefs about professional learning?* Using quality feedback to guide professional learning: A framework for instructional leaders. Tousand Oaks, CA: Corwin; 2016. p.16-29.

78. El-Dib MAB. Levels of reflection in action research: An overview and an assessment tool. *Teaching and Teacher Education*. 2007;23(1):24-35.

79. Health Sciences Education and Research Commons. *Interprofessional Reflection Guide* University of Alberta: University of Alberta; 2016 [Available from: http://www.hserc.ualberta.ca/Resources/CurricularResources/InterprofessionalReflectionGuide.aspx.]

80. Black N, Harden R. Providing feedback to students on clinical skills by using the Objective Structured Clinical Examination. *Medical Education*. 1986;20(1):48-52.

81. Hodder RV, Rivington R, Calcutt L, Hart I. The effectiveness of immediate feedback during the objective structured clinical examination. *Medical education*. 1989;23(2):184-8.

82. Hollingsworth MA, Richards BF, Frye AW. Description of observer feedback in an objective structured clinical examination and effects on examinees. *Teaching and Learning in Medicine: An International Journal*. 1994;6(1):49-53.

83. Cushing A, Abbott S, Lothian D, Hall A, Westwood OM. Peer feedback as an aid to learning - What do we want? Feedback. When do we want it? Now! *Medical Teacher*. 2011;33(2):105-12.

84. Taylor CA, Green KE. OSCE Feedback: A randomized trial of effectiveness, cost-effectiveness and student satisfaction. *Creative Education*. 2013;4(6):9.

85. Gotzmann A, De Champlain A, Homayra F, Fotheringham A, de Vries I, Forgie M, et al. Cheating in OSCEs: The impact of simulated security breaches on OSCE performance. *Teaching and Learning in Medicine*. 2017;29(1):52-8.

86. Lane S, Raymond MR, Haladyna TM, Downing SM. *Test development process*. In: Lane S, Raymond MR, Haladyna TM, editors. Handbook of test development. 2nd ed. New York, NY: Routledge; 2015. p. 3-18.

87. Zenisky AL, Hambleton RK. *A model and good practices for score reporting*. In: Lane S, Raymond MR, Haladyna TM, editors. Handbook of test development. 2nd ed. New York, NY: Routledge; 2015. p. 585-602.

88. Zapata-Rivera JD, Katz IR. Keeping your audience in mind: applying audience analysis to the design of interactive score reports. *Assessment in Education: Principles, Policy & Practice*. 2014;21(4):442-63.

89. Sinharay S, Puhan G, Haberman SJ. Reporting diagnostic scores in educational testing: Temptations, pitfalls, and some solutions. *Multivariate Behavioral Research*. 2010;45(3):553-73.

90. Wise LL, Plake BS. *Test design and development following the standards for educational and psychological testing*. In: Lane S, Raymond MR, Haladyna TM, editors. Handbook of test development. 2nd ed. New York, NY: Routledge; 2015. p. 19-39.

91. Risconscente MM, Mislevy RJ, Corrigan S. *Evidence-Centered Design*. In: Lane S, Raymond MR, Haladyna TM, editors. Handbook of test development New York, NY: Routledge; 2015. p. 40-63.

92. Harden RM. Approaches to curriculum planning. *Medical Education*. 1986;20(5):458-66.

93. Harden RM. Ten questions to ask when planning a course or curriculum. *Medical Education*. 1986;20(4):356-65.

94. Harden R. AMEE Guide No. 21: Curriculum mapping: A tool for transparent and authentic teaching and learning. *Medical teacher*. 2001;23(2):123-37.

95. Canadian Dental Hygienists Association. *Entry-to-practice competencies and standards for Canadian dental hygienists*. 2010. [Available from: https://www.cdha.ca/pdfs/Competencies_and_Standards.pdf]

96. General Medical Council. *Tomorrow's doctors: Outcomes and standards for undergraduate medical education.* 2009.

97. American Dental Hygienists's Association (ADHA), American Dental Education Association (ADEA). *Core competencies for graduate dental hygiene education*. 2010.

98. Australian Dental Council. *Professional competencies of the newly qualified dentist. Melbourne*. 2016.

99. The College of Family Physicians of Canada. *CanMEDS-family medicine: Working group on currciulum review.* 2009.

100. The National Dental Examining Board of Canada (NDEB). *Competencies for a beginning dental practitioner in Canada.* 2005.

101. Harris P, Snell L, Talbot M, Harden RM. Competency-based medical education: implications for undergraduate programs. *Medical Teacher*. 2010;32(8):646-50.

102. Lefroy J, Gay S, Gibson S, Williams S, McKinley R. Development and face validation of an instrument to assess and improve clinical consultation skills. *International Journal of Clinical Skills*. 2012;5:115-25.

103. Kane M. *Validation strategies: Delineating and validating proposed interpretaions and uses of test scores*. In: Lane S, Raymond MR, Haladyna TM, editors. Handbook of test development. 2nd ed. New York, NY: Routledge; 2015. p. 64-80.

104. Kane MT. Current concerns in validity theory. *Journal of Educational Measurement*. 2001;38(4):319.

105. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. *The American journal of medicine*. 2006;119(2):166. e7-. e16.

106. Kane MT. An argument-based approach to validity. *Psychological bulletin*. 1992;112(3):527.

107. Cook DA, Brydges R, Ginsburg S, Hatala R. A contemporary approach to validity arguments: a practical guide to Kane's framework. *Medical Education*. 2015;49(6):560-75.

108. Hsu C-C, Sandford BA. The Delphi technique: making sense of consensus. *Practical assessment, research & evaluation*. 2007;12(10):1-8.

109. Mark RR. *Job analysis, practice analysis and the content of credentialing examinations*. In: Lane S, Raymond MR, Haladyna TM, editors. Handbook of test development. 2nd ed. New York, NY: Routledge; 2015. p. 144-64.

110. Hanson EC. *Analysing qualitative data*. Successful qualitative health research: A practical introduction. New York, NY: Open University Press; 2006. p. 137-60.

111. Pickett FA. *Personal, dental, and health histories*. In: Darby ML, Walsh MM, editors. Dental hygiene theory and practice. 3rd Ed ed. St. Louis, MO: Saunders Elsevier; 2010. p. 149-79.

112. Walsh MM. *Medical emergencies*. In: Darby ML, Walsh MM, editors. Dental hygiene theory and practice. 3rd Ed ed. St. Louis, MO: Saunders Elsevier; 2010. p. 100-26.

113. Sunell S, Richardson F, Udahl B, Jamieson L, Landry D. National competencies for dental hygiene entry-to-practice. *Canadian Journal of Dental Hygiene*. 2008;42(1).

114. Darby ML, Walsh MM. *Dental Hygiene Theory and Practice*. 3rd Ed ed. St. Louis, MO: Saunders Elsevier; 2010.

115. StataCorp. Stata Statistical Software: Release 14. College Station, TX: StataCorp; 2015.

116. Senn S. Change from baseline and analysis of covariance revisited. *Statistics in Medicine*. 2006;25(24):4334-44.

117. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-74.

118. Zieky MJ. *Developing fair tests*. In: Lane S, Raymond MR, Haladyna TM, editors. Handbook of test development. 2nd ed. New York, NY: Routledge; 2015. p. 81-99.

119. Haladyna TM. *Item analysis for selected-response test items*. In: Lane S, Raymond MR, Haladyna TM, editors. Handbook of test development. New York, NY: Routledge; 2015.

120. Bayerlein L. Students' feedback preferences: How do students react to timely and automatically generated assessment feedback? *Assessment and Evaluation in Higher Education*. 2014;39(8):916-31.

121. Nicol DJ, Macfarlane-Dick D. Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*. 2006;31(2):199-218.

122. Brown GTL, Hirschfeld GHF. Students' conceptions of assessment: Links to outcomes. *Assessment in Education: Principles, Policy and Practice*. 2008;15(1):3-17.

123. Deborah L. Butler, Philip H. Winne. Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*. 1995(3):245.

124. Brydges R, Butler D. A reflective analysis of medical education research on self-regulation in learning and practice. *Medical Education*. 2012;46(1):71-9.

125. Sandars J. Future direction for research in self-regulated learning in medical education. *Medical Education*. 2012;46(6):626.

126. Gregory JC, Darin SE. *Setting performance standards on tests*. In: Lane S, Raymond MR, Haladyna TM, editors. Handbook of test development. New York, NY: Routledge; 2015. p. 212-37.

127. Kluger AN, Adler S. Person-versus computer-mediated feedback (English). *Computers in human behavior*. 1993;9(1):1-16.

128. Fedor DB, Rensvold RB, Adams SM. An investigation of factors expected to affect feedback seeking: A longitudinal field study. *Personnel Psychology*. 1992;45(4):779-802.

129. Paul S, Dawson K, Lanphear J, Cheema M. Video recording feedback: A feasible and effective approach to teaching history-taking and physical examination skills in undergraduate paediatric medicine. *Medical education*. 1998;32(3):332-6.

130. Fukkink RG, Trienekens N, Kramer LJ. Video feedback in education and training: Putting learning in the picture. *Educational Psychology Review*. 2011;23(1):45-63.

Appendix A.

Example Validation Scale for Competencies Mapped to Diagnostic Domains

Please rate on a scale of 0 (no fit) to 4 (excellent fit) how well each dental hygiene competency fits in the skill: "**Effective Communication"**

Description: This skill emphasizes "*how*" you are communicating, as opposed to what exactly is being said. Effective communication strategies make the client feel safe and comfortable. It also involves providing information in an appropriate manner the client can follow and understand.

Test Item	No Fit	1	2	3	Excellent Fit
1. Use effective verbal, non-verbal, visual, written and electronic communication. (B1)					
2. Demonstrate active listening and empathy to support client services. (B2)					
3. Select communication approaches based on clients' characteristics, needs, and linguistic and health literacy level. (B3)					
4. Facilitates confidentiality and informed decision-making in accordance with applicable legislation and code of ethics. (B5)					
5. Convert findings in a manner relevant to clients using the principles of health literacy. (C14)					
6. Manage time and other resources to enhance the quality of services provided. (E7)					
7. Create an environment in which effective learning can take place. (G10)					

Please indicate if any of the competencies would fit better under a different skill:

Other skills:

Using a Client-Centered Approach to Care
Eliciting Essential Information
Interpreting Finding

If any competency received a rating of 2 or lower, please provide a comment:

Appendix B.

Instructions to Standardized Patient

D HYG 313 Mock-OSCE Medical History Interviewer

Personal History:

- You are Ms. Smith
- A school bus driver
- No children
- Late fifties/early sixties

Health History:

- You were diagnosed with coronary artery disease after you had a heart attack one year ago (aka. a myocardial infarction, treated with angioplasty).
    - *The student should ask about illnesses – coronary artery disease, and surgeries – heart attack, but only mention the heart attack was one year ago if asked "when".*
- You are now considered stable, but your doctor says you should always carry nitroglycerin (description: very small white tablet, used sublingually/placed under the tongue) with you in case of chest pain (aka. angina). You have brought nitroglycerin with you to this appointment.
    - *Provide information about nitroglycerin when asked about current medications or medications to treat your coronary artery disease. Only tell them you have nitroglycerin with you if explicitly asked.*
- Your doctor also recommended you take low-dose (aka. baby) aspirin (description: round yellow pill) daily (you take it in the morning after breakfast) because it can improve blood flow to the heart and help prevent another heart attack.
    - *Only mention aspirin if explicitly asked, or if asked about* ***over-the-counter*** *medications—i.e. don't mention aspirin as part of coronary artery disease treatment or current medications, many patients forget to mention over the counter medications when asked about "medications" in general.*
- Your doctor believes your heart attack was related to lifestyle factors such as smoking, unhealthy eating habits, and being sedentary. You have been actively changing your diet and taking long walks since your heart attack.
    - *Only mention the smoking if directly asked about smoking/tobacco use.*
- You smoke about two packs a day.
- You are not interested in quitting smoking at this time, but would be open to discussing it at another appointment.
    - *The student should offer tobacco cessation—e.g. are you interested in getting help to quit smoking?—and you will say "no" you are not interested. Afterwards they should ask if you would be willing to discuss it at a later appointment and you should say "yes". If the student continues to press you on quitting smoking after*

you have already said no, you may get frustrated, which should be reflected in your rating of the student.

- You do not drink red wine or coffee, do not use mouthwash, and you do not drink well water.
    - *Besides smoking, these are other possible reasons for teeth staining that the student may ask you about.*
- You are allergic to bananas and kiwi.
    - *These fruit allergies frequently coincide with a latex allergy.*
- You have never been tested for a latex allergy.
    - *If asked about a latex allergy, just say you aren't sure.*

Oral Health History:

- Your last dental/dental hygiene appointment was a little over a year ago (a few months before the heart attack).
- Your main reason for a dental hygiene appointment (i.e. your chief concern) is some brown staining on your teeth.
    - ***Must ask****: Sometime after you're asked about your chief concern/reason for coming in you must ask* ***"What do you think is causing my staining?"***
- Your current oral home care involves brushing twice daily (morning and night) and flossing about once a week.
- You have recently stopped using regular toothpaste and only use baking soda because you read on the internet that it will make your teeth whiter**.**
    - ***Please tell this to the student if asked about your current homecare techniques****.*
- You think your oral health is generally good, except for the staining.

Student interaction:

- The student is a 3rd year dental hygiene student at the University of Alberta dental clinic.
- You are a new patient to the clinic and are meeting the student for the first time. You started coming to this clinic because of the lower fees.
- The student must interview you to determine the status of your medical and dental health.
- You are patient, kind, and easygoing, but not forthcoming in your answers (only answering exactly what is being asked). You may get a little bit agitated if pressured to quit smoking.
- You may reply "no" or "not sure", as appropriate to any other questions with information not listed above.

Example:

Student: Have you had any recent illnesses or surgeries?
You: Yes, I was actually diagnosed with Coronary Artery Disease after having a heart attack.
Student: How long ago was the heart attack?
You: Last year.

Student: Did you have a mitral valve replacement?
You: No I didn't.
Students: What is your current cholesterol level?
You: I'm not sure.

Instructions for SP Ratings:

Following the OSCE the instructor will prompt you to rate the student on two scales, from 0-5 (poor to excellent) based on the following criteria:

1) *Demonstrated good communication skills* → Do you feel the student: 1) exhibited empathy, 2) acted professionally, 3) used appropriate non-verbal communication skills (eye contact and body language), 4) provided information clearly, and 5) was not patronizing?
    - 1 points for each yes.
2) *Demonstrated a patient-centred approach to the encounter* → Do you feel the student engaged with you in the discussion (e.g. talked with you, not talked at you), and respected your needs, opinions, and autonomy?
    - Minus 1 point for each instance where you felt disrespected or left out of the discussion.

Appendix C.

DSR Evaluation project information letter and consent form

INFORMATION LETTER

Evaluation of Online Feedback for a Dental Hygiene OSCE

Research Investigators:
Alix Clarke
MSc Student, School of Dentistry
alix1@ualberta.ca

Minn Yoon
Assistant Professor, School of Dentistry
minn.yoon@ualberta.ca

Background

You are being asked to participate in this study because you are part of the dental hygiene year for which online feedback is being piloted. Feedback is an essential learning tool and this study is intent on improving the type of feedback dental hygiene students receive after a high-stakes clinical examination.

Purpose

To evaluate a new form of online feedback.

Study Procedures

Your participation will be threefold. 1) We request your participation in a mock-OSCE. This assessment will be about 10 minutes long and does not require any preparation (although you are free to do so). It provides an opportunity to practice your clinical examination skills (similar to those used in your year-end OSCE), but will not affect your grade in any way. If you choose to participate you will be assigned to one of two groups. One group will receive feedback in a new online format, the other receive a standard numerical grade. *Group assignment will be random*. 2) Following the mock-OSCE a written reflection regarding your mock-OSCE performance will be solicited. This reflection may be as long as you feel is necessary to adequately describe your experience. 3) Your year-end OSCE results will be collected for comparison to the mock-OSCE.

Benefits

You will directly benefit from this study by being provided an opportunity to practice your examination skills for a high-stakes clinical examination, in a low-stakes environment. Participation in this study will also help us evaluate and improve feedback for the dental hygiene program.

Risk

There are no expected risks of this study. However, if you experience any discomfort you are free to discontinue participation at any time. If you have any concerns feel free to ask any questions.

Voluntary Participation

You are under no obligation to participate in this study. Your participation will have no bearing on your courses or grades, nor your standing in this program. You may discontinue participation any time, no questions asked. Even following participation you may withdraw your data from the study up until the last day of semester (before January 2017). If you would like to withdraw please contact a researcher above, you will need to provide your student ID number so your data can be identified and removed from the study.

Confidentiality & Anonymity

The clinic course coordinator (A. Sheppard) will not be aware of who has agreed to participate in this study, nor have access to individual results. Examiners of the mock-OSCE may be clinic instructors you are familiar with, but they are aware of the confidential nature of the results, and will not use the information in any way that might affect your progress in this program. The data from this study will be collected and stored confidentially using de-identified study ID numbers. All final results and publications will be completely anonymous, with only aggregate data described. Only the research team at the university will have access to the collected data. Hard copy documents will be reviewed then stored in a locked filing cabinet in a locked office at the University of Alberta. Data will be kept in a secure place for 5 years following completion of research project, as per required research protocol.

Further Information

At the end of the project, participants may request to receive a copy of the research findings report. Please send the request and any other questions you may have to a researcher listed above. The plan for this study has been reviewed for its adherence to ethical guidelines by the Research Ethics Board at the University of Alberta. For questions regarding participant rights and ethical conduct of research, contact the Research Ethics Office at (780) 492-2615.

CONSENT FORM for Participants

Title of Project: **Evaluation of Online Feedback for a Dental Hygiene OSCE**		
Part 1: Researcher Information		
Name of Principal Investigator: Alix Clarke, MSc Student Affiliation: School of Dentistry Address: 5-575 ECHA, 11405-87 Ave, Edmonton, AB, T6G 1C9 Contact Information: alix1@ualberta.ca		
Part 2: Consent of Participant		
	Yes	**No**
Do you understand that you are being asked to be in a research study?		
Have you read and received a copy of the information letter?		
Do you understand the benefits and risks of taking part in this research study?		
Have you had an opportunity to ask questions and discuss the study?		
Do you understand that you do not have to participate in the study? You do not have to give a reason.		
Do you understand that you can change your mind and stop being in the study at any time? You do not have to give a reason.		
Do you understand who will have access to your study information?		
Has confidentiality in this study been explained to you?		
Do you understand you will be asked to participate and reflect on a non-graded mock-OSCE?		
Do you understand your final year-end OSCE results will be collected for statistical comparison?		
Do you understand you can contact a researcher listed above to have your information removed from the study up until the end of the semester?		
Part 3: Signatures		
This study was explained to me by: _____ Date: _____		

Signature of Research Participant:

_____

Printed Name:

_____

Date:

_____

I believe the person signing this form understands what is involved in the study and voluntarily agrees to participate.

Signature of the Researcher:

_____

Printed Name:

_____

Date:

_____

Appendix D.

DSR Student Follow-up Survey

Student ID Number: _____

What is your best estimate of your current GPA, or average letter grade, in the dental hygiene program? (e.g. 3.7 and/or A-) _____

Did you received feedback following the mock-OSCE?

No, I received only one overall percentage (go to section 3)

Yes, I received the online feedback with examination breakdown into four categories (go to section 2)

Section 2 (Intervention group only)

Please answer the following, in regard to the feedback you received after the mock-OSCE:

1) Did you review your feedback?
    a. No
    b. Yes, briefly
    c. Yes, thoroughly
2) Did you use the feedback to help you prepare for your final OSCE?
    a. No
    b. Yes, a little
    c. Yes, a lot
3) Did you share your feedback with other students who did not get feedback after the mock-OSCE?
    a. Yes
    b. No

Section 3 (Control group only)

Please answer the following:

1) Would you have liked to receive more feedback following the mock-OSCE?
    a. Yes

b. No

2) Did one of your classmates show you/tell you about the feedback they received after the mock-OSCE, (i.e. the breakdown of scores into four components)?

   a. Yes

   b. No

3) Did you use that feedback to help you prepare for your final OSCE?

   a. No

   b. Yes, a little

   c. Yes, a lot

   d. Not applicable

Section 4 (All students)

Please answer the following referring to the feedback you received after your final OSCE:

1) The feedback I received after the final OSCE was helpful.

Disagree 1 2 3 4 5 6 7 8 9 10 Agree

2) The feedback I received helped me identify my strengths and weakness in history taking.

Disagree 1 2 3 4 5 6 7 8 9 10 Agree

3) The feedback I received will help me improve my performance in clinic.

Disagree 1 2 3 4 5 6 7 8 9 10 Agree

4) I understand how to use my feedback to make self-improvements.

Disagree 1 2 3 4 5 6 7 8 9 10 Agree

Appendix E.

Reflection Quality Grading Rubric[1]

- Code each comment (each sentence or unique idea)

What? (Thinking about the past) – "What did you do well?"		
Code	**Level of Reflection**	**Examples**
0	**Descriptive** – The student simply states what they did without elaboration or analysis of themselves, the client, or the context.	"I introduced myself" "I used open-ended questions" "I asked all 10 questions"
1	**Analytical** – The student interprets the event by adding context: explains why or implications ("so what?"), makes connections to other experiences (e.g. clinic) or coursework/literature.	"I introduced myself to establish a rapport" "I asked questions about her medical condition, so I could make sure there were no contraindications to care"

Now What? (Thinking about the future) – "What could you improve?"		
Code	**Level of Reflection**	**Examples**
0	**Descriptive** – The student acknowledges what they need to improve on, but without elaboration on how or why.	"I need to communicate better" "I could speak slower" "Know contraindications to care"
1	**Analytical/Practical Implications** – The student demonstrates a deeper understanding by establishing context ("why") or by describing a specific plan to improve ("how").	"I need to work on how I communicate with my patients by developing a rapport" "I need to review course materials on all contraindications to care"

Iterative process of rubric development. Started with HSERC levels of reflection and adapted to suit style of reflection data (e.g. two different questions resulting in different types of answers).

[1] Adapted from the University of Alberta HSERC interprofessional reflection guide (2016)

Appendix F.

Rubric for Coding Reflection Content

Domain	Code	Definition	Examples
Communication	C	Anything about improving verbal or non-verbal communication skills.	"Using open-ended questions" "Speaking slower" "Introduce self" "Be more organized"
Client Centered Care	CCC	Anything about engaging client in the conversation, or respecting the client's rights and opinions.	"Respect patient" "Address patient's chief concern" "Use client's name"
Eliciting Essential Information	E	Anything about making sure the proper questions and follow-up questions are used.	"I asked all 10 necessary questions" "Maybe there were more follow-up questions I could have asked"
Interpreting Findings	IF	Anything about using the information provided by the client to make care decisions.	"Know contraindications to treatment" "Know when premedication is required"
Other	O	Anything that does not fit into the above categories.	"Prepare better" "I did well overall"

One comment may relate to several different codes. For example, "I need to make sure I am asking all the right questions, so I can make sure there are no contraindications to care", would be coded as both eliciting essential information and interpreting findings.