Use Authorization

In presenting this thesis in partial fulfillment of the requirements for an advanced degree at Idaho State University, I agree that the Library shall make it freely available for inspection. I further state that permission to download and/or print my thesis for scholarly purposes may be granted by the Dean of the Graduate School, Dean of my academic division, or by the University Librarian. It is understood that any copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Signature _____

Date _____

Dental Hygiene Faculty Calibration to Enhance Inter-Rater Reliability with Calculus Detection: A Pilot Study

By

Lisa J. Santiago, RDH, BS

A thesis

Submitted in partial fulfillment

Of the requirements for the degree of

Master of Science in the Department of Dental Hygiene

Idaho State University

Spring 2015

COMMITTEE APPROVAL

To the Graduate Faculty:

The members of the committee appointed to examine the thesis of Lisa J. Santiago find it satisfactory and recommend that it be accepted.

_____

Jacque Freudenthal, RDH, MHE

Co-Major Advisor

_____

Denise M. Bowen, RDH, MS

Co-Major Advisor

_____

Jim Creelman, PT, DPT, MS. OCS

Graduate Faculty Representative

CERTIFICATE OF EXEMPTION

Idaho State
UNIVERSITY

Office for Research Integrity
921 South 8th Avenue, Stop 8046 • Pocatello, Idaho 83209-8046

June 2, 2014

Lisa Santiago, RDH, BS
234 Hunting Ridge Rd.
Roanoke Rapids, NC 27870

RE: Your application dated 6/2/2014 regarding study number 4107: Dental Hygiene Faculty Calibration to Enhance Inter-rater Reliability with Calculus Detection

Dear Ms. Santiago:

I agree that this study qualifies as exempt from review under the following guideline: 1. Research on educational practices in educational settings. This letter is your approval, please, keep this document in a safe place.

Notify the HSC of any adverse events. Serious, unexpected adverse events must be reported in writing within 10 business days.

You are granted permission to conduct your study effective immediately. The study is not subject to renewal.

Please note that any changes to the study as approved must be promptly reported and approved. Some changes may be approved by expedited review; others require full board review. Contact Tom Bailey (208-282-2179; fax 208-282-4723; email: humsubj@isu.edu) if you have any questions or require further information.

Sincerely,

Ralph Baergen, PhD, MPH, CIF
Human Subjects Chair

Phone: (208) 282-2592 • Fax: (208) 282-4723 • www.isu.edu/research

DEDICATION

This thesis document is dedicated to my family. Without your love and support I would not have made this incredible journey.

ACKNOWLEDGEMENTS

I would like to express my gratitude to each of the following individuals who guided me through this remarkable journey. I begin with Jacque Freudenthal and Denise Bowen, for giving me strength when I was weak and encouraging me through the thesis process. Also, I give my love and thanks to my family who have endured through my schooling and endless hours of patience waiting for me to "get this done." I want to thank Teri Peterson, who helped me set up the statistical portion of my study. To Jim Creelman, thank you for your support and patience. Thank you to my colleagues at Halifax Community College Dental Hygiene Department, for fueling the fire and agreeing to help in any way possible. And thank you to Caroline Harris, for formatting and converting my documents to make this thesis document possible.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

ADA	American Dental Association
ADHA	American Dental Hygienists' Association
BI	Barthel Index
CODA	Commission on Dental Accreditation
CRDTS	Central Regional Dental Testing Service
EMT	Emergency Medical Technician
HCC	Halifax Community College
HIPAA	Health Insurance Portability and Accountability Act
ODU	Old Dominion University
SOAP	Subjective data, Objective data, Assessments, and Plans
UCSF	University of California, San Francisco
UMKC	University of Missouri-Kansas City

ABSTRACT

Purpose: This pre-test post-test pilot study assessed the impact of faculty calibration on inter-rater reliability in calculus detection using two accepted standards: (a) established by CRDTS, and (b) scaling/root planing.

Methods: Participants (N=4) explored calculus on assigned teeth (N=16, 64 surfaces) on each patient (N=4). Calculus scores were calculated before and after calibration training.

Results: Kappa averages based on CRDTS criteria were 0.54 at pre-test and 0.62 at post-test. Kappa scores based on the scaling and root planing standard were 0.21 at pre-test and 0.22 at post-test. Although changes in inter-rater reliability were not statistically significant ($p \geq 0.05$), Kappa scores based on the CRDTS criteria improved from moderate to substantial agreement, but not using the gold standard, scaling and root planing.

Conclusion: Attaining inter-rater reliability using the gold standard may be more challenging. Future studies should use endoscopy in patients during training. Standards used for student evaluation should be evaluated.

Chapter I

Introduction

Challenges exist in dental hygiene programs to ensure reliability, agreement, and consistency in faculty assessments of students' clinical performance. Dahlström, Keeling, Friction, Hilsenbeck, Clark, and Rugh (1994) proposed, nearly two decades ago, that one way to improve reliability in clinical evaluation is to standardize clinical techniques. Over the course of time, researchers have sought methods to increase reliability in evaluation of student performance in many facets of dental education. Studies have concluded that discussion, comparison of technique, and continual practice – all components of faculty calibration training – have the potential to improve reliability (Dahlström et al., 1994; Assaf, Meneghim, Zanin, Tengan, & Pereira, 2006; Jacks, Blue, & Murphy, 2008).

Researchers studying faculty calibration training have postulated that students need to be aware of ideal performance standards and be allowed the opportunity to perform skills that mimic situations similar to those they will perform after graduation to achieve competency (Knight, 1997; Christie, Bowen, & Paarmann, 2007). Each day in clinical training, students can receive conflicting feedback from different faculty members regarding assessments and skills. Although no data exist to document the relationship between faculty calibration and clinical dental hygiene instrumentation learning outcomes, inconsistency in evaluation has the potential to distract student learning, lead to student concerns or frustrations, and detract from the ultimate goal of reaching competency (Henzi, Davis, Jasinevicius, Hendricson, Cintron, and Isaacs, 2005). Students might not understand clinical expectations if clinical assessments are

inaccurate and variable due to individual faculty member's using their own "gold standards" instead of a gold standard established and accepted by the profession. Albino, Young, Neumann, Kramer, Andrieu, Henson, Horn, and Hendrickson (2008) suggested assessment strategies in pre-doctoral dental education should be implemented in a way that will best assess students' readiness to provide patient care without faculty supervision. Faculty assessment methods are necessary components of a clinical evaluation system because these methods assist clinical faculty in making judgments about students' performance in relationship to attaining the "gold standards" accepted within the profession. Faculty should be consistent when making judgments regarding students' performance and need to know and evaluate students based on the gold standard. If faculty judgment is skewed, the chance of rater agreement or consistency is compromised (Knight, 1997).

According to Albino et al. (2008), assessments incorporated in the dental hygiene curriculum are a critical component of measuring the skills, knowledge, and professional values of future dental hygienists. These authors asserted that integral components of dental hygiene education such as proper procedures, instrumentation skills, and professional judgment required for comprehensive patient care need to be continuously evaluated. Further, evaluation of students' capacity to synthesize information within a given context and apply it in unique situations that require critical thinking and problem solving also is necessary.

The problem with the lack of inter-rater agreement among dental hygiene faculty members is widely recognized, and many authors have suggested that problems of inconsistency among faculty can be reduced with calibration training (Dahlström et al.,

1994; Flack, Atchison, Hewlett, & White, 1995; Assaf et al., 2006; Lanning, Best, Temple, Richards, Carey, McCauley, 2006; Jacks et al., 2008). Previous studies have implemented different approaches to calibration depending on the specific problems and needs each institution faced. Research findings documenting the value of dental and dental hygiene faculty calibration and the impact of inter-rater and intra-rater reliability on student learning and satisfaction are limited in scope and depth. Additional studies are needed to determine which methods of faculty calibration generate the best results (Haj-Ali & Feil, 2005; Lanning et al., 2006; Dahlström et al., 1994; Flack, Atchison, Hewlett, & White, 1995; Jacks et al., 2008).

Results of faculty calibration studies are mixed. Successful calibration training has been documented, resulting in reliability, agreement, and consistency among faculty. Researchers Haj-Ali and Feil (2005) conducted a study measuring long- and short-term effects of calibration and inter-rater agreement between faculty grading Class II amalgam preparations. Training faculty for ten weeks showed subsequent improvement in inter-rater agreement; however, following the next ten weeks, improvement deteriorated. Jacks et al. (2008) conducted long- and short-term calibrations for faculty members' capacity to write specific notes, called SOAP notes (Subjective data, Objective data, Assessment, and Plans). Supporting the findings of Haj-Ali and Feil, they concluded that faculty maintained agreement up to one year after the initial workshop. The researchers questioned the frequency faculty should continue to calibrate and recommended additional research on the subject. An evaluation of faculty calibration training to enhance evaluation of ethical reasoning and professionalism was conducted by Christie et al. (2007). Following three years of intermittent faculty calibration training, assessments

showed increased ethical reasoning by students and consistent faculty evaluations of the students.

Other studies have found that faculty calibration was unable to achieve agreement in faculty evaluation. Lanning, Pelok, Williams, Richards, Sarament, Oh, & McCauley (2005) assessed faculty agreement in diagnosis and management of periodontal diseases before and after calibration training. They found significant variability among faculty members' interpretation of the gold standard accepted for evaluation of adequacy in both the pre-test and post-test following a single calibration session. They concluded that accepted guidelines and on-going training efforts should be practiced among faculty (Lanning et al., 2005). In 2006, Lanning et al. conducted another calibration project with a second session to improve the consistency and accuracy of interpretations regarding the amount of visible bone loss shown on dental radiographs. Of the four categories measured, one category had significant consistency and reliability after calibration efforts. These researchers again concluded that further training might enhance consistency and reliability of instructors' radiographic interpretation.

Pippin and Feil examined inter-rater agreement among a large pool of clinical examiners (that explored and scored residual subgingival calculus using patients. Findings indicated only fair agreement after calibration, although substantial agreement (Kappa 0.061-0.80) or nearly perfect agreement (0.81-1.0) is desired for calibration of faculty examiners. A faculty calibration study by Partido, Jones, English, Nguyen, and Jacks (2015) tested intra- and inter-rater calculus scores using a pre-and post-test design with the use of a dental endoscope. Using three typodonts with simulated calculus and a manufacturer's answer key, both groups explored calculus using 11/12 explorers. The

training group received two one-hour calibration sessions using didactic, discussion, and individualized instruction on the use of an endoscope. If any errors were discovered, the subjects were allowed to reconcile errors immediately with re-detection with the endoscope and an 11/12 explorer. Results showed significant increases in inter-rater reliability levels for those in the training group but no significant differences were found with intra–rater reliability calculus detection scores between groups. Garland and Newell (2009) concluded there was no effect on inter-rater reliability levels for simulated calculus detection after faculty calibration training, although rater agreement was substantial before the calibration training leaving little room for improvement. Recommendations for future research included calibration based on evaluation of patients rather than by using artificial calculus deposits on typodonts, a model that represents a hinged mouth with realistic teeth positioned anatomically correct.

Faculty calibration for calculus detection presents a unique challenge relative to the existence of two different accepted standards for evaluation of calculus removal. The standard established by the Central Regional Dental Testing Agency (CRDTS) and other similar agencies is described as "readily detectable calculus, a definite 'jump' or 'bump' which is easily detected with one or two strokes; a deposit that easily 'binds' or 'catches' the explorer; ledges or ring formations; spiny or nodular formations (CRDTS Dental Hygiene Treatment Selection Worksheet, 2014). CRDTS reports successful calibration of its examiners for dental hygiene testing (Haladyna, 2011). The gold standard accepted for evaluation of adequacy of scaling and root planing in clinical practice and education differs because complete removal of calculus is necessary for the health of adjacent tissues. Although there is little evidence that complete root surface smoothness is a

necessity, relative root smoothness is accepted as the best immediate clinical indication that calculus has been removed (Newman, Takei, Klokkevold, & Carranza, 2012). This standard includes not only readily detectable calculus but also light deposits and relative root roughness determined to be calculus by use of tactile sensitivity, elimination of plausible reasons for the roughness, and clinical judgment based on experience. Over instrumentation for the sole purpose of complete root smoothness is not recommended. For this reason, the ultimate evaluation of calculus removal occurs four to six weeks following treatment and is based on tissue response, as the tissues will not heal in the presence of bacterial deposits. Nonetheless, clinical evaluation of calculus deposit removal based on relative root roughness is necessary immediately after instrumentation using visual inspection and tactile inspection with a fine explorer or probe. This gold standard is accepted for clinical evaluation of students' removal of calculus deposits and likely presents a greater challenge for faculty calibration than the gold standard used by testing agencies such as CRDTS for evaluation of minimal competency required for licensure.

Although many authors have been unsuccessful in attaining calibration, all agreed that faculty training was beneficial and concluded that different approaches or multiple calibration sessions might be needed; however, the specific time and mechanism needed for calibration remains undetermined (Assaf et al., 2006; Flack et al., 1995; Dahlström et al., 1994; & Lanning et al., 2005).

Statement of the Problem

Research studies assessing dental and dental hygiene faculty calibration in evaluation of students' performance of clinical skills, including calculus detection, have

not consistently documented an improvement in inter-rater reliability of faculty ratings of students' performance.

Purpose of the Study

The intent of this study was to evaluate the impact of dental hygiene faculty calibration training on inter-rater reliability of faculty evaluation during calculus detection of faculty scores using two different accepted standards.

Professional Significance of the Study

The Commission on Dental Accreditation's (CODA) current Accreditation Standards for Dental Hygiene Programs (2013) requires faculty development through periodic workshops and in-service sessions and regular review and documentation of the program's curriculum review process (Commission on Dental Accreditation, 2013). These standards are intended to support attainment of student competency through proper curriculum review and faculty training. When an institution identifies areas of the curriculum needing improvement, such as evaluation of student performance, changes can be implemented to correct shortfalls. Inconsistencies in evaluation of students' performance by clinical faculty have been identified in the literature as an area in need of improvement in dental and dental hygiene programs.

Null Hypotheses

1. There will be no significant difference in inter-rater reliability of faculty calculus detection scores as measured by Cohen's kappa following calibration training designed to enhance agreement of faculty members' assessment of calculus deposits based on the standard established by the Central Regional Dental Testing Service, readily detectible calculus.

2. There will be no significant difference in inter-rater reliability of faculty calculus detection scores as measured by Cohen's kappa scores following calibration training designed to enhance agreement of faculty members' assessment of calculus deposits based on the gold standard established for evaluation of adequacy of scaling and root planing, relative root smoothness, before and after faculty calibration training.

Operational Definitions

- ***Inter-rater reliability***. The consistency or agreement among two or more raters (Vogt, 2005). In this study, agreement was measured by a Cohen's kappa value greater than or equal to 0.80.
- ***Faculty calibration training.*** A program designed to train faculty members to use the specific gold standard established by the profession and determined by the dental hygiene program administrator, clinical coordinator, or faculty. The faculty are also trained to reproduce the standard repeatedly as practiced during the training sessions (Garland & Newell, 2009). In this study, faculty calibration training included calculus exploring techniques using an 11/12 explorer on typodonts and patients with various types and amounts of calculus, or calcified deposits, on the teeth.
- ***Accepted standards for evaluation of calculus detection.*** Valid criteria and/or rating scales used to achieve "the best" performance comprise the gold standard (Vogt, 2005). Dental professionals provide care for their patients according to the gold standard that is well established in the profession. Helping students develop skills and competence based on the gold standard is a goal of professional

programs. Assessments and evaluations ideally use gold standards as benchmarks or criteria to measure skills and techniques. In this study, two accepted standards were used for comparison with faculty calculus detection scores. The competency standard established by CRDTS is described as "readily detectable calculus, a definite 'jump' or 'bump' which is easily detected with one or two strokes; a deposit that easily 'binds' or 'catches' the explorer; ledges or ring formations; spiny or nodular formations. (Appendix A) The gold standard accepted for evaluation of adequacy of scaling and root planing immediately prior to and following instrumentation is relative root smoothness (Newman et al., 2012).

- ***Calculus detection assessment and scoring*.** The calculus detection assessment score was determined from an evaluation of hard deposits (calculus) on tooth structures using an Old Dominion University (ODU) 11/12 explorer. Patients with varying degrees of calculus accumulation were used for assessing the gold standards and inter-rater reliability of faculty calculus detection before and after calibration training. A calculus detection form used by CRDTS was utilized during the study (Appendix A). The calculus detection form developed by CRDTS has 32 boxes. The boxes represent the 32 teeth in each patient's mouth. Each box has four divided areas that correspond to the following tooth surfaces: M=mesial; D=distal; F=facial; and L=lingual. Each site was worth 1 point, earned by identifying calculus as present or absent on a specific tooth surface. If readily detectible calculus as defined by CRDTS was detected on a particular site, the corresponding box was marked with a "C" if light calculus or root roughness is detected indicating the presence of calculus, the corresponding box was marked

with an “R”, each representing a 1 point value. If no calculus was detected, the corresponding area was left blank, corresponding to 0 point. Each patient had one half of their mouth explored by three subjects, while the other half of their mouth was explored by two subjects. Subjects explored only half of the patient’s mouth to minimize patient discomfort. Half of the tooth surfaces present in each patient’s mouth (7 to 10 teeth) were scored for a total of 28 to 40 points possible at two intervals: (a) before training, and (b) immediately after training.

Summary

The purpose of this pilot study was to test whether faculty calibration training improves inter-rater reliability during calculus detection based on two accepted standards: a standard established by CRDTS and the gold standard accepted for evaluation of adequacy of scaling and root planing. This study was designed to build upon a small but growing body of research on the inter-rater agreement of dental and dental hygiene faculty and the effects of calibration. Research findings are mixed, indicating calibration training results were positive, negative, or had no impact. All previous researchers suggested calibration training was necessary and should be continued. When faculty members are in agreement, consistent, and reliable, students potentially receive feedback that reflects the gold standard of the professional programs.

Chapter 2

Review of the Literature

This literature review begins with a discussion of short- and long-term effects of faculty calibration training and then summarizes studies specifically regarding training to increase reliability of evaluation of periodontal disease assessment including calculus detection. The literature on dental hygiene faculty calibration, inter-rater and intra-rater reliability is limited. Therefore, the search of the literature and this review were expanded to include faculty in the allied health professions and other healthcare disciplines. Lastly, a review of calibration training in other aspects of dental education will be presented. This literature review demonstrates that, to date, dental hygiene faculty calibration training has not been shown to consistently improve inter-rater and intra-rater reliability of faculty evaluation during calculus detection.

Dental and Dental Hygiene Faculty Calibration Training: Short- and Long-Term Effects

The consensus of the results from the literature review indicates low agreement and high variability among educators and evaluators who assess student achievements (Lanning et al., 2005; Lanning et al., 2006; Russian et al., 2008). The diversity of faculty members' interpretations during student assessment leads to questions regarding the validity and reliability of evaluation. Russian et al. (2008) stated, "Clinical instructors possess diverse educational backgrounds and may interpret 'acceptable' clinical performance differently" (p.8). Further, if reliability of evaluators is low, student achievement may be affected. The following section of this literature review includes

studies of the short-term impact of calibration sessions on intra-rater and inter-rater reliability of faculty.

Lanning et al. (2006) reviewed the literature on low agreement of dentists in the evaluation of alveolar bone levels on radiographs and conducted a short-term study to determine the accuracy and consistency of clinical instructors' ratings of bone loss percentage on digital radiographs. The sample included 35 clinical instructors from the University of Michigan School of Dentistry. This quasi-experimental study had a one-group, time series design. Twenty-five radiographic images of acceptable quality were categorized into four categories of bone loss including- (a) none; (b) less than 15%; (c) between 15-30%; and (d) greater than 30%. The standardized measuring tool for bone level was a Schei ruler, measured to the nearest 5% as determined by three of the authors. A pre-test was conducted, followed immediately by a calibration session. Twenty-five radiographs were projected for 30 seconds each, and all subjects (N=35) rated bone loss. A post-test followed the training, with a second training (n=22) conducted immediately for one hour.

Three months later, the same 25 digital radiographs were used for a repeated post-test, followed by another training session (n=17). A kappa coefficient was used for all testing, with mixed results. Inter-rater reliability was greatest in the no bone loss category, with accuracy diminishing with subsequent tests (Lanning et al., 2006). Pre-test results showed a 64.5% agreement between faculty participants. For post-test 1, the inter-rater percent agreement was 76.5%, and for post-test 2, it was 85.2%, showing a positive change from pre-test results. The researchers discussed limitations in their study, such as the sample size bias resulting from a decrease in sample size during testing and the

quality of the digital radiograph projected on the screen for the radiographic bone loss assessment (Lanning et al., 2006). The sample group types, such as dental hygiene faculty (n=6 pre-test, n=4 post-test 1, n=3 post-test 2), graduate students (n=16 pre-test, n=8 post-test 1, n=5 post-test 2), and periodontal faculty (n=13 pre-test, n=10 post-test 1, n=9 post-test 2), were not equally represented in numbers. Also, this research project was designed to provide immediate feedback to the subjects. With an extension of the training program, the researchers believed inter-rater reliability might have improved radiographic interpretation in education assessments (Lanning et al.).

Despite limitations, Lanning et al.'s study showed that, with calibration sessions, faculty may improve inter-rater reliability. Another study conducted in 2006 by Haj-Ali and Feil also showed that some aspects of calibration achieved high inter-rater agreement while other aspects needed improvement.

Haj-Ali and Feil (2006) discovered that a high level of inter-rater agreement can result from a short-term training session regarding evaluation of Class II amalgam tooth preparations. These researchers studied the immediate effects of calibration by means of a quasi-experimental approach, using a one-group study design, with a theoretical framework based on recommendations by Knight (Haj-Ali & Feil; Knight, 1997). A grade sheet with criteria representing the gold standard was generated from Knight's guidelines. Haj-Ali and Feil's assessment tool consisted of a three-level grading scale requiring faculty members to make a judgment of student performance: "ideal" (scored as 2); "acceptable" (scored as 1); and "standard not met" (scored as 0).

The sample consisted of nine volunteer instructors: four full-time faculty, two part-time faculty, and three graduate students in one of the school's graduate programs.

The calibration session was the independent variable, and agreement with the gold standard (criteria on the grade sheet) was the dependent variable. Inter-rater procedures were measured during three sessions: a pre-test, a post-test immediately after calibration training, and another administered ten weeks after training. The pre-test used a random selection of ten prepared teeth completed by students, with raters using periodontal probes, explorers, mirrors, and voluntary use of loupes. Calibration training followed immediately after the pretest, starting with a twenty-minute PowerPoint presentation explaining each of the thirteen criteria with photographs of the three possible outcome scores. The raters then regrouped, conversed about the gold standard, and evaluated two teeth, before assessing ten new Class II preparations using the thirteen criteria representing the gold standard. With no additional training from this point, ten weeks later, an additional ten new class II amalgam preparations were rated using the same criteria. Percentages of intra-rater agreement with the gold standard were 54.5% (σ =21.3) at pre-test, 66.9% (σ=18.0) immediately following the calibration training session, and 64.6% (σ =19.1) at ten weeks post training. Data suggested that, with calibration, raters achieved a higher level of agreement with the gold standard when the scores were 0 and 2. However, no improvement in the mid-point score of 1 was reached. As far as the agreement among faculty members with the gold standard, Haj-Ali and Feil (2006) concluded that "the large standard deviations suggest substantial variability" (p.431).

These mixed results show that there can be problems attaining calibration during faculty training. The ten week timeframe for follow up after calibration training showed that, despite thorough training, faculty could not provide consistent feedback to students.

Time, effort, and consistency are desirable in order to reach a satisfactory level of agreement when calibrating faculty (Haj-Ali & Feil).

The perception of success or failure in faculty calibration can be somewhat dependent upon the clinical judgment of dental and dental hygiene clinical faculty. Researchers Gansky, Pritchard, Kahl, Mendoza, Bird, Miller, and Graham (2004) provided short-term calibration sessions for raters grading a manual dexterity test for potential dental students during the admissions process. The University of California, San Francisco (UCSF) School of Dentistry wanted to determine if dexterity test scores correlated with subsequent grades in preclinical courses. During the application process, dental schools look at a diverse set of tests to help determine if a candidate has the potential to be successful. One test, the block-carving test of dexterity, was administered to students who were admitted to UCSF School of Dentistry during the years 1996-1998 (Gansky et al., 2004). Of the 256 students, 244 were chosen for evaluation due to problems with code number assignments in order to protect the students' identity.

Ganski et al. described the study's methods as follows. Three calibration sessions, each at one week intervals, were performed with eight faculty members (three preclinical dental faculty, three senior dental faculty, and two senior basic science faculty) before the block test was evaluated. The two non-dental members were used as a comparison group to determine if the dental faculty members were being overly critical in grading (Gansky et al.). Each calibration session began with reviewing the four grading criteria followed by an exercise of grading twelve blocks independently. Each block was graded as either "pass" or "fail." Afterwards, the faculty shared and compared grades and discussed discrepancies that arose.

Two grading sessions were performed with the six dental examiners, three examiners per grading session. Twelve pre-graded blocks were used, three each from four categories: "highly acceptable," "acceptable," "marginally acceptable," and "unacceptable," with examiners not knowing how many blocks were from each category. Each examiner graded the 12 blocks independently, assigning Pass/No Pass grades. After calibration, the 244 blocks were graded within 60 minutes. Each instructor began grading on a different laboratory bench, with no talking allowed. Grades were later compared and a consensus was achieved.

Results of intra-rater reliability showed that faculty pre-exam calibration sessions varied in their kappa statistics, ranging from a low of 0.34 to 1.00 (Gansky et al., 2004). Interestingly, the authors observed the two non-dental faculty members used as a control were well within the percentage range of the dental faculty examiners, demonstrating that dental experience was not needed to grade this examination. Further, two of the three examiners gave 13% (n=33) - no Pass grades, with 7% (n=16) in unanimous agreement regarding no Pass. The authors of the study concluded that if dexterity tests are used in the admissions process, faculty calibration training should be used to improve reliability, accountability, and consistency among faculty for the scores they assign to this admissions test.

Knight conducted one of the most cited studies in dental pre-doctoral research in faculty calibration research in 1997 at the University of Detroit Mercy School of Dentistry. This short-term calibration training study was planned in three phases: phase one was criteria development; phase two was the development of a training program; and phase three was faculty ratification of evidence of calibration in order to receive rank and

tenure at the institution. Knight described the pilot study. It was conducted to discover if a long detailed criteria form or an abbreviated version of the criteria form resulted in more accurate faculty evaluations of student performance. He trained a group of faculty, subject number unknown, using a pre-established criteria form to assess prepared teeth. After training, faculty members were randomly assigned to either the long form or the abbreviated form. Each group then performed a pre-test, then immediately completed several calibration exercises. After these practices, Knight conducted post-test showing several results: with limited training; the form type did not influence the evaluation accuracy; and neither group achieved a high level of reliability. Knight's pilot study did not show positive results from calibration training.

From the pilot study Knight hypothesized that, "because faculty are using instructional strategies for teaching novice students to self-evaluate, perhaps these same strategies would work well for faculty" (Knight, 1997, p.945). According to Knight, faculty should begin evaluations by assessing differences of each criterion in a systematic manner, encouraging correct terminology, and refining visual perceptual skills. If clinical faculty members understand the gold standard representation of the product, they could possibly assess errors. The study's findings provided a foundation for many faculty calibration training studies, even though Knight's study showed no effect of calibration training on rater agreement. Other researchers have used Knight's model and hypothesized that longer calibration training could make improvements in rater agreement; however, findings of longer term studies remain mixed.

More research is needed to answer the question regarding the exact approach to frequency of training. Is short-term training not as worthwhile as long-term training for

gaining faculty agreement in ratings of student performance? So far, the findings seem to indicate that faculty calibration should be ongoing and continuous, having as many as possible, if not all, faculty members participate. The following faculty calibration study suggested that short-term calibration may make long-term improvements in inter-rater reliability.

In 2008, Jacks, Blue, and Murphy conducted a study to see if a training session on writing SOAP notes could achieve a high level of short term intra-rater reliability and to determine if inter-rater reliability could be maintained for a year following the training. They concluded, based on this study's findings, that consistency of the clinical evaluators' feedback to students, or standardization of instructors through training sessions, can yield positive results for students. The Jacks et al. study assessed the short- and long-term effects of calibration training on standardization of SOAP notes. The conceptual framework was the SOAP format, a standardized procedure for chart entries incorporating subjective data (8 points), objective data (8 points), assessment (5 points), and plans for treatment (8 points).

Jacks et al. employed quantitative, quasi-experimental research, using a one group pre-test/post-test design. The sample group consisted of eight dental hygiene faculty members. A pre-test based on a fictitious clinical case was followed one week later by a training session for calibration training in writing SOAP notes and an immediate posttest on the case. One year after pretest, a follow-up test was administered using the same case. Jacks et al. reported a mean pretest score of 18.25 out of 29 points possible, the mean score directly after training as 24.63, and the mean post-test score after one year as 22.75. The authors found an improvement of 35% in the consistency of

SOAP note scores was noted between the pretest and post-test scores; however, the follow-up test was slightly lower (78.4%) short-term than post-test training (84.9%).

A significant result from the study showed faculty members were consistent in writing SOAP notes in the correct manner, immediately after and one year from the pre-test. Jacks et al. asked a very important question: at what frequency should recalibration occur? It is unknown whether the time frame is related to the skill being assessed or the standards being applied. This study as well as Haj-Ali and Feil (2006) showed that calibration training might make improvements in percent agreement of faculty scores with the gold standard or in inter-rater reliability, but more research is needed to determine the interval for re-calibration needed for improvements to be ongoing and consistent (Jacks et al., 2008).

Calibration can be used not only in the clinical arena of dental hygiene education, but in other aspects of learning. Educators are constantly trying to improve teaching methods to help their students improve in affective areas such as critical thinking and ethics, to name a few. Many dental hygiene educators attend continuing education classes and workshops to learn new methodology to improve learning and evaluation of affective outcomes. Calibration of dental hygiene educators can assess if such workshops are indeed helping educators understand the gold standard determined by their educational standards. Two long-term studies assessing outcomes of faculty calibration on these affective aspects of evaluation indicated improved faculty attitudes and assessments following calibration training.

Researchers Wallace and Infante (2008), from the University of Texas Health Science Center at San Antonio, wanted to assess the quality of their past faculty

development workshops on clinical teaching strategies for dental hygiene educators. By using a purposeful sampling method, they asked only subjects who attended at least two workshops from 2000-2004 to participate. These researchers used a qualitative Web-based questionnaire developed with Survey Tracker software, was assessed as a pilot with faculty at five dental hygiene programs. The final study had 142 participants from 28 dental hygiene programs. E-mail addresses were confirmed and online assessments with four open ended questions were sent with no further follow-up. Narrative data were analyzed to establish major themes, and the researchers determined the workshops have made a positive impact on participating dental hygiene instructors and their institution's curriculum.

Wallace and Infante (2008) reported that many instructors who attended the workshops began thinking critically about the teaching techniques in place in their programs. The workshops provided a place to try and practice new skills. The researchers found that dental hygiene instructors seemed to benefit from an array of annual workshops compared to singular presentations. Some of the positive changes instructors implemented after the workshops were the redesigning of clinical programs to mirror competency-based concepts and the incorporation of appropriate verbal and nonverbal feedback to students. The number one barrier Wallace and Infante discovered from attendees following the workshops was a resistance to change from non-participating faculty and administration (Wallace & Infante). The researchers incorporated these observations to improve future workshops, strengthening future curricula in programs throughout America. This study provides evidence of the benefits of faculty development workshops.

Dental hygiene institutions re-evaluate their curriculum in order to strengthen the different aspects of education. One such aspect in dental hygiene education is that of ethics and professionalism. Students should have many opportunities during their education to develop and use ethical decision making skills to resolve competing interests that will be encountered during their careers (Christie, Bowen, & Paarmann, 2007). The subject of ethics and professionalism, when judged or evaluated, can be subjective in nature, which can lead to inconsistencies and unreliability in assessment evaluations.

The following study demonstrates how a program evaluated and identified a need for improvement in the area of ethics and professionalism, and incorporated a two-fold plan of implementing a new assessment tool and faculty calibration training to enhance the tool. In 2002, Idaho State University Department of Dental Hygiene recognized the section of ethics and professionalism on student assessments lacked critical evaluation, feedback, and written comments from instructors. Christie et al. (2007) realized in order for students to recognize and use critical thinking skills to assess and act accordingly to ethical and professionalism breaches, a system needed to be developed. This system would be incorporated in students' daily clinical activities where daily ethical decision making transpires. Also, since the subject of ethics and professionalism can be a subjective topic to assess for instructors, the program designed by Christie et al. included a component to re-familiarize faculty to the American Dental Hygienists' Association (ADHA) and American Dental Association (ADA) codes of ethics, which provided the structure upon which the new assessment was based. The faculty development project's goals were to enhance the evaluation of ethical reasoning and professionalism in the

clinical component of the baccalaureate dental hygiene program, and to improve faculty perceptions of apparent grade inflation in the area of professionalism (Christie et al.).

The faculty development workshop included fourteen full-time and part-time instructors, and all were provided materials to read and review before the four-hour workshop (Christie et al., 2007). Faculty members identified core values, also known as ethical principles, and agreed how to express them into comments. A laminated card with word descriptors relating to these core values was dispensed to each faculty participant. Faculty used these descriptors throughout the year during clinical evaluation of students' clinical assessments (Christie et al.).

Christie et al. evaluated results at the end of each school year during the three years the instructors met to discuss the outcomes on the assessments and to evaluate discrepancies. The program director also conducted exit assessment evaluations of the training program from recent graduates. Results indicated the total number of comments provided to students regarding ethical decision making and adherence to the core values during the pre-workshop and three-year post faculty workshop totaled 1,443, with a total of 92.92% (r=72.7-98.8) providing comments in agreement with the examples provided in training and reinforced throughout the time period. Christie et al. concluded students and faculty believed that faculty training was achieving its purpose (Christie et al., 2007) based on the finding that the percent agreement with the gold standard was significantly higher than the pre-workshop agreement of 20.27%.

Christie et al. (2007) stated the following:

> Faculty training supports the culture of ethics and professionalism by promoting awareness, helps clinical evaluators to consider ethical reasoning and dilemmas

where performing daily evaluations, and sends a message to the entire faculty that the issue of ethics is important to the institution (p. 1049).

Faculty calibration not only applies to heightening faculty members' awareness of gold standards for clinical skills but also for less tangible outcomes such as ethical decision making and critical thinking. These curricular components are given increasing importance by accrediting agencies and professional associations.

Since the landmark research conducted by Knight, dental hygiene institutions have realized clinical faculty might not be reliable, consistent, and valid in grading despite the existence of criteria outlining the gold standard. The studies discussed thus far in this literature review have measured the effectiveness of short- and long-term calibration training for dental and dental hygiene faculty (Gansky et al., 2004; Jacks, et al., 2008; Haj-Ali & Feil, 2006; Knight, 1997; Lanning et al., 2006). Most of them found some level of improvement with reaching higher rater reliability after calibration, although results were mixed. The next section will continue reviewing the research on inter- and intra-rater reliability, with a focus on periodontal assessment.

Faculty Calibration in Periodontal Assessment

Since researchers initially began exploring whether calibration would improve the reliability of student assessment in some aspects of dental and dental hygiene education, studies have been conducted to assess calibration techniques in many realms of the curriculum. One highly critical and subjective area of curriculums for oral health professionals is periodontology, or the study of periodontal (gum) disease. For example, when determining the course of treatment for a patient with periodontal disease, practitioners have differing viewpoints on the manner in which the patients should be

treated (Lanning, Pelok, Williams, Richards, Sarment, Tae-Ju, & McCauley, 2005). Faculty in dental and dental hygiene institutions have different levels of experience, education, and perspectives regarding periodontal diagnoses and therapy. These differing perspectives have the potential to confuse interpretation and application of the gold standard for the dental or dental hygiene institution.

Dental and dental hygiene students are exposed to numerous clinical instructors during their education, with moments of frustration and confusion due to low agreement and variation in clinical judgment from faculty (Garland & Newell, 2009). One assessment of students' clinical performance evaluates calculus removal, an outcome essential to periodontal health. Clinical instructors use an instrument called an explorer to assess, or to explore, tooth surfaces to detect calculus not removed by the student. Using a process paralleling a standard used by CRDTS when calibrating board examiners, Garland and Newell (2009) conducted a two-group randomized experimental pilot study to determine if calibration would solve the inter-rater reliability problem informally identified by students in their dental hygiene program.

Garland and Newell (2009) randomly assigned dental hygiene faculty (N=12) to two groups, experimental and control employing the following methods. All faculty members explored three typodonts with different amounts of imitation calculus present. The subjects provided a yes or no answer when calculus was detected on tooth surfaces for the pre-test and post-test on an answer key developed by the researchers. The experimental group received three two-hour training sessions regarding specific standards for calculus detection and assessment. The calculus exploring technique and sequence was similar to the process used by CRDTS, although light deposits and roughness were

simulated as well as readily detectible deposits. Cohen's kappa coefficient was used to determine inter-rater reliability along with ANOVA to determine differences in the scores of the two groups over time. Garland and Newell (2008) found the training sessions did not affect the differences in rater reliability of calculus detection scores between the two groups. The researchers postulated that a high level of inter-rater reliability before the study may have affected the results. They also recommended evaluation of actual calculus deposits in human subjects because of the difficulty in finding a suitable simulation with typodonts. To examine and measure calibration training in future research, a larger, more diverse sample also would be necessary. The authors concluded there was subjectivity with faculty perceptions concerning calculus detection, and the faculty calibration training did not significantly improve rater reliability.

Recently, a faculty calibration study by Partido et al. (2015) tested intra- and inter-rater calculus scores with the use of a dental endoscope. A pre-and post-test was used with a control and training group. Using three typodonts with simulated calculus and a manufacturer's answer key, both groups explored calculus using 11/12 explorers twice to attain a baseline. The training group received two one-hour calibration sessions using didactic, discussion, and individualized instruction on the use of an endoscope. If any errors were discovered, the subjects were allowed to reconcile errors immediately with re-detection with the endoscope and an 11/12 explorer. A scoring of 80% accuracy with the answer key and a 0.80 Kappa score, a similar Kappa average to regional clinical boards, was the pre-established standard for mastery. At post-testing, both groups re-evaluated the original three typodonts with 11/12 explorers. Results showed significant increases in inter-rater reliability levels for those in the training group (Kappa score pre-

test 0.536 and post-test 0.792), but no significant differences were found with intra–rater reliability calculus detection scores between both groups.

Another challenging area to assess in dental and dental hygiene schools is periodontal disease classification. The problem in many dental schools is that of variation among periodontal faculty in diagnosing periodontal disease. "If faculty members are providing an ever-moving target, students' ability to differentiate between accuracy and inaccuracy could be lost" (Lanning, Pelok, Williams, Richards, Sarment, Oh, & McCauley, 2005, p. 336).

Lanning et al. (2005) conducted a qualitative study that measured 27 faculty responses to three Web-based case studies regarding periodontal diagnosis, clinical findings, and treatment planning. The researchers provided a medical history, dental history, chief complaint, and patient assessments in each case. After reviewing the case exercise independently, their subjects proceeded to answer a twenty-item questionnaire, of which only five items were used for the study: subject's position at the school; years of experience; an open-ended question on the periodontal diagnosis of the case, describing the extent, severity, and type of periodontal disease; the treatment plan including the procedure, ADA code used, and quadrant indicated for treatment; and percentage of bone loss displayed in the case. All subjects then attended a department meeting to discuss concerns on variability and accuracy of responses (Lanning et al., 2005).

Lanning et al. reported the following results. The response rate ranged from 62% - 70%, lower than the accepted rate of 80%. Case one had a diagnosis of gingivitis, correctly identified by 67% with no general agreement of treatment plan. Case two had an agreement of 96% of some type of periodontitis, a more advanced form of periodontal

disease, with a 75% agreement on the treatment plan. For case three 12% of the subjects diagnosed gingivitis, and 88 % diagnosed some form of periodontitis. Treatment plan evaluations were varied with no general consensus.

The researchers identified limitations in the study, such as cross-contamination among subjects (Lanning et al., 2005). This cross-contamination invalidated the data, and the use of Web-based cases was an inadequate representation of in vivo patients' diagnosis. Even though the study had flaws, it was the stepping stone for future research in the field of periodontal inter-rater reliability and calibration. Lanning, et al. showed that, with or without calibration training, a huge variance existed in interpretation, hence low consistency and reliability in periodontal diagnosis and treatment planning.

Another ground-breaking study conducted by Taleghani, Solomon, and Wathen (2004) showed that, through calibration training of faculty using a new periodontal assessment protocol, instructors could become more reliable and valid in assessments. The Baylor College of Dentistry implemented a new grading system of clinical student evaluation and tested it with senior dental students in the fall of 2003. This new assessment system required faculty to enter written explanations for evaluations of performance that was outside of clinically acceptable criteria. The school wanted a sufficient way to encourage teaching opportunities between faculty and students as well as discussion time with students (Taleghani et al.).

Teleghani et al. (2004) included an examining team of 19 instructors (seven full-time and 12 part-time) who agreed to participate and were calibrated to the clinical standards agreed to by consensus. They used thirteen forms which reflected the Baylor College of Dentistry's competency document for the assessments. As the research team

provided assessments and instant feedback to students and students could not continue if an aspect of the assessment did not meet minimum requirements. Trends in progress were easily identifiable and faculty calibration sessions could be guided according to these trends (Taleghani et al., 2004).

Teleghani et al. (2004) reported the following results. Before graduation in 2003, the senior students completed evaluation surveys about the effectiveness of the non-graded clinical evaluation with a 99% return rate. Eighty percent of senior students indicated they received more constructive comments from their instructors using the new system. The faculty (95%), with a 100% return rate, stated that they believed that the new system provided an improved learning environment.

Both students and faculty want a productive and healthy learning environment for success of all involved. Through a combination of calibration training and a well-designed assessment tool, faculty may have a reliable and valid template to measure inter-rater reliability.

Rater Reliability in Other Health Professions Education Programs

Dental education is not alone in researching inter-rater and intra-rater agreement to validate the reliability of their assessments. Health professions education has similar problems with subjectivity of faculty and poor rater reliability. Health professions faculty are conducting inter-rater reliability studies to evaluate its impact on consistency, validity, less subjectivity, and proper mentoring (Russian et al., 2008).

A pilot study was developed by Russian et al. (2008) to determine the inter-rater reliability of respiratory care clinical instructors' (N=9) evaluations of their students using the following methods. A sample of nine instructors each received a video-tape of a

student performing three procedures: hand washing, incentive spirometry, and vital signs. These procedures were assigned a score as either "satisfactory," "major or minor unsatisfactory," "not applicable," or "not observed." An answer key was developed by the researchers for the study as the gold standard for evaluation of competency. To ensure that the competencies entailed all elements, they used DataArc,® a biomedical database provider developed to follow several areas associated with clinical instruction in this field. The DataArc® assessment software used a check list evaluation system. This system was designed to list all criteria in a step-by-step manner. When a criterion was not observed, it would be checked. This check list system differed from the rating scale assessment system in which criteria may not be listed in a step-by-step fashion and evaluators must make a judgment call, making the evaluation subjective.

Russian et al. (2008) designed the session which consisted of faculty evaluation of filmed exercises related to competencies which had moments of incorrect behavior to maintain a realistic scenario to the instructors. A check list was completed by faculty members for each of the three procedures and results were returned anonymously in envelopes to avoid evaluator bias. The researchers then calculated the competency elements using the inter-class correlation. The results showed that 12 subjects had an agreement level less than 78%, while 11 subjects had less than an 89% agreement level, and in the hand washing evaluation, eight subjects had 100% agreement. One finding of this pilot study by Russian et al. (2008) suggested that inter-rater reliability was strong for all three competencies evaluated. This pilot study, despite the limitations of small sample size and assessment tool type which prevented it from leading to generalizations, offered helpful suggestions for larger, more complex studies of inter-rater reliability.

In many allied health care educational institutions, practical skill evaluation determines whether a student has reached proficiency to deliver care. Emergency Medical Technicians (EMT) evaluators used checklists and evaluation tools to assess student performance. A landmark study by Snyder and Smit (1998) assessed if trained EMT evaluators scored the performance of students with less variation and more reliability than non-evaluators. This post-test only, nonequivalent group design used a sample of 104 licensed instructor-coordinators (76 trained and 22 not trained), which viewed two video recorded scenarios of student performance, one passing and one failing, performed by students using the Michigan basic EMT practical exam. To test the interactions and differences of evaluator agreement, Snyder and Smit used an ANOVA statistical test was used. Scenario one had an evaluator agreement score of 79.4 %, and scenario two had an evaluator score of 67.8%, not reaching the ideal 80%.

The researchers stated there was evaluator bias concerning which skills are considered important in assessment (Snyder & Smit, 1998). In addition, the authors reported they thought the Michigan EMT exam might have had a reliability problem, which led to the low levels of inter-rater agreement. "Evaluator training programs designed to establish satisfactory levels of inter-rater agreement are a critical component in the development of a reliable observational evaluation process" (Snyder & Smit, 1998, p.44). The authors concluded if the measuring tool is inadequate, instructor-coordinators, no matter their expertise and training, will not or do not have a high level of inter-rater reliability.

There are instances when multiple evaluators can assess a person differently even when the measuring tool is understood by all. When interpretation varies among

evaluators, results can be inconsistent. Consistency has been discussed previously as being necessary in inter-rater agreement, and calibration has the potential to improve consistency. The following study used a single measuring tool and determined the consistency of two groups of practitioners using it to assess the physical status of geriatric patients.

In health care, patients interact with many health professionals who provide care and evaluate progress. Hartigan and O'Mahony (2009) compared the inter-rater reliability of two groups of clinicians, nurses and doctors, in using the Barthel Index (BI) to assess functional change in the rehabilitation of stroke victims when examining older adults in a rehabilitation unit. They wanted to determine if errors of recording the BI were a reflection of poor training in administration, poor technique by the raters, subjective bias, or different interpretations in the items on the BI. There was no gold standard that measured the different functions on the BI; therefore, the researchers first compared two nurses' BI scores and compared them to two doctors' BI scores. The group with the higher inter-rater reliability would serve as the gold standard and the highest level of reliability (Hartigan & O'Mahony).

The sample consisted of 36 nurses and six doctors, and all were trained during a brief BI training session (Hartigan & O'Mahony, 2009). Sixty-five patients consented to participate in this study, and each was measured four times using the BI, by two different nurses and doctors within five days of admittance. A Cohen's kappa coefficient was used to compute the measure of agreement between the individual items on the BI. Results reported by Hartigan and O'Mahony showed that between nurses and doctors, only three out of the ten function items had a 0.61 kappa coefficient, where 1.0 is a perfect score.

The researchers then measured the results within the two groups and found the doctors' group had a wide variation in agreement (-5.1 to 6.2), while the nurses' group had a smaller variation (-4.293 to 4.601), thus indicating more reliability in applying the BI assessment. Since a change of more than 2 points in the total score of the BI reflects a change in a patient's functional ability, the researchers recommended that the future BI assessments be completed by nurses. It was observed that doctors recorded patients' self-reports, while nurses recorded direct observation. The researchers suggested all examiners should assess patients through direct observation during all calibration sessions (Hartigan & O'Mahony).

Many health professions, including oral health professions, are studying methods for achieving a high level of inter-rater reliability of faculty in the evaluation of the students' performance and competence. The studies discussed in this subsection of the chapter represent a sampling of studies related to other health professions. Even though the subjects they measured differ from that in dental hygiene studies, each had an assessment tool that they believed would measure the necessary competencies. Two studies indicated that when using a valid assessment tool, evaluators were reliable (Russian, et al, 2008; Hartigan & O'Mahony, 2009). One study found that even with state licensure examination criteria, evaluators had low inter-rater reliability (Snyder & Smit, 2009). Thus, the issue of faculty calibration remains unresolved, and solutions to gaining higher rater reliability are needed.

The resounding message from the researchers in the above literature review was that time, effort, and consistency is needed for calibration training to be successful in achieving the particular goals of dental, dental hygiene, and health care programs.

Lanning, et al. (2006) concluded an extension of training programs might improve assessment, along with extended time. With this extension, faculty could potentially understand the gold standard representation of its individualized program, analyzing errors so improvements may be made (Knight, 1997). More research is needed to find the proper amount of time and intervals to improve the validity and reliability of faculty members.

Summary of Literature Review

Clinical instructors should not use their own standards when evaluating student performance, but must practice the gold standards that were developed by the profession or the program in which they are teaching. Interpretation of these standards can be varied, so calibration of faculty should be practiced and studied. Calibration has the potential to accomplish high inter-rater and intra-rater reliability, which provides consistency along with inter-rater and intra-rater agreement. Calibration should not be conducted only when necessary, but it should be practiced as an ongoing part of faculty development.

Faculty calibration can be successful long term; however, studies have mixed results. Faculty calibration training for standardization in assessment of calculus deposits has not shown to achieve high inter-rater reliability. Suggestions for future studies made by researchers who were not successful in reaching high inter-rater agreements were to evaluate deposits in the oral cavity of humans, rather than typodonts with simulated calculus. This study addressed recommendations and assessed the impact of a faculty calibration program on percent agreement of faculty scores in the evaluation of calculus detection as well as inter-rater reliability.

Chapter III

Methodology

Overview of Study

Dental hygiene educators recognize student assessment is a vital part of the educational process for all students. Research studies assessing dental hygiene faculty calibration in calculus detection are limited and have not documented an improvement in inter-rater reliability of faculty's evaluation of students' performance. The purpose of this study was to evaluate the impact of dental hygiene faculty calibration training on inter-rater reliability of faculty evaluation during clinical evaluation of presence or absence of calculus. The following null hypothesis was investigated during the study:

1. There will be no significant difference in inter-rater reliability of faculty calculus detection scores as measured by Cohen's kappa following calibration training designed to enhance agreement of faculty members' assessment of calculus deposits based on the standard established by the Central Regional Dental Testing Service, readily detectible calculus.
2. There will be no significant difference in inter-rater reliability of faculty calculus detection scores as measured by Cohen's kappa scores following calibration training designed to enhance agreement of faculty members' assessment of calculus deposits based on the gold standard established for evaluation of adequacy of scaling and root planing, relative root smoothness, before and after faculty calibration training.

Research Design

This pilot study measured inter-rater reliability of dental hygiene faculty at the Halifax Community College (HCC) Dental Hygiene Program before and after calibration training. The study followed a standard pre-test, post-test single one-group design. Four patients with varying degrees of calculus deposits served as patients based on Garland and Newell's primary recommendation that calculus detection evaluation be completed on patients rather than typodonts. An answer key was developed one week prior to the pre-test. The pre-test occurred prior to the faculty calibration training. A calibration workshop was conducted one week after the pre-test with the post-test following immediately after.

Research Context

The study was conducted at Halifax Community College (HCC) in Weldon, North Carolina. The dental hygiene program at Halifax Community College has a clinical facility with sixteen dental operatories equipped with ADEC units. Students have been providing dental hygiene care for the community under faculty supervision in the clinical facility since 2004 (V.V. High, personal communication, July 31, 2013).

This site was selected because the HCC clinical faculty members perform similar assessments with students on the detection of calculus on patients as well as on typodonts. HCC's clinical facility has the necessary clinical equipment to provide safe care for patients. The HCC dental hygiene program provides entry-level education and is accredited by the American Dental Association (ADA). HCC also provides pre-college education, technical training, short-term training, a four-year college transfer option, and two year degrees (www.halifaxcc.edu, 2013).

Research Participants

Sample description. The convenience sample consisted of clinical faculty members at HCC (N=) in Weldon, North Carolina. The sample size of six faculty members was reduced to five when one subject erroneously recorded calculus during the pre-test and was deemed ineligible to include in data analysis. Data collection was limited to the dental hygienist faculty scores (N=4). HCC employs three full-time instructors, five part-time instructors. The participants were invited via a letter (Appendix B) including the study's research questions, expectations of the participants, confidentiality statements, and time frame of participation. The convenience sample of participants who agreed to participate and provided informed consent included four dental hygienists and two dentists. Data collection was limited to the dental hygienist faculty scores (N=4) to control for extraneous variables associated with differences in education and experience. The study included two full-time and two part-time faculty members. One of the participants had two years of clinical teaching experience while the remaining three faculty members had eight to ten years of clinical teaching experience. All subjects had over twenty years of clinical practice experience as licensed dental hygienists. The inclusion criteria for the sample included the following:

- The participant has a dental or dental hygiene license from the state of North Carolina.
- The participant is either a full-or part-time clinical faculty member from HCC dental hygiene program.

Human subjects' protection.

An application for a certificate of exemption for the study was presented and exempted by the institutional review board, the Human Subjects Committee, at Idaho State University on June 2, 2014. HCC completed an administrative review and granted approval for the project to be conducted at HCC. According to Human Subjects: a Manual and Guide for Investigators (Human Subjects Committee, Idaho State University, 2013), this study qualified for an exemption due to the following:

> Normal Educational Practices and Settings. This exemption includes research conducted in established or commonly accepted educational settings, involving normal educational practices. Examples are research on regular and special educational strategies, or research on the effectiveness of or the comparison among instructional techniques, curricula, or class management methods. (p.7)

All HCC dental hygiene faculty members received an invitation to participate in the study, which described the purpose of the study, study procedures, potential risks and discomforts during the study, anticipated benefits to patients and to society, and alternatives to participation (Appendix B). All participants were able to withdraw from the study at any point for any reason. There was no monetary compensation, legal or monetary penalties for participants before, during, or after this study. Confidentiality was protected during testing. Discussion of calibration results by the faculty and researcher may occur during the calibration training. All study participants received a four-digit code assigned to them to ensure confidentiality of the subject's scores during data processing, analysis, and storage. The list of participant's names was destroyed as soon as codes were assigned.

Study data forms were maintained separately from the HCC's patient files in the clinic records department. This department is monitored during clinical hours and kept locked and secured during non-clinical hours. Clinical records are protected by the Health Information Portability and Accountability Act of 1996 (HIPAA). All data were collected exclusively by the primary researcher, placed in an envelope immediately after testing, and stored securely in a locked file cabinet and computer file by the researcher. Upon completion of the study, these records were submitted to the major thesis advisor who will assure safe and secure storage for a period of seven years in the Department of Dental Hygiene at Idaho State University.

The risk involved for the participants was a possible short-term loss of productivity, as the study was conducted during class preparation and office hours. The benefits to the faculty participants were training to enhance the understanding of intra-rater reliability and calibration and the potential application of knowledge and skills acquired to enhance clinical teaching and evaluation after participating in calibration training. All research participants were protected by procedures outlined in Human Subjects: a Manual and Guide for Investigations (Human Subjects Committee, Idaho State University, 2013).

Patients were protected by the Health Insurance Portability and Accountability Act of 1996 (HIPPA) and the HCC's clinical standards of care and emergency procedures that are followed daily at the clinic (Appendix C). Although the patients were not subjects, their protection in the training process was the responsibility of the faculty calibration training program coordinator.

Data Collection

Procedure. Patients were recruited from a pool of personal contacts known by the principal investigator (PI). An initial telephone inquiry regarding interest was followed by a letter formally inviting their participation and explaining all procedures (Appendix B). An orientation for the patients was conducted by the PI in the reception area of HCC's dental hygiene clinic (Appendix D). Patients received a patient consent form once they agreed to participate (Appendix E). Patients were informed of the commitment dates and specific times of the appointments consisting of one appointment for establishing the answer key for calculus scoring based on the gold standard, and two appointments for the faculty's pre- and post-test scoring. The procedures of the study were explained, along with the potential risks and the possibility of discomforts. All patients had the right to discontinue participation as patients for the faculty calibration at any point with no explanation. Dental hygiene therapies were offered at no cost to patients who completed the process at the conclusion of the study if they wished. Patients were given the opportunity to receive dental hygiene therapy as compensation at no cost to the patient or HCC's institution. Treatment options included care provided by HCC's dental hygiene students, supervised by faculty, or provided by the PI during HCC'S normal clinic hours. Five patients met the following inclusion criteria for the study based on the CRDTS patient acceptability and health history requirements, but only four patients were utilized due to one patient missing the pre-test session.

- Patients were at least eighteen years of age with a minimum of 24 teeth and did not require local anesthesia for the study.

- Each patient had a full mouth series of radiographs exposed within 12 months of the study.
- The patients' last dental prophylaxis was not completed within the last six months of the study starting date.
- Blood pressure was no higher than 160/100.
- The patient was willing to attend three sessions during the study.

After the patients agreed, signed the consent form, and accepted per the study's criteria, they were scheduled for the first session where Evaluator 1 and 2 established inter-rater reliability with calculus detection. Evaluator 1 is PI and Evaluator 2 is a licensed hygienist who has practiced dental hygiene for twenty two years and has a Master in Dental Hygiene degree and was a student instructor for two semesters.

The four patients had a minimum of 24 teeth and a maximum of 28 teeth present, excluding third molars. All teeth have four tooth surfaces: M=mesial; D=distal; F=facial; and L=lingual for a total of 96 to 112 surfaces. The four patients had varying amounts of calculus present. All patients had subgingival calculus, calculus present under the gingival margin, and/or supragingival calculus, calculus present above the gingival margin. Calculus may have been present in different amounts including light calculus, roughness, graininess, spicules, or as defined by CRDTS (2013) as "explorer-detectable, subgingival calculus which is distinctive, obvious, and can be easily detected with a #11/12 explorer as it passes over the calculus" (p.20). No patients had a dental prophylaxis within the last six months of the beginning of the study. These patients' conditions reflected typical experiences students encounter during their education. If readily detectable calculus as defined by CRDTS was detected on a particular site, the

corresponding box was marked with a “C” and if lighter calculus or root roughness indicating the presence of calculus was marked with an “R”, each representing a 1 point value.

During the sessions scheduled to establish faculty inter-rater reliability, patients waited in HCC's dental hygiene reception area. Coffee, water, soda, and light snacks were available. If a patient wished to brush their teeth prior to being seen, a toothbrush and toothpaste was provided by the researcher.

At the faculty calibration orientation (Appendix D), all study participants received a consent and confidentiality form before the pre-tests were conducted. The researcher explained the content of the forms, all potential risks and benefits, answered any questions, and provided time to complete the forms. All subjects were informed that confidentiality of calculus scores would be maintained. The researcher explained the data collection form and answered any questions. This procedure took approximately thirty minutes to complete.

The four patients were presented for the pre-test exploring and scoring of calculus by all participants. Before the pre-test was conducted, all patients had a medical history review and vitals were taken by dental hygiene students and participants, with radiographs displayed on view boxes. The time spent was approximately twenty minutes (Appendix D). Calculus detection score sheets (Appendix A) with two randomly assigned quadrants identified were available for recording calculus at each of the dental units where the patients were seated. The selected teeth were present in all the four patients. The participants used new 11/12 explorers and mirrors to explore each tooth surface in the designated half mouth on each patient. Study participants noted if calculus was

present on a tooth surface, and recorded it on the form in the corresponding area. If no detectible calculus was present, the corresponding area was left blank. If readily detectable calculus as defined by CRDTS was detected on a particular site, the corresponding box was marked with a "C" and if lighter calculus or root roughness indicating the presence of calculus was identified, it was marked with an "R," each representing a 1 point value. There was an area where participants could make notations about location, size, or type of calculus. Two quadrants of each patient's dentition (6 teeth on the right and 10 teeth on the left side were randomly assigned by the trainer by flipping a coin) was explored by each study participant. Assignments were made alphabetically using the subject's last names. The pre-test was completed approximately 20 to 30 minutes per patient for a total of 2.5 hours (Appendix F).

On session three, one week later, subjects received training. The calibration session lasted 75 minutes (Appendix D). The trainer discussed the following for the calibration training:

- PowerPoint presentation on the definition and purpose of calibration, and explorer instrumentation techniques.
- Reviewed CRDTS criteria for calculus detection (Appendix F).
- A video by Neild-Gehrig (2013) demonstrating correct technique of the ODU 11/12 explorer.
- Scenarios were discussed by the trainer and the participants that depicted the most challenging instances subjects have encountered during their experiences with calculus detection such as:

1. Anatomical landmarks of a tooth which can be mistaken as calculus (e.g. CEJ, root concavities).
2. Existing restoration margins.
3. Root proximity and roughness.
4. Malaligned teeth.

- The use of UMKC models with explorer-detectible calculus and new ODU 11/12 explorers to practice detection and recording of calculus.

A post-test was conducted after the calibration training session with the participants using the same four patients from the pre-test. Prior to the post-test, all patients had their medical histories updated and vital signs taken by dental hygiene students and subjects while being supervised by the PI, and radiographs displayed on viewers. The time required for patient preparation was approximately twenty minutes (Appendix D). Using a separate but identical CRDTS calculus detection form as the form used for the pre-test (Appendix A), participants began exploring each patients' assigned dentition. If calculus was identified on a particular tooth surface, the corresponding area was marked on the form. If no detectible calculus was present, the corresponding area was left blank. If readily detectable calculus as defined by CRDTS was detected on a particular site, the corresponding box was marked with a "C" and if lighter calculus or root roughness indicating the presence of calculus was detected it was marked with an "R", each representing a 1 point value. There was an area where participants could make notations about location, size, or type of calculus. The post-test was completed in two hours (Appendix D). The researcher then reviewed the pre-test and post-test CRDTS calculus detection forms and input the data into an Excel spreadsheet for statistical

analysis. The patients were invited to return for the researcher to disclose findings and to receive preventive dental hygiene care at their convenience.

Instruments.

Calculus detection form. CRDTS uses a Dental Hygiene "Full Mouth Patient Screening" worksheet (Appendix A) that has boxes representing specific teeth and tooth surfaces of mesial (M), distal (D), facial (F), and lingual (L) listed beside each tooth number. The corresponding box was marked with a "C" if light calculus or root roughness is detected indicating the presence of calculus, the corresponding box was marked with an "R", each representing a1 point value. If no calculus was detected, the corresponding area was left blank, corresponding to 0 points. The CRDTS calculus detection form has established validity and reliability, based upon analysis completed by Dr. Thomas M. Haladyna, Professor Emeritus, Arizona State University in 2011 (CRDTS).

The calculus detection form developed by CRDTS has 32 boxes, representing 32 teeth, including third molars, a normal adult dentition. Each box has four divided areas that represent the following tooth surfaces present on all teeth: M=mesial; D=distal; F=facial; and L=lingual. The CRDTS Dental Hygiene "Full Mouth Patient Screening" worksheet was used for all three sessions including (a) establishing the answer key, (b) pre-test, and (c) post-test.

Reliability and validity

Data were collected using a process based upon and established procedure used by the CRDTS for detecting calculus (CRDTS, 2013). CRDTS is a regional testing service made up of 16 state boards that test dental and dental hygiene candidates for

licensure. The data collection protocol for this study was based on the CRDTS procedures and a similar study previously conducted by Garland and Newell (2009).

Limitations

This study had possible limitations and threats to internal and external validity. To help prevent diffusion of treatment, where the faculty groups discuss outcomes with each other, subjects were asked to respect confidentiality by refraining from discussing calculus scores during the testing. Each participant also signed a confidentiality form.

Findings cannot be generalized to other groups of dental hygiene faculty due to the use of a small, nonrandomized, convenience sample. The study will potentially serve as a pilot program for others interested in faculty calibration for calculus detection.

Proposed Statistical Analysis

Inter-rater reliability was determined by comparing reliability of pre-test and post-test calculus detection scores of the four faculty participants. A Cohen's kappa was used to assess inter-rater reliability.

Chapter Three Summary

The purpose of this study was to test whether faculty calibration training improves inter-rater reliability during calculus detection using two accepted standards. This study was designed to build upon a small but growing body of research on the inter-rater agreement of dental and dental hygiene faculty and the effects of calibration. This pre-test post-test design measured the impact of calibration training on inter-rater reliability. The findings of previous faculty calibration studies have been mixed indicating that calibration training results are positive, negative, and have no impact.

References

Albino, J.E., Young, S.K., Neumann, L.M., Kramer, G.A., Andrieu, S.C., Henson, L., Horn, B., & Hendricson, W.D., (2008). Assessing dental students' competence: best practice recommendations in the performance assessment literature and investigation of current practices in predoctoral dental education, *J Dent Educ., 72*(12), 1405-1435.

American Dental Hygienists' Association, (2007). National dental hygiene agenda, Retrieved on Monday, November 30, 2009 at http://adha.org/research/nra.htm

Assaf, A.V., Meneghim, C., Zanin, L., Tengan, C., & Pereira, A.C., (2006). Effects of different diagnostic thresholds on dental caries calibration- a 12 month evaluation, *Community Dental Oral Epidemiol*, *34*, 213-219. doi: 10.1111/j1600-0528.2006.00278.x

Central Regional Dental Testing Service, (2014). *CRDTS dental hygiene full mouth patient screening worksheet*. Retrieved on March 3, 2014 at http://www.crdts.org/uploads/CRDTS%202014%20Full%20Mouth%20%20Patient%20Screening%20Worksheet.pdf

Central Regional Dental Testing Service, (2014). *CRDTS dental hygiene "treatment selection" worksheet*. Retrieved on March 3, 2014 http://www.crdts.org/uploads/CRDTS%202014%20Treatment%20Selection%20%20Worksheet%20w%20Example%20on%202nd%20page.pdf

Christie, C., Bowen, D., & Paarmann, C., (2007). Effectiveness of faculty training to enhance clinical evaluation of student competence in ethical reasoning and professionalism, *J Dent Educ., 71*(8), 1048-1057.

Commission on Dental Accreditation, 2013, Retrieved on July 27, 2013 at http://www.ada.org/117.aspx

Dahlström, L., Keeling, S.D., Friction, J.R., Hilsenbeck, S.G., Clark, G.M., & Rugh, J.D., (1994). Evaluation of a training program intended to calibrate examiners of tempomandibular disorders, ACTA Odontol Scand, *52*, 250-254.

Flack, V.F., Atchison, K.A., Hewlett, E.R., & White, S.C., (1995). Relationships between clinician variability and radiographic guidelines. *J Dent Educ,75*(2), 775-782.

Gansky, S.A., Pritchard, H., Kahl, E., Mendoza, D., Bird, W., Miller, A.J., & Graham, D., (2004). Reliability and validity of a manual dexterity test to predict preclinical grades, *J Dent Educ., 68*(9), 985-994.

Garland, K.V., & Newell, K.J., (2009). Dental hygiene faculty calibration in the evaluation of calculus detection, *J Dent Educ., 73*(3), 383-389.

Haj-Ali, R., & Feil, P., (2006). Rater reliability: short-and long-term effects of calibration training, *J Dent Educ., 70*(4), 428-433.

Haladyna, T.M., (2011). An evaluation of the central regional dental testing service's national dental hygiene examination. Retrieved on March 3, 2014 at http://www.crdts.org/uploads/2010%20CRDTS%20DH%20Technical%20Report.pdf

Hartigan, I., & O'Mahony, D., (2009). The Barthel Index: comparing inter-rater reliability between nurses and doctors in an older adult rehabilitation unit, *Science Direct*, Retrieved on Monday, January 25, 2010 at www.sciencedirect.com

Idaho State University Office of Research (2012a). *Human subjects manual for investigator, 2008*. Retrieved November 20, 2013, at http://www.isu.edu/research/hsc.shtml

Jacks, M.E., Blue, C., Murphy, D., (2008). Short and long-term effects of training on dental hygiene faculty member's capacity to write soap notes, *J Dent Educ.,72*(6), 719-724.

Knight, G.W., (1997). Toward faculty calibration, *J Dent Educ,, 61*(12), 941-946.

Lanning, S.K., Best, A.M., Temple, H.J., Richards, P.S., Carey, A., & McCauley, L.K., (2006). Accuracy and consistency of radiographic interpretation among clinical instructors in conjunction with a training program, *J Dent Educ.*, 70(5), 545-556.

Lanning, S.K., Pelok, S.D., Williams, B.C., Richards, P.S., Sarament, D.P., Oh, T., & McCauley, L.K., (2005). Variation in periodontal diagnosis and treatment planning among clinical instructors, *J Dent Educ., 69*(3), 325-337.

Merriam-Webster Dictionary Online, (2009). Retrieved on December 12, 2009 at www.merriam-webster.com/dictionary/intra

Newman, M.G., Takei, H.H., Klokkevold, P.R., & Carranza, F., (2012). *Carranza's clinical periodontology*, ed. 11, St. Louis, Saunders.

Partido, B.B., Jones, A.A., English, D.L., Nguyen, C.A., Jacks, M.E. (2015). Calculus detection calibration among dental hygiene faculty utilizing dental endoscopy: a pilot study. *J Dent Educ., 90*(2): 124-32.

Pippin, D.J., & Feil, P. (1992). Interrater agreement on subgingival calculus detection following scaling. *J Dent Educ., 56*(5): 322-26.

Russian, C.J., Harkins, L., Marshall, S.G., Engelhardt, M., & Shamrock, B.A., (2008). Inter-rater reliability among respiratory care clinical instructors: pilot study, *Respiratory Care Education Annual*, *17*, 7-12.

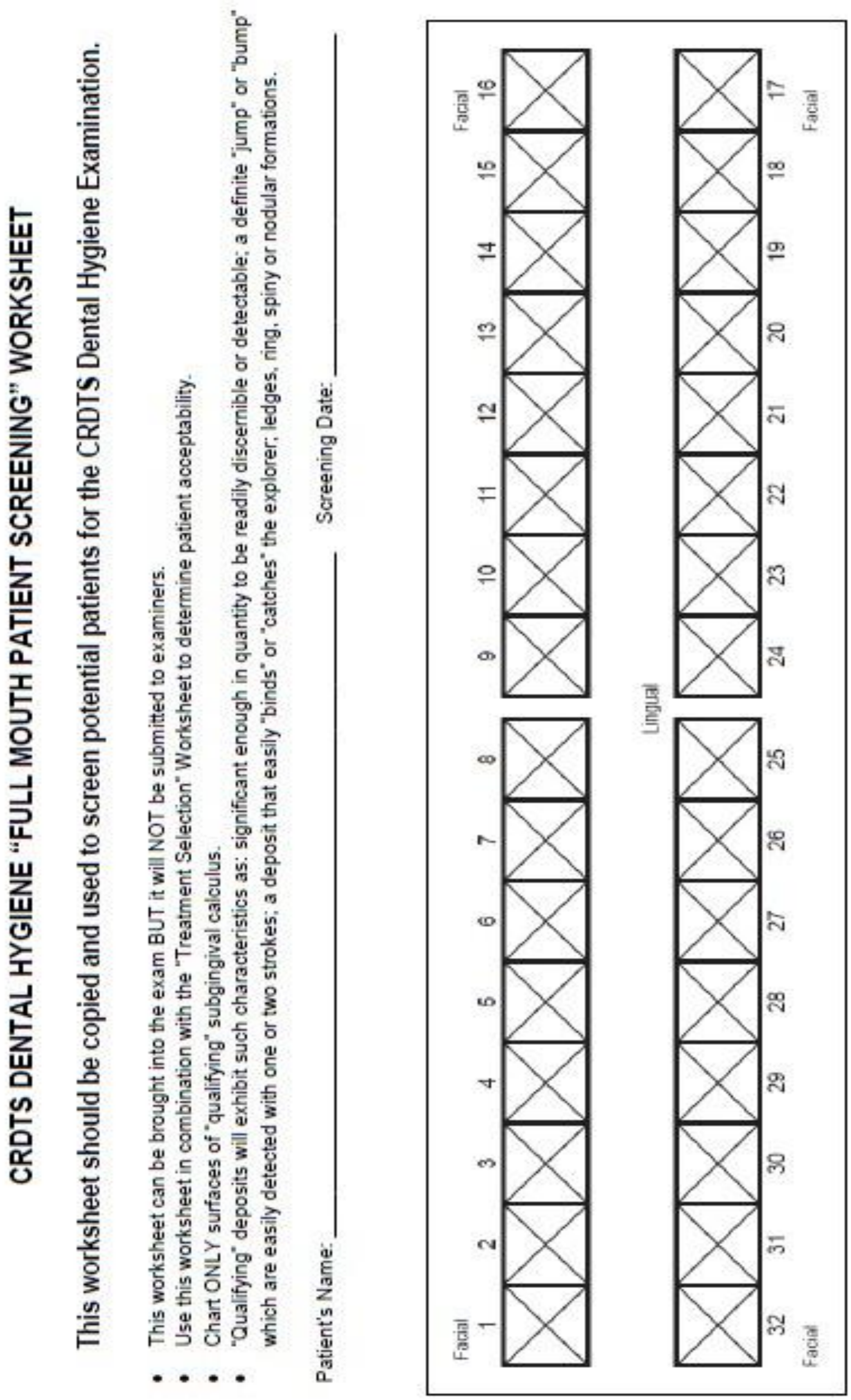
Snyder, W., & Smit, S., (1998). Evaluating the evaluators: Interrater reliability on EMT licensing examinations, *Prehospital Emergency Care*, *2*(1), 37-46.

Taleghani, M., Solomon, E.S., & Wathen, W.F., (2004). Non-graded clinical evaluation of dental students in a competency-based education program, *J Dent Educ.*, *68*(6), 644-655.

Vogt, W. P. (2005). *Dictionary of statistics & methodology: A nontechnical guide for the social sciences*, (3rd ed.). London, United Kingdom: Sage.

Wallace, J. S., & Infante, T. D. (2008). Outcomes assessment of dental hygiene clinical teaching workshops, *J Dent Educ, 72*(10), 1169-1176.

Appendix A: Central Regional Dental Testing Service Calculus Form



CRDTS DENTAL HYGIENE "FULL MOUTH PATIENT SCREENING" WORKSHEET

This worksheet should be copied and used to screen potential patients for the CRDTS Dental Hygiene Examination.

- This worksheet can be brought into the exam BUT it will NOT be submitted to examiners.
- Use this worksheet in combination with the "Treatment Selection" Worksheet to determine patient acceptability.
- Chart ONLY surfaces of "qualifying" subgingival calculus.
- "Qualifying" deposits will exhibit such characteristics as: significant enough in quantity to be readily discernible or detectable; a definite "jump" or "bump" which are easily detected with one or two strokes; a deposit that easily "binds" or "catches" the explorer; ledges, ring, spiny or nodular formations.

Patient's Name: ______________________ Screening Date: ______________________

Appendix B: Subjects' Invitation Letter

May 2, 2014

Dear Halifax Community College Dental Hygiene Faculty,

I am pleased to offer an invitation to participate in a faculty calibration program titled "Dental Hygiene Faculty Calibration to Enhance Inter-Rater Reliability with Calculus Detection" conducted by Lisa J. Santiago, RDH, B.S., a master's thesis candidate at Idaho State University. This program will be conducted at Halifax Community College in the dental hygiene clinic. As a portion of my thesis work, I will be assessing the outcomes of this project to answer the following questions:

1. Does percent agreement between faculty scores and the answer key scores representing (a) the gold standard established by the Central Regional Dental Testing Service, readily detectible calculus, and (b) the gold standard accepted for evaluation of adequacy of scaling and root planning, relative root smoothness, and to assess change after faculty calibration training in calculus detection?
2. Is there a change in inter-rater reliability of faculty calculus detection scores as measured by Cohen's kappa scores following calibration training designed to enhance agreement of faculty members' assessment of calculus deposits based on two accepted gold standards?

Your participation in the program should take approximately three hours and thirty minutes. It will involve a pre-test and post-test of calculus detection of five patients with varying amounts of calculus. A workshop has been designed to take place between the pre-test and post-test to provide calibration training. If you do decide to participate, it will cost only your time. You have the right to refuse to participate at any time with no

repercussions. If you happen to have any questions during the study, please feel free to ask me at any time. All data will be kept confidential. The names of participants will not be disclosed. Results will be reported and disseminated in group form only. This pilot program and its results may help others in future dental hygiene faculty calibration training or studies. I appreciate your consideration in participating, and I look forward to working with you. Please feel free to contact me at any time if you have questions or comments (lsantiago730@halifaxcc.edu/252-538-4338).

Appendix C: Halifax Community College Standards of Care

Standards of Care

Halifax Community College

Dental Hygiene Program

The HCC Dental Hygiene Program has adopted the following policies and procedures concerning the treatment services provided to all patients in the HCC Dental Hygiene Facility:

Assuring Comprehensive, Quality, Client-Centered Care

- During evaluation and grading of the dental hygiene care plan each faculty member will verify that the care plan:
    1. Includes all educational and clinical services necessary to meet the comprehensive/individual treatment needs of the client.
    2. Includes all referrals for medical or dental care/evaluation.
    3. Is client-centered.
    4. Is developed based on scientific evidence.
- At the end of each clinic session a faculty member will sign the treatment record with his/her last name to signify that the he/she has:
    1. Assisted the student in immediately modifying the care plan to correct any deficits found in the care plan.
    2. Verified that all treatment planned for that date is complete.
    3. Noted any deficits in client services and the measures to correct that deficit in the treatment record. The dental hygiene care plan must be amended to reflect the change.

- Before assigning a grade as an end-product evaluation, the faculty member will verify that only completed, quality treatment has been delivered.
- As a faculty member works with a student he/she will review the previous treatment record entries to make sure quality, client-centered, comprehensive treatment is delivered and all treatment completed.

Referring Clients for Medical/Dental Consultations

- Each client's medical/dental history is reviewed by the student operator with a faculty member at the beginning of every appointment.
- If the student and faculty member have no medical or dental concerns, relating to that client, the faculty member signs the "Permission to Proceed" (PTP) line on the clinical evaluation form and the student may begin to treat the client.
- Any client with questionable medical/dental concerns will be referred for a medical consult and the client will be treated once they have a written release from their physician/dentist. The referral will be signed by the student, client, and dentist.
- Any client with a highly questionable medical/dental status will have his/her medical/dental history reviewed by the student, faculty, and dentist. The client will be expected to clarify, when possible, any questions concerning the history. These clients may be denied dental care in the HCC Dental Hygiene Clinic due to advanced medical/dental complications. The consulting dentist will make the final determination in accepting a client for care in the clinic.
- Any client may be refused treatment in the HCC Dental Hygiene Clinic due to the presence of:

1. Progressive disease state that makes it unsafe to deliver routing dental hygiene care.
2. Serious communicable diseases that creates an unsafe environment for students, faculty, staff or clients.
3. Advanced dental disease that will not respond to non-surgical periodontal treatment.

- A client will be referred for a dental consult when:
    1. Periodontal disease does not respond to non-surgical periodontal therapy.
    2. Treatment is needed for dental concerns beyond the scope of a dental hygiene educational clinic.

Assuring Completed Care

- Letters will be sent to all clients who receive substandard care. These clients will be informed of the treatment shortfall and offered a plan to correct the deficit. A copy of the letter, signed by the Dental Hygiene Department Head, will be kept in the client's file. The client will be given 30 days to respond to the letter. For example, if x-rays are of less than diagnostic quality, the client will be sent a letter and offered an appointment to correct the deficits in radiographic series.
- Clients who receive incomplete care because they do not return to the clinic will be contacted by the Clinic Manager with two attempts by phone. The Clinic Manager will document the contact in the treatment record. After making the phone attempts with no success, the Clinic Manager will send a follow-up letter, addressed to the client, giving the client 30 days to respond to the letter or schedule an appointment to complete care. The Clinic Manager will transfer the letter to HCC letterhead

stationery and sign the letter. A copy of the letter is maintained in the client's folder. The Clinic Manager will document the mailing of the letter in the client's treatment record. If the client does not respond within the 30 days the client's record will be filed as "Inactive".

Appendix D: Sequence of Procedures

Sequence of Procedures

Session 1

Inter-Rater Agreement and Answer Key Development Session	Time
*Patients arrived at clinic. Patients checked in with PI. An orientation session began with PI discussing consent forms, expectations on study dates and times. A question/answer session was conducted.	30 minutes
*Examiner 1 and 2 sat one patient at a time in a dental operatory located in the clinical area. The medical history was reviewed and vital signs were taken. Radiographs were displayed. Patients were checked if they qualified for the calibration session using the inclusion and exclusion criteria established. A calculus detection sheet was placed at each patient station.	50 minutes (10 min. per pt.)
*Examiner 1 and 2 explored the patients' full dentition. Examiners recorded any areas of calculus detected on calculus detection key. If readily detectible calculus as defined by CRDTS is detected on a particular site, the corresponding box was marked with a "C" and if lighter calculus or root roughness indicating the presence of calculus was marked with a "R", each representing 1 point value. If no calculus is detected, the corresponding area was left blank, corresponding to 0 point.	2 hours (20-25 min. per pt.)
*Patients waited in waiting room as examiners determine if 100% score of reliability has been achieved.	10 minutes
*If 100% score was achieved, patients were dismissed.	
*If 100% score was not been achieved: Evaluator 1 and 2 re-explored those surfaces of disagreement. Evaluator 1 and 2 recorded detectable calculus on calculus detection key.	2 hours (20-25 min. per pt.)
*Patients waited in waiting room as PI determines if 100% percent agreement score was achieved.	10 minutes
TOTAL	210min/340min

Session 2

Faculty Participants' Pre-test Calculus Detection and Scoring Session	Time
*Study participants meet for orientation. PI discussed consent forms, expectations for study, dates, and times. A question/answer session did occur.	30 minutes
*Patients arrived at clinic. Patients checked in with PI who answered any questions they had.	15 minutes
*PI escorted patients into clinic area. Review of medical history and vitals was taken by dental hygiene students under supervision of the PI. Radiographs were displayed. A calculus detection sheet was placed at every patient station.	60 minutes (20 min. per pt.)
*Study participants explored two designated quadrants assigned for each patient with new 11/12 explorers. If readily detectible calculus as defined by CRDTS is detected on a particular site, the corresponding box was marked with a "C" and if lighter calculus or root roughness indicating the presence of calculus was marked with a "R", each representing 1 point value. If no calculus is detected, the corresponding area was left blank, corresponding to 0 point. Participants made notations if needed. Strict clinical protocol was followed throughout the patient examination sessions.	2.5 hours (20-30 min. per pt.)
* Patients had the opportunity to ask PI any questions. Patients were dismissed.	10 minutes
* Study participants were dismissed.	10 minutes
*PI and dental hygiene students broke down and disinfected dental units. All 11/12 explorers and dental mirrors were sterilized and stored.	20 minutes
TOTAL	5 hrs

Session 3

Faculty Participants' Post-test Calculus Detection and Scoring Session	Time
*Study participants met for calibration training.	60-90 minutes
*Patients arrived at clinic. Patients were checked in with PI. A question/answer session was done.	15 minutes
*PI escorted patients back into clinic area. Review of medical history and vitals were taken by dental hygiene students under supervision of the PI. Radiographs were displayed. A calculus detection sheet was placed at every patient station.	60 minutes (20 min. per pt.)
*Study participants explored designated quadrants assigned to patient with 11/12 explorers. If readily detectible calculus as defined by CRDTS was detected on a particular site, the corresponding box was marked with a "C" and if lighter calculus or root roughness indicating the presence of calculus was marked with a "R", each representing 1 point value. If no calculus was detected, the corresponding area was left blank, corresponding to 0 point. Participants made notations if needed.	2-2.5 hours (20-30 min. per pt.)
* Patients had the opportunity to ask PI any questions. Patients were dismissed.	10 minutes
* Study participants were dismissed for lunch.	10 minutes
*PI and dental hygiene students broke down and disinfect dental units. All 11/12 explorers and dental mirrors were sterilized and stored.	20 minutes
*PI compared pre- and post-test scores. PI determined if 80% score of reliability has been achieved.	60 minutes
* Study participants returned. PI shared findings.	10 minutes
*PI and study participants discussed findings and concluded test.	20 minutes
TOTAL	7.25 hours

Appendix E: Patient Consent Form

CLINIC INFORMATION/PATIENT CONSENT FORM

You have been asked to participate in this faculty development project in which licensed dental hygiene faculty members are participating in training and calibration for calculus detection. The calibration is intended to enhance consistency of student and patient evaluation. These assessment procedures are normal dental hygiene procedures used in the clinic, although multiple evaluations will occur for training purposes. You will be asked to attend three separate visits. During your first visit, the faculty trainer, Lisa J. Santiago, a masters in dental hygiene degree candidate at Idaho State University and licensed dental hygiene faculty member at Halifax Community College, and a former dental hygiene faculty member who is also licensed, will assess for the presence of calculus (tarter) deposits on the teeth in one-half of your mouth using a dental hygiene instrument. The assessments should not be painful but may cause some temporary sensitivity of the gums. The morning session should take 2-2.5 hours. There will be drinks such as bottled water, juice, and soda available along with light snacks. Before dismissal, the evaluators will ask you to wait 15 minutes to tally the scores. It is possible they may need to reassess calculus deposits.

On session two, licensed dental hygiene faculty members will assess for calculus similarly to the researcher. Before the session, there will be time for any questions to be answered. The assessment should not be painful but may cause some temporary sensitivity of the gums. The morning session should take 2-2.5 hours.

On session three, one week later, licensed dental hygiene faculty members will assess for calculus similarly to the researcher. Before the session begins, time will be allowed for

questions to be answered. The assessment should not be painful but may cause some temporary sensitivity of the gums. The morning session should take 2-2.5 hours.

An incentive for participation in this study would be the preventive care being provided at no cost to you by either the faculty trainer or by dental hygiene students after the conclusion of the study.

Please realize that participation is voluntary. If you choose to decline to participate at any time during the faculty development study, there will be no negative consequences. You may continue receiving dental hygiene care at HCC and your relationship with HCC's dental hygiene faculty will remain positive.

Please read the following information carefully so that you will understand the condition under which patients will be treated in this clinic during this faculty development study. After you have had a chance to ask questions at the first session, you will be asked to please sign your name at the bottom of the reverse side of this page, indicating that you understand these conditions.

I understand the following

1. The treatment will be provided by a licensed dental hygienist and dentist.
2. The treatment will be limited to preventive treatment and is not intended to take the place of a dental examination by a dentist. It is recommended that you visit and have a dental examination twice a year, or as recommended by your dentist or dental hygienist.
3. While optimal dental treatment can be expected, the results of this preventive dental health care cannot be guaranteed.

4. This faculty development project will be conducted for three days, which should last approximately 3.5-5.5 hours on the first session and 3-3.5 hours for sessions two and three.

5. The researcher is required to obtain a medical and dental history of each patient before initiating services. Such information is confidential and considered essential for adequate dental hygiene care and for the purpose of this faculty development project.

6. All records are property of the college; however, radiographs may be sent to my private dentist upon request by him/her. Radiographs will be kept on file indefinitely.

7. You have the right to terminate participation in this faculty development project at any time with no penalty and no questions asked.

8. If the use of topical anesthesia is indicated, I consent to the administration of such as the clinical dental hygienist and dentist may deem advisable and proper.

9. I consent to the use of my radiographs (x-rays) or any part of my treatment record for dental, scientific or educational purposes and to professional observation of treatment for the purposes of advancing dental hygiene education.

10. For compensation for participation, I may receive dental hygiene therapy (dental cleanings) free of charge and may be performed by HCC's dental hygiene students or by the researcher, who is a licensed dental hygienist.

Having read the above, I verify that I understand the information contained herein, and I grant authority to Halifax Community College Dental Hygiene Program to perform those diagnostic and treatment procedures deemed necessary.

Signature

Patient:_____Date:_____/_____/____

Appendix F: Central Regional Dental Testing Service Criteria Form

CRDTS DENTAL HYGIENE "TREATMENT SELECTION" WORKSHEET

This worksheet should be copied and used to screen patients to determine if the patient selected meets the CRDTS criteria for patient acceptability.

Candidates should bring this completed Treatment Selection Worksheet with them to the exam. Candidates can also use this worksheet to prepare an Alternate Selection to use in case their Initial Treatment Selection is rejected.

Instructions for Completing the Treatment Selection Worksheet:

- Write the teeth numbers you want to submit in the corresponding tooth number boxes.
- Chart ONLY surfaces of qualifying subgingival calculus present in the Treatment Selection by filling in the appropriate bubble (M,F,D,L)
- If qualifying subgingival calculus is found on the line angle of a tooth, mark it on the nearest interproximal surface. (ie: A surface of qualifying subgingival calculus on the mesial-facial line angle of a tooth should be marked on the mesial surface.)
- If a tooth is rotated, the surfaces of subgingival calculus should be charted as the tooth appears in the mouth.
- ALL criteria must be met or the Treatment Selection will be deemed Unacceptable.

CHART ONLY QUALIFYING CALCULUS

Qualifying deposits will exhibit such characteristics as:
- significant enough in quantity to be readily discernible or detectable;
- a definite "jump" or "bump" which is easily detected with one or two strokes;
- a deposit that easily "binds" or "catches" the explorer;
- ledges or ring formations;
- spiny or nodular formations.

TREATMENT SELECTION CRITERIA CHECKLIST

Does your Treatment Selection meet ALL this criteria?

Teeth:

___At least 6 teeth?
___If submitted, no more than 3 anterior teeth?
___If submitted, 3rd molars are fully erupted?

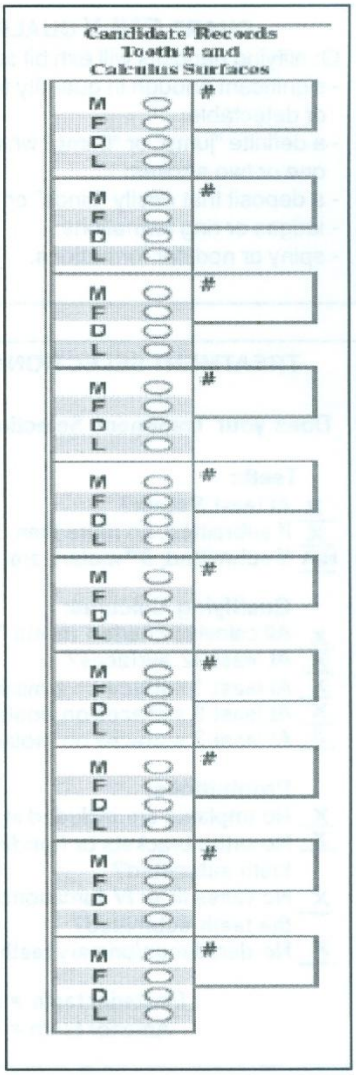
Qualifying Calculus:

___All calculus charted meets "qualifying" definition?
___At least 12 surfaces?
___At least 1 surface on a minimum of 6 teeth?
___At least 8 surfaces on posterior teeth?
___At least 3 surfaces on molar(s)?

Prohibitions:

___No implants are included in the teeth submitted?
___No ortho brackets or bonded retainers are in the teeth submitted?
___No Class III or IV furcations, mobility or disease are in the teeth submitted?
___No deciduous/primary teeth are in the teeth submitted?

Posterior teeth = molars and premolars
Anterior teeth = canines and incisors

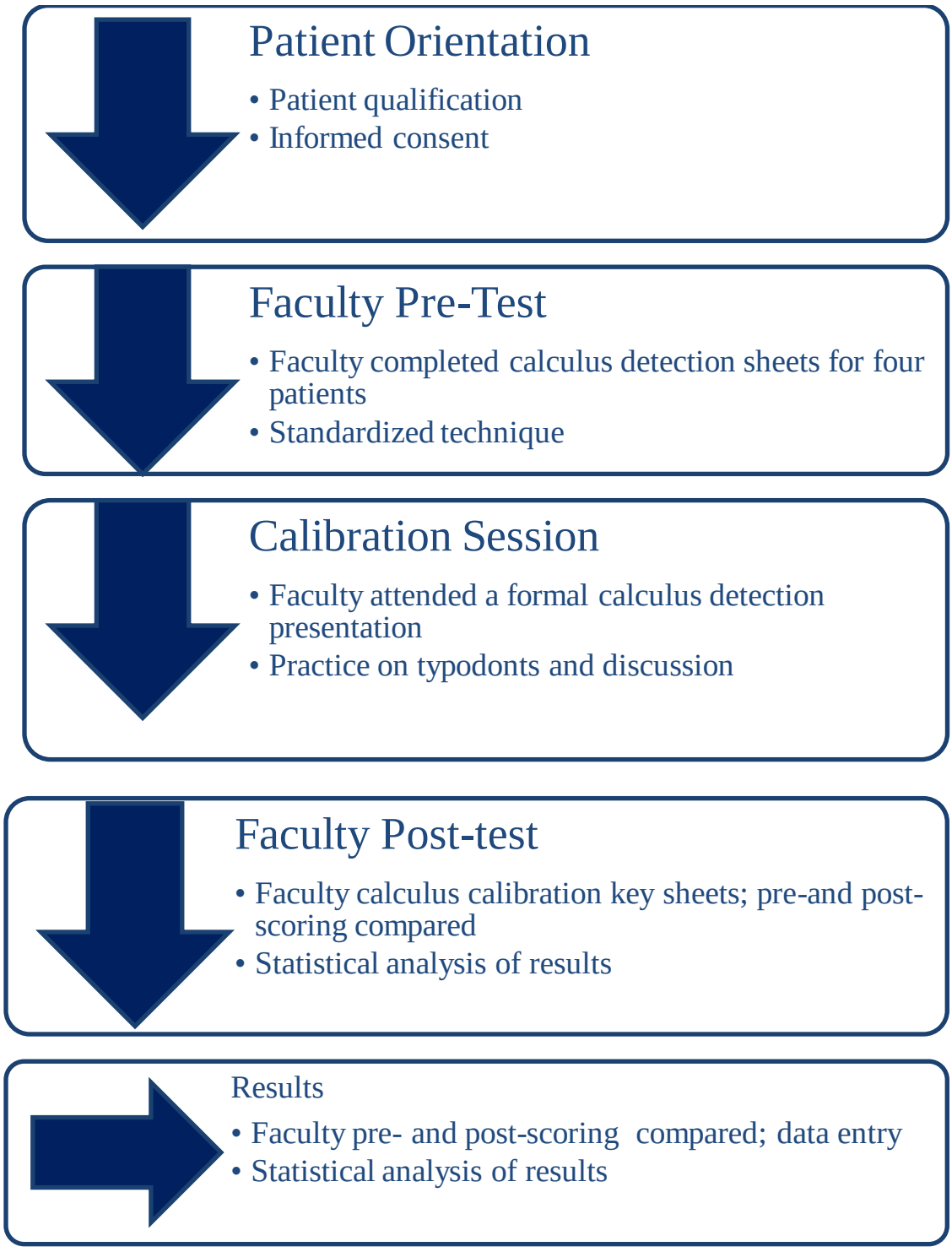


Appendix G: Table 1

Table 1. Kappa Scores for Testing Dental Hygiene Faculty Inter-rater Reliability

Standard Used for Answer Key	Pre/Post Test	Kappa Statistic	Standard Error	p value
1 CRDTS	Pre	0.535	0.054	0.001
	Post	0.622	0.622	0.001
2 Root Planing	Pre	0.206	0.074	0.001
	Post	0.221	0.066	0.001

Appendix H: Figure 1. Faculty Calculus Detection Calibration Study Protocol



Appendix I: Journal of Dental Education Author Guidelines

INFORMATION FOR AUTHORS

The *Journal of Dental Education (JDE)* is a peer-reviewed monthly journal that publishes a wide variety of educational and scientific research in dental, allied dental and advanced dental education. Published continuously by the American Dental Education Association since 1936 and internationally recognized as the premier journal for academic dentistry, the *JDE* publishes articles on such topics as curriculum reform, education research methods, innovative educational and assessment methodologies, faculty development, community-based dental education, student recruitment and admissions, professional and educational ethics, dental education around the world and systematic reviews of educational interest. The *JDE* is one of the top scholarly journals publishing the most important work in oral health education today; it celebrated its 75th anniversary in 2011.

I. Types of Manuscripts Considered and Requirements for Each

The Editor will consider the following types of manuscripts for publication:

Submissions for Peer Review:

Original Articles (see below for categories within this type)

Review Articles

Solicited or Pre-approved by the Editor:

Letters to the Editor (solicited or pre-approved by the Editor)

Guest Editorials (solicited by the Editor)

Perspectives (pre-approved by the Editor)

Brief Communications (pre-approved by the Editor)

Point/Counterpoint (solicited by the Editor)

Special Reports:

Miscellaneous (submitted by ADEA staff)

Submissions for Peer Review

1. Original Articles

This type of article addresses subject matter in the following categories:

a. Predoctoral Dental Education

b. Advanced Dental Education

c. Allied Dental Education

d. Interprofessional Education

e. Community-Based Dental Education

f. Global Dental Education—Manuscripts pertaining to global health education or issues pertinent to the global dental education community. (Not intended solely for submissions from international authors. International authors should submit manuscripts under pertinent topic areas provided in this section.)

g. Use of Technology in Dental Education

g. Use of Technology in Dental Education

h. Assessment

i. Faculty Issues/Development

j. Continuing Education

Original Articles should report the results of hypothesis-based research studies and may be either qualitative, quantitative or of a mixed methods nature. Manuscripts must address how the findings advance our understanding of the questions asked in the study and make a novel contribution to the literature. The limitations of the study should also be addressed. Small studies of local relevance/interest, limited to one class/course, or

small course/student-based surveys may not meet the criteria to be published as an Original Article.

Original Articles should be no more than 3,500 words, excluding the abstract, illustrations and references. A maximum of six figures and tables can be submitted (the figures can be multi-panel), and the number of references should not exceed 50 (unless the article is a systematic review).

Original Articles should have the following general organization (see "Document Preparation, Organization and Formatting" below for more detailed instructions):

Title: An informative and concise title limited to 15 words with no more than 150 characters.

Abstract: For research studies, a structured abstract of no more than 250 words should be submitted with the following subheads:

Purpose/Objectives: Briefly summarize the issue/problem being addressed.

Methods: Describe how the study was conducted.

Results: Describe the results.

Conclusion(s): Report what can be concluded based on the results, and note implications for dental education.

Abstracts for other types of manuscripts should be in paragraph form, with no subheads.

Introduction: Provide a succinct description of the study's background and significance with references to the appropriate published literature. Detailed literature review/discussion should be reserved for the discussion section. Include a short paragraph outlining the aims of the study.

Materials and Methods: A statement that the study has been approved or exempted from oversight by a committee that reviews, approves and monitors studies involving human subjects **MUST** be provided at the beginning of this section, along with the IRB protocol number.

In this section, provide descriptions of the study design, curriculum design, subjects, procedures and materials used, as well as a description of and rationale for the statistical analysis. If the design of the study is novel, enough detail should be given for other investigators to reproduce the study. References should be given to proprietary information.

Results: The results should be presented in a logical and systematic manner with appropriate reference to tables and figures. Tables and figures should be chosen to illustrate major themes/points without duplicating information available in the text.

Discussion: This section should focus on the main findings in the context of the aims of the study and the published literature. The authors should avoid an extensive review of the literature and focus instead on how the study's findings agree or disagree with the hypotheses addressed and what is known about the subject from other studies. A reflection on new information gained, new hypotheses and limitations of the study should be included, as well as guidance for future research.

Conclusion: The article should end with a short paragraph describing the conclusions derived from the findings and implications of the study for dental education.

Acknowledgments: The acknowledgments should report all funding sources, as well as any other resources used or significant assistance.

Disclosure: Authors must disclose any financial, economic or professional interests that may have influenced the design, execution or presentation of the scholarly work. If there is a disclosure, it will be published with the article.

Clinical Trials: Any educational research studies that are designed as "clinical trials" must register the trial before submitting to the *Journal of Dental Education*. The registration number must be provided in the manuscript.

The studies can be registered at U.S. National Institutes of Health Clinical Trials Registry, EU Clinical Trials Register, or WHO International Clinical Trials Registry Platform.

2. Review Articles

The *JDE* will not consider articles that consist of a general review of topics or published information that is more appropriate for a textbook. However, systematic reviews that focus on trends, issues, new programs or innovations in dental education that are of global interest are welcome. These reviews should not be exhaustive reviews of the literature, but should be concise and address important and relevant questions that affect dental education. Reviews should be presented in a scientific format and use the methods of a systematic review. Authors can refer to the *Cochrane Handbook for Systematic Reviews of Interventions* for more details. In addition, the Editor asks authors of reviews to make sure they follow the PRISMA checklist and flow diagram to ensure the highest quality of systematic reviews and meta-analyses.

For review articles, a structured abstract of 250 words or fewer that addresses the question of interest must precede the review. A brief background and significance section with a review of the literature should be provided. The question being asked and the

justification for the review should be addressed. As with any systematic review, the search strategy and the inclusion and exclusion criteria should be outlined. The authors should describe the findings of the search and the quality of the studies retrieved. The discussion section should compare the findings of the study to the literature at large. Limitations and future areas of interest/research should be identified. Review articles should be limited to 3,500 words with no more than 80 references. No more than six tables and figures should be included. Acknowledgments and any conflicts of interest should be documented as described in the Original Article section.

Solicited or Pre-approved by the Editor

1. Guest Editorials

Each issue opens with a "From the Editor" note or a Guest Editorial solicited by the Editor, usually consisting of a short commentary on articles in that issue or on critical topics of interest to readers. The Editor's annual report about the journal will be published in the January issue.

2. Letters to the Editor

Letters to the Editor should be responses to articles published in the *JDE* in the previous three-month period. They should add to the discussion in a scientific manner, without being personal reflections or reactions. On occasion, letters that deal with the profession, education and training, as well as issues critical to dental education, will be considered. Letters should be brief, focused on one or a few specific points or concerns, and can be signed by no more than four individuals. The letter should be limited to 400 words and six references in *JDE* format. Authors should submit letters directly to the Editor (JDEeditor@adea.org).

3. Perspectives

Perspectives articles should provide an opinion-based but well-supported commentary on controversies, innovations or emerging trends in dental education. On occasion, manuscripts addressing historical figures/perspectives that are impacting current practices will also be considered. Perspectives articles may also be solicited by the Editor on issues that are critical in dental education. Authors who want to independently submit a commentary should contact the Editor ahead of time by e-mail. These articles will be limited to 2,000 words, no more than 10 references, and no more than two figures and/or tables.

Perspectives articles should consist of a) an introduction that addresses why this topic is of general interest to a North American and/or global audience; b) a main section that contains the information relevant to the area being discussed, the author's perspective on it and the grounds for that perspective; and c) a summary that describes the importance of the commentary/perspective to the current and future status of the topic and recommendations concerning how these items can be addressed.

Authors should submit inquiries for submission of perspectives directly to the Editor (JDEeditor@adea.org).

4. Brief Communications

Brief Communications should be used to inform readers about significant findings in studies based on a limited data set, such as a topic of local relevance/interest or limited to one class/course. These communications will typically contain novel items/findings that are time-sensitive. These articles should include an unstructured abstract of 150 words or fewer. This category of article will be limited to 1,000–1,500 words, no more than 10

references and no more than two tables and/or figures. Authors should submit inquiries for submission of Brief Communications directly to the Editor (JDEeditor@adea.org).

5. Point/Counterpoint

Point/Counterpoint articles will be solicited by the Editor, who will provide those authors with information about required length and format.

Special Reports

In addition to the above types of manuscripts, the *JDE* occasionally publishes several types of articles and reports that fall outside the standard peer-review process. These include Association Reports (which are written by ADEA staff members) and special reports/sections/issues (which are the result of special activities or studies conducted by ADEA or other groups and are considered on a case-by-case basis by the Editor). Each year, the ADEA Annual Proceedings and the abstracts of poster and TechExpo presentations at the ADEA Annual Session & Exhibition are also published in the *JDE*. All these types of documents undergo systematic internal review and selected external review as determined by the Editor.

II. Requirements and Policies for Submitted Manuscripts

The *JDE* considers only manuscripts that are in MS Word and submitted electronically (see "Submission and Production Procedures" below for the submission process). All manuscripts submitted to the journal should follow the "Uniform Requirements for Manuscripts Submitted to Biomedical Journals," compiled and published by the International Committee of Medical Journal Editors (ICJME). Authors are also encouraged to refer to the code on good publication practice produced by the Committee on Publication Ethics.

No Prior Publication or Duplicate Submissions. Manuscripts are considered for publication only if they are not under consideration by other journals and have not been published previously in the same or substantially similar form. Submitting authors should attest to their compliance with this requirement in their cover letters. Should a prior or duplicate publication be discovered, the Editor will address the matter with the affected author/s and the other journal's editor following guidelines published by the ICJME and by the Committee on Publication Ethics.

Plagiarism. Plagiarism is a violation of scholarly standards and will not be tolerated. If a case of plagiarism is alleged or discovered, the Editor will address it with the affected author/s, following ICJME guidelines. Authors should exercise extreme care in quoting or paraphrasing material from published sources, so as not to risk plagiarism.

Conflict of Interest. A conflict of interest exists when professional judgment concerning a primary interest may be influenced by secondary interests (professional, personal, financial, etc.). Forms declaring any conflict of interest must be submitted for each author when the manuscript is submitted for consideration. The form can be found on ScholarOne Manuscripts in the upper right-hand corner under ''Instructions & Forms.''

Human Subjects. It is the author's responsibility to obtain approval or exempt status from his or her institution's Institutional Review Board for studies involving human subjects; this approval or exempt status must be mentioned at the very beginning of the Methods section. Failure to meet these requirements is likely to place the manuscript in jeopardy and lead to a rejection.

Editorial Assistance. Manuscripts considered for submission must be written in standard academic English that is comprehensible to English-speaking readers. The American

Medical Writers Association (AMWA) offers a Freelance Directory with contact information for editors who provide assistance in the writing of medical literature, especially for authors whose first language is not English. Please visit their website for further information.

III. Document Preparation, Organization and Formatting

Manuscripts submitted for consideration should be prepared in the following parts, each beginning on a new page:

Title page

Abstract and keywords

Text

Acknowledgments

References

Tables

Figures

Figure titles if figures are provided as images

Blinding. Both blinded and non-blinded manuscripts should be prepared once the original manuscript has been completed. All institutional references should be removed from the body of the manuscript to produce the blinded version; please indicate in the file name which version is blinded.

Document Format. Create the documents on pages with margins of at least 1 inch (25 mm) and left justified with paragraphs indented with the tab key, not the space bar. Use double-spacing throughout and number the pages consecutively. Do not embed tables and figures in the body of the text but place them after the references; include callouts for

each table or figure in the text (e.g., see Table 1). Unless tables vary significantly in size, include all in one document. If any figures are large files, submit them as separate documents.

Title Page. The title page should carry 1) the title, which should be concise but descriptive, limited to 15 words and no more than 150 characters; 2) first name, middle initial and last name of each author, with highest academic degrees; 3) an affiliations paragraph with the name of each author or coauthor and his or her job title, department and institution, written in sentence style; 4) disclaimers if any; 5) name, address, phone and email of author responsible for correspondence about the article and requests for reprints; and 6) support or sources in the form of grants, equipment, drugs, etc. See published articles for examples.

Individuals listed as authors must follow the guidelines established by the ICMJE: 1) substantial contributions to conception and design, or acquisition of data or analysis and interpretation of data; 2) drafting the article or revising it critically for important intellectual content; and 3) final approval of the version to be published. It is the submitting author's responsibility to make sure that authors have agreed to the order of authorship prior to submission.

Abstract and Key Words/MeSH terms. The second page should carry the title and an abstract of no more than 250 words. For research studies, the abstract should be in the structured form described above. Abstracts should be written in the third person, and references should not be used in the abstract. The abstract should include the year of the study and, for survey-based research, the response rate. Below the abstract, provide three to five key words or phrases that will assist indexers in cross-indexing the article and will

be published with the abs tract. At least three terms should come from the Medical Subject Headings listed at the National Library of Medicine. Guidelines for words found in the Medical Subject Headings can be found here. Authors should confirm these terms still exist in the Index Medicus or should search for more accurate terms if not found in our list. **NOTE:** Authors will also be prompted to identify Key Words when submitting their manuscripts in ScholarOne. These Key Words may differ from the items presented here. The Key Words identified in ScholarOne are generated from a list that will best match the submitted manuscript to a Peer Reviewer with expertise in the area(s) identified.

Text. Follow American (rather than British) English spelling and punctuation style. Spell out numbers from one to ninety-nine, with the exception of percentages, fractions, equations, numbered lists and Likert scale numbers. The body of the manuscript should be divided into sections preceded by appropriate subheads. Major subheads should be typed in capital letters at the left-hand margin. Secondary subheads should appear at the left-hand margin, be typed in upper and lower case and be boldfaced. Tertiary subheads should be typed in upper and lower case and be underlined. For authors whose first language is not English, please use a medical writer or a native English-speaking colleague to edit the manuscript prior to final submission. Manuscripts will be rejected prior to peer review if there are numerous usage or grammatical errors.

Please Note: In preparing the main document for submission, save the original file with the word "unblinded" at the end of the file name. Please also remove all author names and affiliated institutions from the original manuscript, and save this version with the word "blinded" at the end of the file name.

References. Number references consecutively in the order in which they are first mentioned in the text. Each source should have one number, so *be careful not to repeat sources in the reference list*. Identify references by Arabic numerals, and place them in the text as superscript numerals within or at the end of the sentence. Do not enclose the numerals in parentheses, and be sure to follow American rather than British or European style conventions (e.g., the reference number follows rather than precedes commas and periods). Two important reminders: 1) references should not be linked to their numbers as footnotes or endnotes and 2) references to tables and figures should appear as a source note with the table/figure, not numbered consecutively with the references for the article. Follow the style of these general examples. Titles of journals should be abbreviated according to the Index Medicus style. Do not use italics or boldface anywhere in the references. If the publication has one to four authors, list all of them; if there are more than four authors, list the first three followed by et al.

Book

1. Avery JK. Essentials of oral histology and embryology: a clinical approach. 2nd ed. St. Louis: Mosby, 2000.

Chapter in an Edited Volume

2. Inglehart MR, Filstrup SL, Wandera A. Oral health and quality of life in children. In: Inglehart MR, Bragramian RA, eds. Oral health-related quality of life. Chicago: Quintessence Publishing Co., 2002:79-88.

Article in a Journal

3. Seale NS, Casamassimo PS. U.S. predoctoral education in pediatric dentistry: its impact on access to dental care. J Dent Educ 2003;67(1):23-9.

Report

4. Commission on Dental Accreditation. Accreditation standards for dental education programs. Chicago: American Dental Association, 2010.

Web Source

5. American Dental Hygienists' Association. Position paper: access to care. 2001. At: www.adha.org/profissues/access_to_care.htm. Accessed: November 27, 2012.

Figures. Figures may be charts or graphs, photographs, or scientific images; any illustration that consists of text should be called a table (see below). Each figure should have a title, numbered consecutively with Arabic numerals in the order in which they appear in the text. Figures may be provided pasted into an MS Word document or as a separate TIFF or JPEG. Do not put the title on the image itself. Rather, if the image is in a Word document, place the title below the image; if the image is in a TIFF or JPEG, provide the figure titles in a list at the end of the manuscript. For graphs, be sure to label both axes. Include a key to symbols, patterns or colors in the figure either as a legend on the image or as a note below the figure. Any sources should appear in a Source note below the figure. Remember that the total number of figures and tables submitted with an article must not exceed six.

Figures should be used selectively to illustrate major points that cannot be expressed well in textual format. Authors should be able to articulate (for themselves, not as part of the submission) why a figure is necessary and what it adds to the understanding of the points made in the manuscript. Figures should be of the highest possible quality—typically 1,000 dots per inch (dpi) for monochromatic images and 600 dpi for images including

halftones. Illustrations should not exceed 8½ x 11 inches, and all lettering should be at least 1½ mm high. If your article is accepted, we may request illustrations in higher resolution than those you've submitted.

Display of Quantitative Information:JDE readers expect authors to employ the highest standards of information design to display information in figures. It is recommended to review the seminal work by Edward R. Tufte, "The Visual Display of Quantitative Information," before designing figures that display quantitative information: Tufte, Edward R., The visual display of quantitative information. 2nd ed. Cheshire, Connecticut: Graphics Press; 2001, ISBN-13: 978-0961392147.

Illustrations: Illustrations should be employed to showcase complex relationships that can be explored by the reader to gain additional insight beyond what was already presented in the manuscript. While illustrations are part of the manuscript, they need to fulfill a purpose for themselves and must have value as standalone elements—telling a particular story or showcasing a relationship not easily expressed in words. It is recommended to review works on information design, such as "The Functional Art: an Introduction to Information Graphics and Visualization" by Alberto Cairo, before designing illustrations: PeachPit Press, 2012, ISBN-13: 978-0321834737.

Figure Checklist:

1. Planning:

· Small, noncomparative and highly labeled data sets belong in tables rather than figures.

· Show data variations, not design variations.

· The number of information-carrying (variable) dimensions depicted should not exceed the number of dimensions in the data; i.e., no 3D bars for pocket depths in mm.

· Above all else show the data (data ink) not design variations.

· Range frame should replace non-data-bearing frame.

· The same ink should often serve more than one graphical purpose.

· Organize and order the flow of graphical information presented to the eye.

(adapted from E. Tufte: The visual display of quantitative information.)

2. Design:

· Variations in font size reflect importance and have meaning.

· Data sets are labeled directly, avoiding cognitive overhead for the reader to decode patterns or shades.

· All symbols (*, #, etc.) are explained in the legend.

3. Execution:

· All source files are available on request, and minimal resolution guidelines have been followed.

· If JPEG images or other compressed formats are used, export has been done with maximal quality setting.

· Color is not used.

· Vector graphics are preferred (using drawing or illustration programs such as Adobe Illustrator).

Tables. Each table should have a title, numbered consecutively with Arabic numerals in the order in which they appear in the text. All tables should be in column format. Arrange column headings so that their relation to the data is clear. Indicate explanatory notes to

items in the table with symbols or letters (note that asterisks should be used only with p-values) or in a general note below the table. Any sources should appear in a Source note below the table. All percentages in tables should include the % sign.

Note that tables may be uploaded in PDF form for initial consideration and peer review; however, *tables must be uploaded as MS Word documents for final review and, if accepted, for production*. Remember that the total number of figures and tables submitted with an article must not exceed six.

Permissions. Any aspect of the article that is not the author's original work (e.g., figures or tables from other publications) must be fully credited to the original publication. It is the author's responsibility to acquire permission to reprint the material and pay any fees. Evidence of required permissions must be in the author's hands before the article can be published.

Manufacturers. Manufacturers of equipment, materials and devices should be identified with the company name and location in parentheses immediately after the first mention.

Commercial Products. Do not use brand names within the title or text, unless the paper is comparing two or more products. If identification of a product is needed, a generic term should be used and the brand name, manufacturer and location (city/state/country) mentioned in parentheses.

IV. Submission and Production Procedures

Submissions should be made via the ScholarOne system, following these steps:

1. Launch your web browser and go to the *JDE*'s submission homepage at http://mc.manuscriptcentral.com/jdentaled.

2. Log-in, or click the "Register here" option if you are a first-time user of ScholarOne Manuscripts. Follow the instructions to create a new account. If you have forgotten your login details, go to "Password Help" on the journal's ScholarOne Manuscripts homepage and enter your email address. You will be sent instructions on how to reset your password.

3. Prior to starting the process of submission, please review your manuscript against the Author Submission Checklist and make sure you have the following items prepared for uploading:

    a) Separate title page (with all author information/titles as requested)

    b) Original manuscript (NOTE: MeSH terms must be provided as requested after abstract)

    c) Blinded version of manuscript as described

    d) Figures

    e) Tables

    f) IRB letter

    g) Conflict of interest form

4. After logging in, select "Author Center." Click the "Submit a Manuscript" link. Enter data and answer questions as prompted. Click on the "Next" button on each screen to save your work and advance to the next screen. Keep advancing until you reach the "upload" page.

5. To upload your files, click on the "Browse" button, locate the file on your computer and select the appropriate designation. Click the "Upload" button when

all files have been selected. Please review your submission (in both PDF and HTML formats) before sending to the Editor. Click the Submit button.

Review Process. Manuscripts submitted as Original Articles, Perspectives, Brief Communications and Review Articles will be peer-reviewed by individuals, selected by the Editor or Associate Editor, who have expertise and experience pertinent to the topic. The journal follows a blind peer review process. The Editor and/or Associate Editor also review all manuscripts. The review process can take up to four months.

From Review to Acceptance. If the manuscript is accepted or changes are recommended, it will be returned to the author with the reviewers' comments for the author's responses and revisions. After the author has made changes, the manuscript is returned for final review to the Editor. If the Editor finds it acceptable, he notifies the author of its formal acceptance and assigns it to an issue. Currently, the time from acceptance to publication is approximately eight to ten months.

Agreement to Publish. On acceptance or provisional acceptance of the manuscript for publication, the author will be asked to sign a publication agreement, which must be signed and submitted before the article is published. This form is a legal document specifying that the article is original and that the author holds all rights in it and grants the journal the exclusive first serial rights to it, for both paper and online publication. If the article is coauthored, all authors must sign the agreement.

Page Proof Review. Corresponding authors will receive page proofs of their articles by email from the Managing Editor. Corresponding authors should remember to update their email addresses in ScholarOne if it changes after the article is accepted. Changes at the

page proof stage will be limited to correction of errors and updates to authors' titles or institutions. Authors will typically have two to three business days to review their proofs.

Reprints and Permissions. Authors are given the opportunity to order reprints of their articles and are urged to do so at the time the issue is printed for the most timely and efficient service; however, reprint orders will be accepted at any time after the issue is published. The price of reprints varies with the page count of the article and the quantity of reprints ordered. The Managing Editor sends detailed information and an order form to the corresponding author with the article proofs. A copy of an individual article may also be acquired online, whether by the authors or other readers, by visiting the *JDE* website. Electronic versions can also be downloaded if you are a subscriber or have access to the *JDE* through a library.

The *JDE* permits the photocopying of articles for the noncommercial purpose of educational and scientific advancement.

V. Key Contacts

General questions (not for submission of manuscripts; see below). Contact Dr. Nadeem Karimbux, Editor, *Journal of Dental Education*, Tufts University School of Dental Medicine, One Kneeland St., DHS-15, Boston, MA 02111; JDEeditor@adea.org.

Submission. Direct questions about submission of manuscripts through ScholarOne to Sarah Burstyn, Publications Associate, *Journal of Dental Education*, 655 K Street, NW, Suite 800, Washington, DC 20001; 202-289-7201 phone; 202-289-7204 fax; burstyns@adea.org.

1. **Proofs and production.** Direct questions about proofs and other matters after article acceptance to Lynn Page Whittaker, Managing Editor, *Journal of Dental Education*, 127 Autumnwood Avenue, Athens, GA 30606; whittakerl@adea.org.
2. **Reprints and copyright permission.** Address correspondence relating to copyright and other business matters to Susan Kimner, Director of Publishing, American Dental Education Association, 655 K Street, NW, Suite 800, Washington, DC 20001; 202-289-7201 phone; 202-289-7204 fax; kimners@adea.org.

Section II: Publishable Manuscript

A manuscript entitled, "Dental Hygiene Faculty Calibration to Enhance Inter-rater Reliability with Calculus Detection: A Pilot Study," is included in the thesis in lieu of Chapters IV and V to report results, discussion, and conclusions of the thesis research study. This manuscript will be submitted for publication in the Journal of Dental Education. The manuscript section of the thesis reflects the specifications outlined in the author guidelines (Appendix L).

Dental Hygiene Faculty Calibration to Enhance Inter-rater Reliability with Calculus Detection: A Pilot Study

Lisa Santiago, RDH, BS; Jacqueline J. Freudenthal, RDH, MHE; Teri Peterson. EdD: Denise M. Bowen, RDH, MS

Lisa J. Santiago is currently a Dental Hygiene Instructor at Halifax Community College and was a graduate student during the time of the study at Idaho State University; Jacqueline Freudenthal, RDH, MHE is an Associate Professor, Department of Dental Hygiene, Idaho State University; Teri Peterson, EdD is presently Assistant Professor, College of Business, formerly Statistical Consultant, Division of Health Sciences, Idaho State University- Denise M. Bowen is Professor Emeritus, Department of Dental Hygiene, Idaho State University.

Direct correspondence and requests send to Lisa J. Santiago, Halifax Community College, Dental Hygiene Program, 100 College Drive, P.O. Drawer 801, Weldon, North Carolina, 27890; 252-538-4339 phone, lsantiago730@halifaxcc.edu.

The authors have no disclaimers or disclosures related to this study or manuscript.

Abstract

Purpose: A pre-test post-test design was used to assess the impact of faculty calibration on inter-rater reliability of calculus detection scores using two accepted standards: (a) established by testing agencies (i.e., CRDTS), readily detectible calculus, and (b) the gold standard for scaling/root planing, relative root smoothness.

Methods: Participants (N=4) explored calculus on assigned teeth (N=16, 64 surfaces) in patients (N=4). Faculty calculus detection scores were calculated pre-and post calibration training.

Results: Kappa averages using CRDTS criteria were 0.54 at pre-test, 0.62 at post-test. Kappa scores using the scaling /root planing standard were 0.21 at pre-test, 0.22 at post-test. Scores indicated improvement from moderate (Kappa=0.41-0.60) to substantial agreement (Kappa =0.61-0.80) following training using the CRDTS standard. Although this result differs qualitatively, and kappas are significantly different from 0, differences in pre- to post-Kappas for patient-rater dyads using CRDTS (p=0.340), were not statistically significant. There was no difference (p=0.340) in Kappa scores pre- to post-training using the scaling/root planing standard.

Conclusion: Training improved inter-rater reliability to substantial agreement using the CRDTS standard, but not using the gold standard, possibly due to greater difficulty in attaining agreement regarding relative root roughness. Future studies should include multiple sessions and use of endoscopy in patients during training.

Key Words

Key words: calculus detection, calibration training, dental hygiene faculty, dental hygiene education, faculty development

Introduction

Challenges exist in dental hygiene programs and dental schools to ensure reliability, agreement, and consistency in faculty assessments of students' clinical performance. Over the course of time, researchers have sought methods to increase reliability in evaluation of student performance in many facets of dental and dental hygiene education. Two decades ago, findings of early faculty calibration studies suggested that standardization in clinical techniques may improve reliability in clinical evaluation.[1-3] Experts have postulated that students need to be aware of ideal performance standards and be allowed the opportunity to perform skills that mimic situations similar to those they will perform after graduation to achieve competency.[4] Studies have concluded that discussion, comparison of technique, and continual practice – all components of faculty calibration training – have the potential to improve reliability.[2,5-10]

Although no data exist to document the relationship between faculty calibration and clinical dental hygiene instrumentation learning outcomes, inconsistency in evaluation potentially could distract student learning, lead to student concerns or frustrations, and detract from the ultimate goal of reaching competency.[11,12] Students' understanding of clinical expectations may be impacted if clinical assessments are inaccurate and variable due to individual faculty member's using their own "gold standards" instead of the gold standard established and accepted by the profession.[13] Faculty assessment methods are necessary components of a clinical evaluation system because these methods assist clinical faculty in making judgments about students' performance in relationship to attaining the "gold standards" within the profession. Faculty should be consistent when making judgments regarding students' performance

and need to know and evaluate students based on the gold standard. If faculty judgment is skewed, the chance of rater agreement or consistency is compromised.[2]

Faculty calibration for calculus detection presents a unique challenge relative to the existence of two established standards for evaluation of student competence in calculus detection and removal. The standard established by the Central Regional Dental Testing Agency (CRDTS) and other similar agencies is described as "readily detectable calculus, a definite 'jump' or 'bump' which is easily detected with one or two strokes; a deposit that easily 'binds' or 'catches' the explorer; ledges or ring formations; spiny or nodular formations.[14] CRDTS reports successful calibration of its examiners for dental hygiene testing using this standard.[15] The gold standard accepted for immediate evaluation of adequacy of scaling and root planing or periodontal debridement in clinical practice and education differs because complete removal of clinically detectable calculus is necessary for the health of adjacent tissues. Although there is little evidence that complete root surface smoothness is a necessity, relative root smoothness is accepted as the best immediate clinical indication that calculus has been removed.[16] This gold standard is commonly accepted for clinical evaluation of students' removal of calculus deposits and likely presents a greater challenge for faculty calibration than the standard used by testing agencies such as CRDTS for evaluation of minimal competency required for licensure.

The problem with the lack of inter-rater agreement among dental hygiene faculty members is widely recognized, however, results of faculty calibration studies are mixed. Haj-Ali and Feil[17] assessed faculty calibration for Class II tooth preparation and concluded that, with one training session, inter-rater agreement with the gold standard

can improve and remain consistent after ten weeks. Jacks et al. conducted long- and short-term evaluation of a single calibration session related to writing SOAP (Subjective data, Objective data, Assessment, and Plans) notes.[7] They concluded that faculty evaluations maintained agreement up to one year after the initial workshop supporting the findings of Haj-Ali and Feil.[17] An evaluation of faculty calibration training involving multiple sessions to enhance evaluation of ethical reasoning and professionalism was conducted by Christie et al. Following three years of intermittent faculty calibration training, assessments showed increased ethical reasoning by students and consistent faculty evaluations of the students.[5]

Recently, a faculty calibration study by Partido et al. tested intra- and inter-rater calculus scores with the use of a dental endoscope.[18] These investigators conducted a pre- and post-test with control and training groups. Using three typodonts with simulated calculus and a manufacturer's answer key, both groups explored calculus using 11/12 explorers twice to attain a baseline. The training group participated two one-hour calibration sessions using didactic, discussion, and individualized instruction on the use of an endoscope. If any errors were discovered, the subjects were allowed to reconcile errors immediately with re-detection with the endoscope and an 11/12 explorer. Partido et al. pre-established a standard for mastery at 80% accuracy with the answer key and a 0.80 Kappa score, parallel to the Kappa average required for regional clinical board examiners. At post-testing, both groups re-evaluated the simulated manufacturer-produced calculus on the original three typodonts with 11/12 explorers. Results reported by Partido et al. showed significant increases in inter-rater reliability levels for those in the training group pre-test to post-test (Kappa score pre-test 0.536 and post-test 0.792).

No significant differences were found with intra–rater reliability of calculus detection scores between the two groups as both groups improved.[18]

Findings of several other studies have indicated that calibration was unable to improve agreement in evaluation of periodontal assessment and calculus detection among faculty. A study designed to assess faculty agreement in diagnosis and management of periodontal diseases before and after calibration training found significant variability among faculty members' interpretation of the gold standard in both the pre- and post-test following a single calibration session and concluded that accepted guidelines and on-going training efforts should be practiced.[8] Garland and Newell [12] concluded there was no effect on inter-rater reliability levels of clinical faculty (N=12) for calculus detection after faculty calibration training using simulated calculus and typodonts, although faculty Kappa averages indicated substantial agreement prior to calibration training, leaving little margin for improvement. These authors' primary recommendation for future studies of faculty calibration for evaluation of calculus was to conduct training using patients for more realistic evaluation results. An earlier one-year study by Pippin and Feil[19] examined inter-rater agreement among clinical examiners (N=10) that explored and scored residual subgingival calculus using patients. Findings indicated only fair agreement with a low Kappa score of 0.33 and 0.34 respectively after calibration, although substantial agreement (Kappa 0.061-0.80) or nearly perfect agreement (0.81-1.0) is desired for calibration of faculty examiners.[19] A companion study using simulated calculus had similar results. Authors of these previous faculty calibration studies which did not use emerging technologies such as endoscopy suggests the need for objective measures of

calculus detection or the use of an endoscope in patients to attain effective faculty calibration training for subgingival calculus detection techniques.[12,19]

Research findings documenting the value of dental and dental hygiene faculty calibration and the impact of inter-rater and intra-rater reliability on student learning and satisfaction are limited in scope and depth and have presented conflicting findings. Although many authors have been unsuccessful in attaining calibration, all agreed that faculty training was beneficial and concluded that different approaches or multiple calibration sessions might be needed. Many agreed, however, that the specific time and mechanism needed for calibration remains undetermined.[1,6,8,12] The aim of this pilot study was to evaluate the impact of dental hygiene faculty calibration training on inter-rater reliability of faculty evaluation during calculus detection using two different accepted standards. Calibration training was conducted clinically with patients to address the main recommendation made by Garland and Newell in their study of faculty calibration for inter-rater reliability in calculus detection.[12] In addition, two standards were used to evaluate the effectiveness of the faculty calibration training. Relative root roughness is the gold standard for scaling and root planing or periodontal debridement that is used by most, if not all, dental hygiene and periodontal educational programs for evaluation of calculus removal. This standard may not be equally replicated when using patients and replacement teeth for typodonts purchased with simulated calculus. Endoscopy was not used in this pilot study due to the anticipated increase in time and procedures required for the patients. Use of endoscopy for faculty development training also is not feasible for many dental hygiene programs due to the cost of the dental endoscope.

Materials and Methods

This pilot study measured inter-rater reliability of dental hygiene faculty who work at the Halifax Community College (HCC) Dental Hygiene Program before and after calibration training. A standard pre-test, post-test single group design was used with a total of three sessions (Figure 1). Following exemption from the sponsoring university's institutional review board, the Human Subjects Committee, all full-and part-time clinical faculty members at HCC (N=8) in Weldon, North Carolina were invited to participate via a letter.

Patients (N=4) were selected with varying degrees of calculus deposits and met the following inclusion criteria: (a) ≥ 18 years old; (b) last prophylaxis more than six months prior to the study; (c) availability of a full mouth series of radiographs exposed within the last 12 months; (d) blood pressure ≤ 160/100; and (e) a minimum of 24 teeth present in the dentition. After obtaining informed consent, patient eligibility was confirmed.

During the pre-test session, patients' health histories and blood pressure readings were updated, and the clinical faculty members were assigned alphabetically to one of two groups. The ODU 11/12 explorer and full mouth radiographs present on each of the patients' view boxes were used to determine if calculus was present. Each group was randomly assigned two quadrants to explore for presence or absence of calculus deposits for each patient. This half-mouth assignment was made for each faculty participant in each of the two groups for the comfort of the patients by limiting the number of times each surface would be explored. On the right side of patients' mouths, six (24 surfaces) were explored and on the left side of patients' mouths, ten teeth (40 surfaces) were explored to assure that the selected teeth were present in all of the patients' mouths.

Using the CRDTS Dental Hygiene "Full Mouth Patients Screening" Worksheet,[14] participants explored the two quadrants that were randomly assigned to them and recorded a C for calculus deposits or a R for rough tooth surface. The box was left blank where no calculus or rough surface was detected. No discussions of findings were allowed between faculty participants during pre- or post-testing. The pre-test was completed in approximately 20 to 30 minutes per patient for a total of two hours.

A calibration workshop was conducted one week after the pre-test with the post-test immediately following the training. The calibration session was conducted for one hour and fifteen minutes. The calibration training included: (a) the definition and purpose of calibration training; (b) a video by Nield-Gehig[20] demonstrating the proper technique for using the ODU 11/12 explorer; and (c) the use of a UMKC model with detectible calculus for practice and application of techniques discussed. Participants openly discussed pre-developed scenarios that depicted the most challenging instances encountered during calculus detection such as anatomical landmarks of a tooth which can be mistaken as calculus, existing restorations, root proximity and roughness, and malaligned teeth. Open discussion about the pre-test occurred, and the use of air was emphasized.

Patient preparation, armamentarium, and recording protocols for the post-test were identical to those described for the pre-test. Each clinical faculty member explored the same side of each patients' mouths as they did during the pre-test. The post-test scoring was completed in two hours. All patients were invited to return for the researcher to discuss findings and to receive preventive dental hygiene care at their convenience.

A data set comprised of all pre-test and post-test calculus detection data was constructed in an Excel spreadsheet for statistical analysis using SPSS version 21.0.0 (2012). Cohen's kappa coefficients were calculated at the pre-and post-test for each of the standards to test the following two hypothesis: 1) there is no difference in inter-rater reliability of faculty calculus scores as measured by Cohen's kappa scores following calibration training designed to enhance agreement of faculty members' assessment of calculus deposits based on the standard established by CRDTS, readily detectible calculus; and 2) there is no difference in inter-rater reliability of faculty calculus scores as measured by Cohen's kappa scores following calibration training designed to enhance agreement of faculty members' assessment of calculus deposits based on the standard established for evaluation of adequacy of scaling and root planing, relative root smoothness, before and after faculty calibration training. A paired sample t-test was used to test the statistical significance of changes in kappa scores pre- to post-training at the 0.05 alpha level.

Results

This pilot study used a pre-test post-test design to assess the impact of a faculty calibration program on inter-rater reliability of dental hygiene faculty scores based on two accepted standards: (a) an established standard by CRDTS, readily detectible calculus, and (b) the gold standard accepted for evaluation of scaling/root planing, relative root smoothness. The convenience sample of participants who agreed to participate and signed the informed consent form included four dental hygienists and two dentists. Data collection was limited to the dental hygienist faculty scores (N=4) to control for extraneous variables associated with differences in education and experience.

The study included two full-time and two part-time faculty members. One of the participants had two years of clinical teaching experience while the remaining three faculty members had eight to ten years of clinical teaching experience. All participants had over twenty years of clinical practice experience as licensed dental hygienists.

A Cohen's kappa was used to assess inter-rater reliability of faculty calculus detection scores both before and following calibration training based on the standard established by CRDTS, readily detectible calculus.[14] Table 1 shows the pre-test Kappa score 0.535 and 0.622 at post-test indicating improvement from moderate (Kappa=0.41-0.60) to substantial agreement (Kappa =0.61-0.80).[21] Although this result is qualitatively different, and the kappas are significantly different from 0, when comparing the pre- to post- Kappas for patient-rater dyads using CRDTS (p=0.340) there was no statistically significant increase in the Kappa scores. This lack of statistical significance could be a result of the small sample size in this pilot study.

Table 1 also shows the Kappa scores based on the gold standard established for immediate evaluation of adequacy of scaling and root planing or periodontal debridement, relative root smoothness. These Kappa scores were 0.206 (p=0.001) at pre-test indicating slight to fair agreement and 0.221 (p=0.001) at post-test, indicating slight to fair agreement, indicating no qualitative or quantitative improvement. Using a matched pairs t-test, this minor change was not statistically significant (p=0.652): therefore, the null hypothesis could not be rejected concerning inter-rater reliability concerning using the gold standard.

Discussion

Studies of faculty calibration for calculus detection presents a unique challenge in design and measurement relative to the existence of two different standards used for the evaluation of calculus removal. The problem with the lack of inter-rater agreement among dental and dental hygiene faculty members has been widely recognized in several areas of clinical evaluation. To evaluate the effectiveness of any calibration effort, subject scores should be compared pre-training to post-training.[12] Specific to faculty calibration for calculus detection, some authors have suggested that problems of inconsistency among faculty can be reduced with calibration training.[18] whereas others have found it to be effective.[12,19] These authors have recommended using patients for faculty calibration of calculus detection and/or using emerging technologies such as dental endoscopy for training. In addition, different standards or types of calculus have been used for calibration of faculty and testing agency examiners. To address the potential impact of the use of these two different standards, this study evaluated the impact of faculty calibration training on calculus detection in patients using both standards.

This pilot study followed the recommendations of Garland and Newell[12] to evaluate calculus in patients with varying amounts of deposits rather than typodonts with simulated calculus in order to provide realistic tactile sensations for faculty calibration. The patients selected for this study represented the different degrees of calculus deposits students may encounter in the clinical setting, from light and grainy to readily detectable. The results indicated there was a difference in the faculty's inter-rater reliability when the CRDTS criteria were used; however, there was no difference when all types of deposits were assessed, despite time spent discussing the associated challenges during training.

The studies by Garland and Newell[12] and Partido et al.[18] used typodonts that included varied numbers of teeth with simulated calculus for faculty calibration training and evaluation. Garland and Newell hand crafted the calculus to simulate light, moderate, and heavy deposits while Partido et al. used typodont teeth prepared with calculus deposits by the manufacturer. These manufacturer-produced deposits tend to be moderate to heavy, similar to the type of deposits described by the CRDTS standard. The gold standard for evaluating of adequacy of scaling and root planing in clinical practice and education differs from that used by CRDTS and other dental hygiene testing agencies because removal of all *clinically detectable* calculus is necessary for the health of adjacent tissues.[16] There is no expectation that *complete* root smoothness be attained, so *relative root smoothness* and tissue reevaluation to assess healing is recommended. This gold standard of relative root smoothness is accepted universally for immediate clinical evaluation of students' removal of calculus deposits and likely presents a greater challenge for faculty calibration than the standard for removal of readily detectable calculus used by testing agencies such as CRDTS for evaluation of minimal competency required for licensure. The results of this pilot study seem to indicate that, while calibration can be attained for detection of readily detectable calculus, it is not easily attained for evaluation of relative root roughness. Students have anecdotally reported frustration with differences in evaluation of scaling and root planing, or periodontal debridement,[11] a skill requiring a much higher level of skill and judgment than removal of readily detectable calculus.

The results of calculus detection following faculty calibration using the CRDTS criteria indicated substantial agreement of faculty calculus detection scores after

training.[21] CRDTS reports successful calibration of examiners using this standard and requires a Kappa average of ≥ 0.80.[23] The calibration training in this pilot study utilizes similar training techniques used by CRDTS, such as a Power Point presentation and discussion regarding policy, procedures, and protocol for evaluation and recording of calculus deposits detected during exploration.[23] Also, the use of a typodont with artificial calculus for practice and application of principles discussed in the workshop, and demonstration of proper use of the ODU 11/12 explorer is employed. The level of training provided by CRDTS, however, exceeds the training provided during this study. Multiple training sessions are conducted by CRDTS for continuous examiner calibration as they move forward in training, rather than a single session.[23] Results of previous faculty studies in dental and dental hygiene programs have concluded that multiple training sessions over time may be indicated to attain agreement. [1, 10, 17,19]

Faculty training prior to post-testing did not emphasize individualized technique training in detecting root roughness other than the video demonstration of the use of the ODU 11/12 explorer and the discussion of case studies presenting related challenges. The subjects explored examples of root roughness on the typodont, and the majority of time was spent discussing situations that would help faculty differentiate root roughness from other conditions and past experiences in determining root roughness. Perhaps exploring these types of deposits in patients' mouths and discussing agreement and differences, as well as evaluating exploring techniques used for calculus detection, would enhance calibration training for student evaluation using the gold standard. The participants informally shared qualitative comments after the post-testing, indicating that

exploring for calculus deposits using the scaling and root planing standard, in particular, was very stressful, as they tended to "second guess" themselves.

This study provided an adequate representation of typical calculus that a dental or dental hygiene student may encounter during patient care with current radiographs displayed. CRDTS examiners only identify calculus present on a tooth surface if it is easily distinguishable. Determining whether relative root roughness is a calculus deposit requiring additional periodontal instrumentation poses a significant challenge; however, dental hygiene students' performance commonly is evaluated using this standard in the clinical setting. CRDTS reports having determined that it is very difficult to achieve 80% agreement with relative root roughness due to widely varying opinions regarding the presence and extent of root roughness.[23] CRDTS is an entry-level competency assessment required for licensure, therefore, examiners are required to demonstrate a clear understanding of the defined criteria for qualifying calculus and calculus removal.[23] When measuring inter-rater agreement for calculus detection using the CRDTS criteria, this study attained substantial agreement post-training. When analyzing the data concerning the standard of scaling root planing or periodontal debridement, inter-rater agreement improved marginally, from slight to fair agreement, but did not achieve the goal of substantial agreement. This difficulty in evaluation of relative root roughness provides the foundation of the shift from final evaluation of scaling and root planing upon completion to periodontal debridement requiring an ultimate evaluation based on tissue response four to six weeks following nonsurgical periodontal therapy.[14]

In dental and dental hygiene institutions, competency assessments commonly are completed on the detection of calculus based on the gold standard used in this study.

Based on the results this pilot study, the question arises, should calculus detection competencies be based on students' ability to identify readily detectible calculus instead of evaluating root roughness immediately after completion of instrumentation with the final competency assessment being based on tissue response at reevaluation? Another alternative worth consideration might be to base competency examinations on the standard of readily detectable calculus coupled with a non-graded evaluation system for relative root roughness for initial evaluation of periodontal debridement or root planing, with discussion and assistance provided the faculty. Student competencies should be developed to consider the different standards and challenges for experienced faculty let alone inexperienced students. The question of how to fairly and accurately assess calculus deposits on calculus detection evaluations is one that should be contemplated by dental and dental hygiene educators and studied in the future.

The limitations to this study include the small sample size (N=4): however, the number of tooth surfaces (N=64) evaluated did provide adequate power for statistical analysis. The use of a convenience sample of clinical faculty members from a single dental hygiene program also limits the ability to generalize results to other institutions.

Future research should incorporate the use of technologies such endoscopy, similar to Partido et al.,[16,24] during training using patients for faculty calibration in evaluation of sub-gingival calculus detection. This approach would enhance external reliability of results as the full range of calculus deposits evaluated when assessing the immediate outcomes of scaling and root planing or periodontal debridement would be represented. The presence of calculus could be evaluated and discussed by faculty utilizing the endoscope with patients during an additional session of the faculty

calibration training following the initial discussion and training session using typodonts. If the use of endoscope is not possible, validity could be improved by utilizing a third examiner to independently confirm deposit detection and ratings when disagreement occurs between faculty members during training. A larger sample size from multiple institutions may provide additional data to support or deny the effectiveness of calibration using the gold standard; however, additional pilot studies may be needed to confirm feasibility of attaining agreement with the gold standard prior to the conduct of larger scale studies. Calibration training should include the use of exploration techniques using color-coded 11/12 explorers, rather than the standard ODU 11/12 explorers.[23] Training sessions should include pre-study materials, be conducted with patients at regular intervals to discuss agreement and disagreement about calculus deposits present, and provide clinical faculty the opportunity to review and improve exploration techniques as well as discuss cases and scenarios.

Conclusion

This study compared two accepted calculus detection standards used in examiner calibration with mixed results. Although changes pre- to post-training were not statistically significant, successful faculty calibration occurred when measuring calculus detection using the CRDTS standard, while non-substantial results occurred with the gold standard of relative root roughness. The challenge of faculty calibration is not unique to dental education. Unfortunately, there has been little documentation provided for the approaches or number of sessions capable of producing the consistency needed to maintain high rater reliability. The question that remains is what standard should be employed when evaluating competency of dental and dental hygiene students in calculus

removal when considering the apparent difficulty in calibration faculty detection scores to the commonly-used gold standard, relative root roughness, and the need for clinical competency using this standard.

Acknowledgments

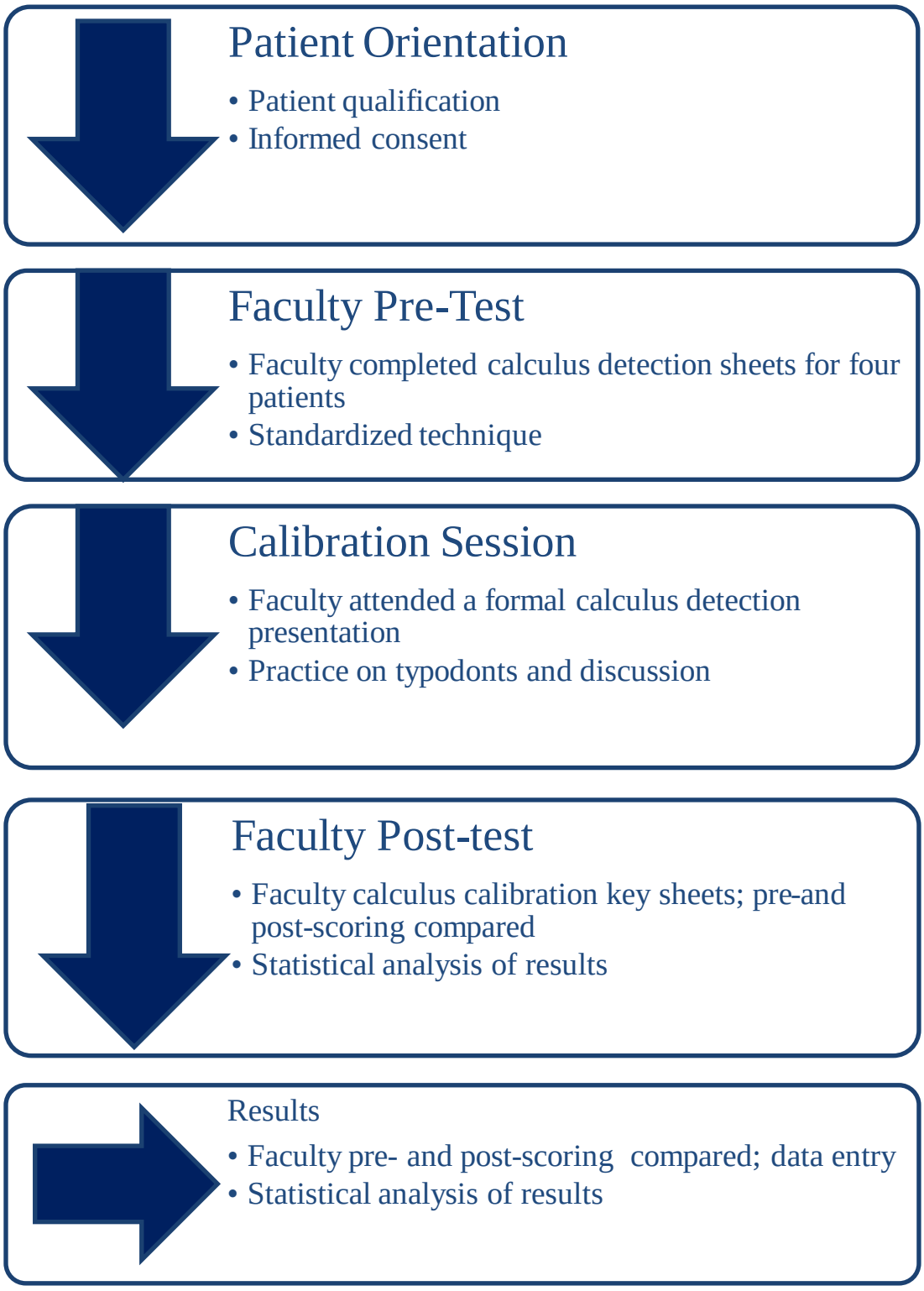
The authors would like to thank the Halifax Community College dental hygiene faculty members for participation in this project.

Table 1. Kappa Scores for Testing Dental Hygiene Faculty Inter-rater Reliability

Standard Used for Calibration	Pre/Post Test	Kappa Statistic	Standard Error	p value
1 CRDTS	Pre	0.535	0.054	0.001
	Post	0.622	0.622	0.001
2 Root Planing	Pre	0.206	0.074	0.001
	Post	0.221	0.066	0.001

Figure 1

Faculty Calculus Detection Calibration Study Protocol



References

1. Dahlström L, Keeling SD, Friction JR, Hilsenbeck SG, Clark GM, Rugh JD. Evaluation of a training program intended to calibrate examiners of tempomandibular disorders. ACTA Odontol Scand 1994; *52*, 250-4.
2. Flack VF, Atchison KA, Hewlett ER, White SC. Relationships between clinician variability and radiographic guidelines. J Dent Educ 1995; 75(2); 775-82.
3. Biller IR, Kerber PE. Reliability of scaling error detection. J Dent Educ 1980; 44(4): 206-10.
4. Knight GW. Toward faculty calibration. J Dent Educ 1997; 61(12): 941-46.
5. Christie C, Bowen D, Paarmann C. Effectiveness of faculty training to enhance clinical evaluation of student competence in ethical reasoning and professionalism. J Dent Educ 2007; 71(8): 1048-57.
6. Assaf AV, Meneghim C, Zanin L, Tengan C, Pereira AC. Effects of different diagnostic thresholds on dental caries calibration- a 12 month evaluation. Community Dental Oral Epidemiol 2006; *34*, 213-19. doi: 10.1111/j1600-0528.2006.00278.x
7. Jacks ME, Blue C, Murphy D. Short and long-term effects of training on dental hygiene faculty member's capacity to write soap notes. J Dent Educ 2008; 72(6): 719-24.
8. Lanning SK, Pelok SD, Williams BC, Richards PS, Sarament DP, Oh T, McCauley LK. Variation in periodontal diagnosis and treatment planning among clinical instructors. J Dent Educ 2005; 69(3): 325-37.
9. Lanning SK, Best AM, Temple HJ, Richards PS, Carey A, McCauley LK. Accuracy and consistency of radiographic interpretation among clinical instructors in conjunction with a training program J Dent Educ 2006; 70(5): 545-56.

10. Alexander AG, Leon AR, Ribbons JW, Morganstein SI. Evaluation of a training programme for the PMA and the Gingival Index. J Periodontal Res 1972; 7(4): 341-5.
11. Henzi D, Davis E, Jasinevicius R, Hendricson W, Cintron L, Isaacs M. Appraisal of the dental school learning environment: the students' view. J Dent Educ 2005; 69(10): 1137-47.
12. Garland KV, Newell KJ. Dental hygiene faculty calibration in the evaluation of calculus detection. J Dent Educ 2009; 73(3): 383-9.
13. John V, Lee SJ, Prakasam, S, et al. Consensus training: an effective tool to minimize variations in periodontal diagnosis and treatment planning among dental faculty and students. J Dent Educ 2013; 77(8): 1022-32.
14. Central Regional Dental Testing Service, CRDTS dental hygiene full mouth patient screening worksheet. 2014. At: http://www.crdts.org/uploads/CRDTS%202014%20Full%20Mouth%20%20Patient%20Screening%20Worksheet.pdf. Assessed: March 3, 2014.
15. Central Regional Dental Testing Service. An evaluation of the central regional dental testing service's national dental hygiene examination. 2011. At: http://www.crdts.org/uploads/2010%20CRDTS%20DH%20Technical%20Report.pdf. Assessed: March 3, 2014.
16. Newman MG, Takei HH, Klokkevold PR, Carranza, F. Carranza's clinical periodontology. 11th ed. St. Louis: Saunders, 2012.
17. Haj-Ali R, Feil P. Rater reliability: short-and long-term effects of calibration training. J Dent Educ 2006; 70(4): 428-33.

18. Partido BB, Jones AA, English DL, Nguyen CA, Jacks ME. (2015). Calculus detection calibration among dental hygiene faculty utilizing dental endoscopy: a pilot study. J Dent Educ 2015; 90(2): 124-32.
19. Pippin DJ, Feil P. Interrater agreement on subgingival calculus detection following scaling. *J Dent Educ 1992; 56*(5): 322-26.
20. Neild-Gehrig JS. Fundamentals of periodontal Instrumentation and advanced root instrumentation. 7th ed. Philadelphia: Lippincott Williams & Wilkins. 2013.
21. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. Family Medicine 2005; 37(5): 360-3.
22. Gansky SA, Pritchard H, Kahl E, Mendoza D, Bird W, Miller AJ, Graham D. Reliability and validity of a manual dexterity test to predict preclinical grades. J Dent Educ 2004; 68(9): 985-94.
23. Personal communication, K Cobb, Executive Director, Data and Dental Exam Administrator of CRDTS, January 2015.