University of Louisville
ThinkIR: The University of Louisville's Institutional Repository

Electronic Theses and Dissertations

5-2017

Using an audience response system to calibrate dental faculty assessing student clinical competence.

Sean A. Aiken
University of Louisville

Follow this and additional works at: https://ir.library.louisville.edu/etd

 Part of the Oral Biology and Oral Pathology Commons

Recommended Citation

Aiken, Sean A., "Using an audience response system to calibrate dental faculty assessing student clinical competence." (2017). *Electronic Theses and Dissertations.* Paper 2636.
https://doi.org/10.18297/etd/2636

This Master's Thesis is brought to you for free and open access by ThinkIR: The University of Louisville's Institutional Repository. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of ThinkIR: The University of Louisville's Institutional Repository. This title appears here courtesy of the author, who has retained all other copyrights. For more information, please contact thinkir@louisville.edu.

USING AN AUDIENCE RESPONSE SYSTEM TO CALIBRATE DENTAL FACULTY ASSESSING STUDENT CLINCAL COMPETENCE

By

Sean A. Aiken
B.S., The College of William & Mary, 2012
M.S., University of Louisville School of Medicine, 2013

A Thesis
Submitted to the Faculty of the
School of Dentistry of the University of Louisville
in Partial Fulfillment of the Requirements
for the Degree of

Master of Science in Oral Biology

Department of Oral Biology
University of Louisville School of Dentistry
Louisville, Kentucky

May 2017

Copyright 2016 by Sean Alexander Aiken

All rights reserved

USING AN AUDIENCE RESPONSE SYSTEM TO CALIBRATE DENTAL FACULTY ASSESSING STUDENT CLINCAL COMPETENCE

By

Sean A. Aiken
B.S., The College of William & Mary, 2012
M.S., University of Louisville School of Medicine, 2013

A Thesis Approved on

September 6, 2016

By the following Thesis Committee

_____
Dr. Michael Metz

_____
Dr. Cynthia Miller

_____
Dr. Weishao Lin

_____
Dr. Theresa Mayfield

_____
Dr. Marcelo Durski

ACKNOWLEDGMENTS

I would like to thank my family back home in Atlanta and my wonderful fiancée Taylor for encouraging me to be inquisitive and to chase the answers to problems that present themselves. I would also like to thank Dr. Cyndi Metz for instilling a passion for educational research and a commitment to bettering the learning environment of those who will come after me. You truly are an inspiration to us students and I am blessed to get to call you a friend. Finally, I would like to thank Dr. Michael Metz, for being a companion, a mentor and a colleague in dentistry these past four years. He sets an example of what I hope to aspire to in the field of dentistry and who I wish to become as an individual.

ABSTRACT

USING AN AUDIENCE RESPONSE SYSTEM TO CALIBRATE DENTAL FACULTY ASSESSING STUDENT CLINCAL COMPETENCE

Sean A. Aiken

September 6, 2016

In order to best prepare students to become competent and confident practitioners in a clinical environment upon graduating dental school, it is imperative for them to receive consistent and productive feedback from the supervising faculty. Through academic engagement, and more specifically faculty calibration, it may be possible to eliminate the disconnect that sometimes exists between faculty expectations and terminology, and those of the students. In terms of definitions, academic engagement reflects faculty scholarly development activities that support integration of relevant, current theory of best practices consistent with the school's mission, expected learner outcomes, and supporting strategies.[1-6, 32]

The difficulty lies in finding an effective, yet cost efficient way to conduct that faculty calibration and ensure that students are receiving consistent and reliable feedback in order to mold them into the most competent clinicians they are capable of becoming. It can be stated that professional faculty engagement is the cornerstone of providing consistent and calibrated clinical instruction to students for patient centered care learner outcomes.[7-11] A significant part of faculty engagement with professional students is to provide foundational knowledge, attitude and skills for both formative and summative assessment of competence.[12-18]

We hypothesize that the introduction of faculty calibration to the clinical faculty will result in more consistent feedback, leading to more predictable results and ultimately more competent clinicians. This, in turn, will increase student perception of clinical faculty yielding an increase in the belief that they are receiving quality, accurate and consistent instruction.[24-30]

TABLE OF CONTENTS

PAGE

INTRODUCTION

A prominent goal of faculty scholarly development activities is to support integration of relevant, evidence-based best practices consistent with the school's mission, expected learner outcomes, and supporting strategies.[1-6] Professional faculty engagement is the cornerstone of providing consistent and calibrated clinical instruction to students for patient centered care learner outcomes.[7-11] A significant part of faculty engagement with professional students is to provide foundational knowledge, attitude and skills for both formative and summative assessment of clinical competence.[12-18] The idea of faculty development or calibration is not a new concept. Many research projects have focused on ways to collectively centralize instructional information for improving student learning outcomes. However, missing from the current literature is a method for conducting calibration sessions that notes weaknesses where a consensus on terminology or concepts is lacking.

Two conceptual educational models help us understand how learning outcomes or objectives relate to learners' professional development as they move along the novice to expertise continuum.[19-23] It is imperative to understand these in order to truly appreciate the research being done. The first is found in Bloom's Taxonomy of Objectives in the Cognitive Domain (1956), which describes how learning objectives related to cognitive development increase in complexity as learners develop deeper understanding, start to apply this knowledge, and ultimately synthesize and evaluate what they have learned.[19-21] While originally published in 1956, the inception of Bloom's Taxonomy was a landmark study in categorizing educational research following a series of conferences from 1949 to 1953, which were designed to improve communication between educators on the design of curricula and examinations. Essentially, the taxonomy divides the learning into three distinct domains. The cognitive domain, which is knowledge based, the affective domain, which is emotive based, and finally the

psychomotor domain, which is action based, make up the three domains of learning according to the Taxonomy.

When revised in 2001, Anderson et al. did an excellent job summarizing the need for Bloom's Taxonomy in educational research. Their response was as follows:

"The authors of the revised taxonomy suggest a multi-layered answer to this question, to which the author of this teaching guide has added some clarifying points:

1 Objectives (learning goals) are important to establish in a pedagogical interchange so that teachers and students alike understand the purpose of that interchange.
2 Teachers can benefit from using frameworks to organize objectives because
3 Organizing objectives helps to clarify objectives for themselves and for students.
4 Having an organized set of objectives helps teachers to:
    - "plan and deliver appropriate instruction";
    - "design valid assessment tasks and strategies"; and
    - "ensure that instruction and assessment are aligned with the objectives."

The professional learning environment that is dentistry, and the way in which our curriculum is structured, provides an excellent infrastructure in which to study Bloom's Taxonomy in the context of faculty calibration and student outcomes. By incorporating iClicker calibration sessions into faculty development, we can affect the first domain of learning Bloom identified, which is the cognitive domain. As previously mentioned, this cognitive domain is knowledge-based. Thus, by using an audience response system to calibrate dental faculty assessing student clinical competence, we can add an additional layer of control over knowledge acquisition in this domain. If the knowledge acquisition stage can become more predictable and effective, then it makes

sense to assume that the other steps would follow suit. As the ultimate goal of our research is to show marked increases in clinical competence and confidence in the students, this domain is fundamental. The progression between knowledge (Cognitive Domain), acceptance of that knowledge (Affective Domain), and action (Psychomotor Domain) is imminently clear in dental education and in clinical feedback on operative procedures and competencies.

In 2001, former students of Bloom published a revised Taxonomy using verbs rather than Bloom' s original nouns.[20] These were also listed from low order thinking skills (LOTS) to high order thinking skills (HOTS) to represent the complex process of learning.[20] The revisions published in 2001 serve to aid in further stratifying the domains into smaller subdomains so that we have the ability to microanalyze the efficacy of different learning styles and strategies. For example, instead of simply viewing the aforementioned cognitive domain as the knowledge acquisition domain, the revised terminology stratifies the knowledge dimension into four unique subsets of knowledge. Factual knowledge is defined as being comprised of the ability to list, summarize, classify, order, rank and combine. The other dimensions of knowledge attainment have similar compositions, but the other three types of knowledge headers are as follows: conceptual knowledge, procedural knowledge, and meta-cognitive knowledge.

While the research on Bloom's Taxonomy and dental education could be discussed for many pages and countless hours, it may be useful to provide a general overview of the immediacy of importance that it plays in the research being done here at ULSD. The overarching message is simple, and enhancing student development is the end goal. The ability to navigate through the three domains of learning as described by Bloom begins with the ability to effectively and efficiently establish a knowledge base (cognitive domain). The next step is in the transition from knowledge to intellectual commitment to that knowledge. This is emblematic of the Affective Domain, which we previously described as emotive based. If the information students are receiving in the Cognitive Domain is inconsistent and non-calibrated, the confidence in the student of the knowledge base they received in the Cognitive Domain is compromised. The final

domain of learning is the Psychomotor domain. If the domains prior to this have been invalidated, then the degenerative trend continues into the psychomotor domain, resulting in compromised outcomes and diminished patient care by lesser qualified clinical technicians ultimately.

The research at hand aims to tackle the issue of knowledge acquisition in the Cognitive Domain. By calibrating the faculty prior to and during student development, we can standardize feedback given on preparations. If students receive standardized feedback, they are more apt to truly buy into the feedback. This decreases wasted time sorting through which feedback is trustworthy and allows for a more tangible emotive-based comprehension of the concepts at hand. By improving outcomes in the emotive-based Affective Domain, the opportunities for success in the Psychomotor Domain abound. When these things all fall into place concurrently, the result is better clinical outcomes and foundational knowledge, attitude and skills for both formative and summative assessment of clinical competence are improved.

Another model that is particularly useful for thinking about learning outcomes in relation to assessment of clinical competence is Miller's (1990) pyramid.[22] Developed in 1990 by renowned psychologist Dr. George Miller, this model is similar to Bloom's Taxonomy in that there is a marked shift from being able to demonstrate knowledge that underpins clinical competence to patient application. However, what lends additional credence to Miller's study was that his subjects were clinical physicians. By taking the learning out of the classroom and into the clinic, more advanced learning styles were tapped into. No longer did learning simply involve cognition, now it had an astutely obvious psychomotor counterpart, allowing for deeper indoctrination of learning styles and methods. By considering the underlying thought process introduced by Bloom and integrating the clinical aspects of Miller's study, it lays the foundation for our study involving standardizing intellectual outcomes in order to engender better clinical outcomes. In Dr. Miller's described learners' theory (intellectual skills), psychomotor skills and professional attitudes are synthesized and internalized into a seamless routine that can be carried out in different contexts.[22,23]

Across our dental school curricula, dental students are exposed to both pre-clinical and clinical operative dentistry courses where they receive formative instruction from various dental school faculty. The formative assessment of student performance on operative dentistry terminology, preparations and restorations begin in pre-clinical laboratory sessions through objective grading criteria used by faculty assigned to that course. While this method is perhaps the most efficient method possible in an academic setting, all clinical faculty are not awarded the opportunity to be assigned to the pre-clinical operative dentistry courses. By bringing calibration sessions to clinical faculty asked to grade operative procedures in the clinic, the gap can be closed between graders in appropriate operative dentistry terminology, preparation design and restoration design.

One novel technique to bring real-time calibration instruction to covering faculty is the use of an audience response system. The results from this system can help tailor continuing education topics in areas of weakness noted across the departmental faculty and operative competency examination graders. Additionally, the use of an audience response system could help to improve faculty calibration, clinical assessment and student perception in other areas of general dentistry instruction. It is imperative to understand the progression of learner outcomes through consistent objectives so that calibrated and realistic expectations of our dental student's clinical experiences are established. The progressive transformation of novice provider to competent clinician must include calibrated faculty assessment to ensure a deeper understanding of the knowledge, attitude and skills needed for patient centered care.[19-23]

The purpose of this research project was to calibrate departmental faculty and competency graders' knowledge base in operative dentistry terminology and concepts while providing clinical instruction. By using an audience response system, facilitators are provided with immediate feedback in order to stimulate conversation amongst faculty instructors and competency graders. These discussions may help to further solidify the process and equilibration of clinical opinions amongst faculty. Additionally, these calibration sessions may allow a more calibrated grading assessment during

patient care for student performance feedback. By accomplishing these goals, students in turn could have an improved opinion of objective feedback practices and a more positive perception of operative instruction across faculty and courses. The specific aims for this study were to answer the following research questions: Can the use of one year of faculty calibration sessions using an audience response system:

1. Improve departmental and competency grading faculty scores in a discussion forum?

2. Improve faculty interrater agreement scores during student clinical assessments?

3. Improve student perception of faculty calibrated instruction during formative and summative clinical operative assessments?

METHODS

Sample

This study was approved by the University of Louisville Institutional Review Board (IRB) and determined to be exempt as human subjects research: IRB Tracking # 14.1003. The convenience sample used for this study included all general dentistry departmental faculty (part-time and full-time) assigned to pre-clinical and clinic formative instruction in operative dentistry (n=43). From within this sample, operative dentistry competency graders assigned to summative competency assessments received additional sessions (n=10). An interrater reliability analysis using the Kappa statistic was performed to determine consistency among raters as compared to the Director of Clinical Operative Dentistry (n=25). A convenience sample of D3 and D4 dental students was used to provide perception of faculty calibration across 12 months of implementation (n=100).

Participants

Faculty Participants

Quarterly departmental and competency grader sessions were held in which audience response system calibration sessions took place (November 2014, March 2015, July 2015 and December 2015). A total of eight sessions were held; four departmental and four competency grader sessions. In areas where a non-calibrated consensus occurred, an open dialogue was initiated by the instructor (Director of Clinical Operative Dentistry).

During the departmental calibration sessions, faculty members within the Department of General Dentistry (n=43) responsible for covering daily formative operative procedures were assigned a specific i<clicker to be used across all sessions. A

series of questions were displayed via Microsoft PowerPoint presentation where the participants answered the most appropriate multiple choice answer using their assigned i<clicker technology. The same questions were asked at each of the four sessions. A real-time answer graph was displayed for the instructor (not visible to audience) to determine areas with a lack of consensus. The correct answer was given so that the participants could self-assess and hear the reasoning behind the answer. The goal of this format was to spark conversation and stimulate collegial interactions. It was elected not to display the results to participants to prevent embarrassment when a small minority of faculty members submitted incorrect responses. Question topics included clinical applications of direct restorative materials, pulpal protection, preparation design/ terminology and restoration design via current evidenced-based operative dentistry textbooks. A sample question for preparation design was as follows:

Primary retention form for an ideal class II preparation to be restored with dental amalgam comes from:

A. Flat Pulpal Floor
B. Rounded Axial-pulpal Line Angle
C. Converging Proximal Walls
D. Proximal Retention Grooves

The November 2014 sessional score obtained by each participant was considered their baseline knowledge score (control value).

During the competency grader calibration sessions, all designated competency grading faculty (n=10) within the department of general dentistry performing summative assessments were assigned a specific i<clicker to be used across all sessions. The same process was followed as described for departmental faculty. However, these questions were more specific and focused than the questions used during departmental calibration sessions. Question topics included clinical applications of direct restorative materials, dental material properties, pulpal protection, preparation design/ terminology, restoration design and the paperwork associated with operative competency examinations. A sample question for pulpal protection was as follows:

The primary reason for using calcium hydroxide as an indirect pulp cap is to

A. Provide Sedation
B. Stimulate Dentin Repair
C. Provide Water Insoluble Layer
D. Provide Mechanical Support for Restoration
E. All of the above

The November 2014 sessional score obtained by each participant was considered their baseline knowledge score (control value).

Interrater reliability was evaluated pre and post calibration to assess potential clinical effectiveness. Pre-calibration data was collected by the Director of Clinical Operative Dentistry via two methods: 1. For departmental faculty performing formative assessments of operative dentistry, a second independent score sheet was completed by the Director of Clinical Operative Dentistry on a random sample of 15 operative procedures with 15 different faculty (n=15) in October 2014. 2. For competency graders, scores sheets were pulled for all 10 competency graders from October 2013-2014 where the Director of Operative Dentistry was one grader (n=10). Post-calibration data was collected via the same methods in December 2015 and from October 2014-2015.

Student Participants

One hundred and thirty dental students (n=130) in active clinical care voluntarily completed a questionnaire to evaluate student perception of faculty calibration on operative dentistry concepts. A 10 question Likert scale pre-calibration and post-calibration questionnaire was administered anonymously via audience response system to evaluate student perception of instructional consistency within daily formative (questions 1-5) and summative competency assessments (questions 6-10). The ten questions used for the student questionnaire are located in Table 5. One open-ended question, prompted by the statement “Do you have any further comments” was available for scripted feedback on their perception of grading consistency.

DATA COLLECTION AND ANALYSIS

Faculty Sessional Scores

Data from the i<clicker software was recorded for both departmental faculty calibration sessions (4 sessions) and operative competency grader calibration sessions (4 sessions). The raw data was imported into SPSS (Statistical Package for the Social Sciences, IBM, Chicago, IL) for descriptive and inferential statistical reporting and analysis. Raw faculty calibration session scores were evaluated over four gatherings (1 year) and reported as mean scores ± standard deviations. A separate analysis was performed for the departmental calibration session and the operative competency grader calibration sessions. A one-factor repeated measures analysis of variance (ANOVA) was used to detect any overall differences between related means ($p<.05$). A test for the homogeneity of sphericity assumption was performed. Mauchly's Test of Sphericity tests the null hypothesis that the variances of the differences are equal.[24,27] Thus, if Mauchly's Test of Sphericity is statistically significant ($p < .05$), the null hypothesis can be rejected and the alternative hypothesis is accepted that the variances of the differences are not equal (i.e., sphericity has been violated). A test of the main effect using the Bonferroni correction was performed. Bonferroni correction is a method used to counteract the problem of multiple comparisons and to control the familywise error rate.[25]

Interrater Reliability

Pre and post-calibration data was recorded for both department and competency clinical sessions. The raw data was imported in SPSS to determine interrater reliability using the Kappa statistic to determine consistency among raters. Interrater reliability analysis aims to determine how much of the variance in the observed scores is due to variance in the true scores after the variance due to measurement error between coders has been removed. For example, an interrater reliability estimate of 0.80 would indicate that 80% of the observed variance is due to true score variance or similarity in ratings between coders, and 20% is due to error

variance or differences in ratings between coders. Interrater reliability is not an inferential statistic and therefore can't test a null hypothesis. For categorical data, this may be expressed as the number of agreements in observations divided by the total number of observations. The pre and post calibration data was recorded by two independent raters as superior (3), acceptable (2) or unacceptable (0) on twelve areas of an operative dentistry procedure with the max grade being 36. An interrater reliability analysis using the Kappa statistic was performed to determine consistency among raters with a significant level set as $p<0.05$. A statistical measure of interrater reliability is Cohen's Kappa which ranges generally from 0 to 1.0 (although negative numbers are possible) where large numbers mean better reliability, values near or less than zero suggest that agreement is attributable to chance alone. As a rule of thumb values of Kappa from 0.40 to 0.59 are considered moderate, 0.60 to 0.79 substantial, and 0.80 outstanding.[26] Most statisticians prefer for Kappa values to be at least 0.6 and most often higher than 0.7 before claiming a good level of agreement.

Student Questionnaires

Raw data from the student pre-calibration and post-calibration questionnaires was entered into SPSS for descriptive and inferential statistical reporting and analysis on the individual item level. The open-ended question was reviewed and themes were hand coded using NVivo qualitative software (QSR software) for both pre-calibration and post-calibration surveys. Direct student quotes are entered into the software program, which analyzes responses for specific themes and concepts. The pre-calibration and post-calibration Likert data was recorded as mean responses ± standard deviations on the individual item level. The internal reliability of the instrument was evaluated by using Cronbach's Alpha coefficient. Likert scale responses were coded in SPSS as 1= strongly disagree, 2= disagree, 3= neutral, 4= agree and 5= strongly agree. Therefore, a higher number was associated with a more positive student perception of faculty calibrated instruction. Content validity was ensured by 100% agreeance within the authorship that the construct of student perception was measured. A dependent paired samples t-test was used to compare the mean pre and post calibration scores for

each question with significance level set as $p<0.05$. The open-ended question was reviewed and themes were hand coded for both pre-calibration and post-calibration surveys.

Null Hypotheses Tested

- Null Hypothesis RQ 1: There will be no difference in session scores regardless of quarter reported.
- Null Hypothesis RQ 3: There will be no difference in student perception of faculty calibration regardless of calibration training.

RESULTS

Departmental Calibration Sessions

Descriptive Statistics: There is a general trend in increasing faculty departmental grader i<clicker calibration scores across the quarterly calibration sessions (Table 1). Quarter 1: (79.60 ± 5.49), Quarter 2: (81.98 ± 4.80), Quarter 3 (86.06 ± 5.90) and Quarter 4: (88.46 ± 6.10). The standard deviations of the quarterly mean scores seem to be similar with quarter 1 having the lowest spread in scores and quarter 4 with the highest spread in scores. The largest increase in mean scores seems to occur between quarter 2 and quarter 3.

Inferential Statistics: Mauchly's Test of Sphericity indicated that the assumption of sphericity had been violated, χ^2 (2) = 118.30, p = 0.000. The p value of 0.000 is statistically significant at α=0.05 level. Therefore, the Greenhouse-Geisser test was used with p= 0.511, p> 0.05. This test indicates that the assumption of sphericity has not been met requiring the use of the alternative test to accept the null hypothesis that quarters 1 through 4 share similar variances about their mean quarterly faculty i<clicker score values. A significant main effect was found for departmental calibration training; F (1.534, 64.448) = 125.15 with p= 0.000 at α=.05 level. The effect size was determined to be large at 0.749.

Based on the findings of a significant main effect for departmental calibration training, a pairwise comparison was performed using the Bonferroni correction to control type I error rates. It was determined that all quarters (1-4) were statistically significantly different (p< 0.05) from one another. Different lower case letters in Table 1 represent statistically significant differences in mean scores. There was a statistically significant increase in i<clicker departmental calibration scores at each quarterly training

session. The null hypothesis for research question 1 can be rejected: There is a difference in sessional scores across all quarters reported.

Competency Grader Calibration Sessions

Descriptive Statistics: There is a general trend in increasing faculty competency grader i<clicker calibration scores across the quarterly calibration sessions that appears linear in nature (Table 2). Quarter 1: (83.90 ± 6.38), Quarter 2: (87.60 ± 6.60), Quarter 3 (90.90 ± 5.87) and Quarter 4: (93.80 ± 6.05).The standard deviations of the quarterly mean scores seem to be similar with quarter 3 having the lowest spread in scores and quarter 2 with the highest spread in scores.

Inferential Statistics: Mauchly's Test of Sphericity indicated that the assumption of sphericity had not been violated, χ^2 (2) = 10.044, p = 0.076. The p value of 0.075 is not statistically significant at α=0.05 level. Mauchly's test indicates that the assumption of sphericity has been met and we must accept the null hypothesis that quarters 1 through 4 share similar variances about their mean quarterly faculty i<clicker score values. A significant main effect was found for grader calibration training; F (3,27) = 74.02 with p= 0.000 at α=.05 level. The effect size was determined to be large at 0.892.

Based on the findings of a significant main effect for grader calibration training, a pairwise comparison was performed using the Bonferroni correction to control type I error rates. It was determined that all quarters (1-4) were statistically significantly different (p< 0.05) from one another. Different lower case letters in Table 2 represent statistically significant difference in mean scores. There was a statistically significant increase in i<clicker competency grader calibration scores at each quarterly training session. The null hypothesis for research question 1 can be rejected. There is a difference in sessional scores across all quarters reported.

Interrater Reliability

The interrater reliability analysis was assessed using the Kappa statistic (Tables 3 and 4). The pre-calibration interrater agreement with departmental faculty ranged from as high as 0.85 (grader 1) to as low as 0.15 (grader 13) when compared to the Director of Clinical Operative Dentistry (Table 3). The results indicated that there were two outstanding, three substantial, four moderate, five fair and one poor agreement(s). The post-calibration interrater agreement with departmental faculty ranged from as high as 0.92 (grader 1) to as low as 0.21 (grader 13) when compared to the Director of Clinical Operative Dentistry (Table 3). There were five outstanding, eight substantial, one moderate, one fair and zero poor agreement(s). The general trend in data shows improvement of interrater reliability of the departmental faculty across 12 months of calibration implementation.

The pre-calibration interrater agreement with competency grader faculty ranged from as high as 0.91 (grader 5) to as low as 0.59 (grader 10) when compared to the Director of Clinical Operative Dentistry (Table 4). The results indicated that there were five outstanding, four substantial, one moderate, zero fair and zero poor agreement(s). The post-calibration interrater agreement with competency grader faculty ranged from as high as 0.97 (grader 5) to as low as 0.79 (grader 10) when compared to the Director of Clinical Operative Dentistry (Table 4). There were nine outstanding, one substantial, zero moderate, zero fair and zero poor agreement(s). The general trend in data shows improvement of interrater reliability of the competency grading faculty across 12 months of calibration implementation.

Student Perception Questionnaires

Descriptive Statistics: The paired samples statistics are presented as individual item level means ± standard deviations (Table 5). One hundred and thirty (n=130) dental students participated in the pre-calibration Likert scale questionnaire. One hundred (n=100) of the same dental students participated in post-calibration Likert scale questionnaire as tracked by the i<clicker software. Only the students participating in

both sessions were used in this comparison. The pre-calibration individual question Likert mean scores were paired with the same post-calibration individual question Likert mean scores anonymously using i<clicker registration numbers. The general trend in the pre questions for both formative daily assessments and competency assessments were that the student perception was poor for clinical operative dentistry experiences. Although the pre questions for the competency grading experience were slightly higher than the daily experiences, all are below neutral perception. The general trend shows that all of the post questions showed improvement in student perception across 12 months of implementation. Psychometric evaluation to the reliability of the instrument using Cronbach's Alpha coefficient determined moderate level of internal consistency for the scale used measuring the construct of student perception (0.683).

Inferential Statistics: A paired samples t-test was used to compare the pre-calibration Likert scale responses to the post-calibration responses (n=100) for each individual question. It was determined that for both daily formative (questions 1-5) and summative competency (question 6-10) clinical experiences, all the post-calibration Likert scale responses were statistically significantly higher than the pre-calibration responses ($p< 0.05$). Different lower case letters within each paired question in Table 5 represent a statistically significant difference in mean scores. Likert scale responses were coded in SPSS as 1= strongly disagree, 2= disagree, 3= neutral, 4= agree and 5= strongly agree. Therefore, a higher number was associated with a more positive student perception of faculty calibrated instruction. The results indicate a more positive perception of faculty calibration instruction for clinical operative dentistry instruction by dental students in current active patient care after 12 months of i<clicker faculty calibration training. The null hypothesis for research question 3 can be rejected. There was a significant increase in student perception following 1 year of calibration sessions.

Open-Ended Questions

Of the 100 surveys responded to, there were 22 pre-calibration comments and 41 post-calibration comments. Qualitative analysis of the 22 pre-calibration comments

determined the following words were used most frequent: delayed, confused, unsure, poor instruction and wrong materials. Qualitative analysis of the 41 post-calibration comments determined the following words were most frequent: thanks, efficient, better, consistent, correct sequence and correct materials. Specific comments will be addressed in the discussion section.

DISCUSSION

The results from this study have shown not only an increase in calibration scores at each session from the initial baseline, but a significant increase across twelve months of implementation. The use of the audience response system to engage faculty in real-time discussions of operative dentistry terminology and concepts was shown to be effective in facilitating a discussion forum, calibrating clinical assessment and improving student perception of instruction. Similarly, previously published literature has shown improvement in understanding various dental concepts after some form of calibration discussions.[28-31] Professional faculty engagement through calibration gatherings forced open forum discussions of terminology and concepts that were historically misused or erroneous. Areas of confusion were lessened at each progressive gathering as all faculty started using similar language for instruction. The essence of faculty engagement with professional students is to provide foundational knowledge, attitude and skills for both formative and summative assessment of competence.[1-4] The results of this study show a significant increase in both departmental and competency grader faculty recognition and use of operative dentistry terminology and concepts in a discussion forum.

The results for the interrater reliability showed an increase in both formative and summative evaluations across clinical assessment as well. The Director of Clinical Operative Dentistry was used as the comparison for the interrater reliability. He has 16 years of clinical experience and is Board Certified by the American Board of Operative Dentistry. Additionally, he is recognized as an expert in the field by writing operative dentistry questions for the American Dental Association/ Joint Commission on National Dental Examinations. The faculty knowledge and conceptual understanding carried from the calibration discussion forums into clinical student assessment was crucial. The daily formative assessments were key to students receiving a calibrated and unified clinical

evaluation. All formative and summative evaluators exhibited a more calibrated assessment over the twelve months of implementation during patient centered care.

Just as important, dental student perception to the quality of instruction being received is paramount to their foundational knowledge, attitude and skills.[5-7] Realistic objectives imposed on dental students must be consistent throughout their curricular instruction for them to successfully transition from a fledgling student to a competent health care provider.[19-23] The results from the study show a significant increase in positive student perception to the consistency of the instruction received in operative dentistry for both formative and summative assessment. All ten Likert scale questions were evaluated at the pre-calibration level and twelve months after its implementation (post-calibration). All ten Likert scale responses were significantly more positive following implementation of faculty calibration. This information suggests that the inconsistency in operative dentistry instruction was not limited to a few faculty but woven throughout the department. For students to adequately provide an accurate self-assessment of their performance, a consistent instruction is paramount to improve perception of learning needs, promote change in learning activity, and improve clinical practice and patient outcomes.[32]

Some of the pre-calibration survey comments were:

- "Dr. ___ send me to the window for Durelon (zinc phosphate cement) to place a base in my class II preparation. We don't use that material or place bases at the dental school."
- "Dr. ___ told me that there is no retention needed for a class III resin composite when you have clearly taught us that it does."
- "My group manger keeps referring to Dycal (calcium hydroxide) as a base when you taught us that it is too brittle to be a base."
- "Dr. ___ does not understand the application of Hibiclens (chlorhexidine gluconate) in the sequence of restoration placement. He says you do it before removing the smear layer!!!!!"

- "Dr. ___ told a patient that dental amalgam is toxic and that resin composite is the best material for all restorations."
- "I was told by a covering faculty that pin placement for retention is malpractice and should be banned from dental education."
- "Dr. _ told me that resin-modified glass ionomer (Fuji II LC) can be used for all restorations, even under significant occlusal load."
- "According to Dr. _, rubber damn placement is not necessary in the real world as it slows you down."
- "Drs. _ and _ are seriously confused when it comes to qualifying operative lesions for a class II competency. One said yes and the other said no! Me and my patient were both confused."

From these comments, a clear vision can be acquired to the problem that existed amongst clinical instruction of our students. A major obstacle was to remove personal opinion from the covering faculty minds and replace it with evidence-based teaching protocol from quality peer-reviewed publications during calibration. Henzi et al. (2006) found that this student perception of inconsistency in instruction occurs across the nation and posit that calibration is crucial for success.[33]

Some of the post-calibration survey comments were:

- "I feel like the instructors I work with now understand the concepts taught in our operative dentistry curriculum. Dr. ___, thanks for teaching the faculty to be consistent during clinic time. It makes the appointment run smoothly and I feel like I am actually learning something."
- "There has been a significant improvement in the understanding of the materials used at the school for operative dentistry."
- "It helps so much that the instructors get the same information as we do."
- "I no longer feel apprehensive asking for material at the window because the instructors know what we use."
- "During operative competency examinations, the faculty are more in sync with qualifying lesions clinically and radiographically."

- "All faculty using the same terminology for preparation modifications helps competency examinations run smoothly and efficiently."

From these comments, operative dentistry calibration is now more positively perceived by the students in a providing them a consistent clinical experience.

It is possible that other confounding variables could be the reason for the results obtained in this study. To eliminate as many confounding variables as possible, the calibration sessions were all held in the same classroom at the same time of the day with the same instructor. However, the faculty were not blinded to the study and could have memorized the concepts while not fully understanding them. The faculty could also have consulted a neighbor for the answer without fully understanding key principles. During the interrater reliability evaluations, the faculty could purposefully have decided to grade more like the director that day. The students could have over self-reported their perception of instructional consistency trying to please the faculty. These and many more biases could have occurred but all attempts were made to adequately control the study. Statistically, type I errors were controlled for using Bonferroni correction and tested assumptions during the One-Factor ANOVA comparing session scores. The clinical implications of the sessions were evaluated with interrater reliability using the Director of Clinical Operative Dentistry for comparison. Additionally, the internal reliability and content validity of the Likert scale student questionnaire were evaluated in measuring the construct of student perception.

As reported in current literature, [5,7,11,15] the elimination of counter-teaching and/or misusing terminology and conceptualization has shown improve deep understanding. Our initial results using an audience response system have shown promising results as well. Professional faculty engagement through real-time interactions has appeared to be beneficial in calibrating faculty members both in a discussion forum and in clinical instruction.[13-17] In turn, student perception was shown to become more positive towards reception of clinical instruction. With the results from this study, a continued quarterly training program will be a vital part of professional

faculty development for both full and part-time faculty at our institution in all disciplines.

CONCLUSION

The implementation of an audience response system for calibrating both departmental and competency graders in operative dentistry terminology and concepts has shown to be effective across twelve months of training. Clinical interrater reliability has been shown to improve for both formative and summative clinical competency assessments. Additionally, student perception to the quality and consistency of operative dentistry clinical instruction was shown to become more positive across twelve months of training. It is paramount that all dental schools continue to provide a trackable, vested and profound professional development program to ensure consistent instruction for assessing dental student competence.

REFERENCES

1. Colthart, I., Bagnall, G., Evans, A., Allbutt, H., Haig, A., Illing, J., & McKinstry, B. (2008). The effectiveness of self-assessment on the identification of learner needs, learner activity, and impact on clinical practice: BEME Guide no. 10. *Med Teach, 30*(2), 124-145.
2. Koffel, J., & Reidt, S. (2014). An interprofessional train-the-trainer evidence-based practice workshop: design and evaluation. *J Interprof Care*, 1-3.
3. Steinert, Y., Mann, K., Centeno, A., Dolmans, D., Spencer, J., Gelula, M., & Prideaux, D. (2006). A systematic review of faculty development initiatives designed to improve teaching effectiveness in medical education: BEME Guide No. 8. *Med Teach, 28*(6), 497-526.
4. Steinert, Y., Naismith, L., & Mann, K. (2012). Faculty development initiatives designed to promote leadership in medical education. A BEME systematic review: BEME Guide No. 19. *Med Teach, 34*(6), 483-503.
5. Hewson MG. Clinical teaching in the ambulatory setting. J Gen Intern Med. 1992;7:76–82.
6. Novack DH, Suchman AL, Clark W, Epstein RM, Najberg E, Kaplan C. Calibrating the physician: personal awareness and effective patient care. JAMA. 1997;278:502–9.
7. Chin R, Benne KD. General strategies for effecting changes in human systems. In: Bennis WG, Benne KD, Chin R, Corey KE (eds). The Planning of Change. 3rd ed. New York: Holt, Rinehart and Winston, 1976.
8. Knowles M. The Adult Learner: A Neglected Species. 4th ed. Houston, TX: Gulf Publishing, 1990.
9. Kolb DA. Experiential Learning: Experience as the Source of Learning and Development. Englewood Cliffs, NJ: Prentice—Hall, 1984.
10. Schon DA. Educating the Reflective Practitioner. San Francisco, CA: Jossey—Bass, 1987.

11. Harris IB. New expectations for professional competence. In: Curry L, Wergin JF, et al (eds). Educating Professionals: Responding to New Expectations for Competence and Accountability. San Francisco, CA: Jossey—Bass, 1993.
12. Levinson W, Branch WT, Kroenke K. Clinician—educators in academic medical centers: a two-part challenge. Ann Intern Med. 1998; 29:59–64.
13. Roberts KB, DeWitt TG. Faculty development of pediatric practitioners: complexities in teaching clinical precepting. Pediatrics. 1996;97:389–93.
14. Irby DM. Teaching and learning in ambulatory care settings: a thematic review of the literature. Acad Med. 1995;70:898–931.
15. Wilkerson LA, Irby DM. Strategies for improving teaching practices: a comprehensive approach to faculty development. Acad Med. 1998;73:387–96.
16. Copeland HL, Hewson MG. Developing and testing a clinical teaching effectiveness instrument in an academic medical center. Acad Med. 2000;75:161–6.
17. Hewson MG, Rehm SJ. Onsite faculty development—coaching clinical teaching. J Gen Intern Med. 1996;11(4 suppl):101.
18. Hewson MG, Copeland HL. Outcomes assessment of a faculty development program in medicine and pediatrics. Acad Med. 1999;74, 10 suppl:S68–S71.
19. Bloom B.S. (1956). Taxonomy of Educational Objectives, Handbook I: The Cognitive Domain. New York: David McKay Co Inc.
20. Anderson, L.W., Krathwohl, D.R., Airasian, P.W., Cruikshank, K.A., Mayer, R.E., Pintrich, P.R., Raths, J., Wittrock, M.C. (2001). A Taxonomy for Learning, Teaching, and Assessing: A revision of Bloom's Taxonomy of Educational Objectives. New York: Pearson, Allyn & Bacon.
21. Krathwohl, D.R., Bloom, B.S., Masia, B.B. (1973). Taxonomy of Educational Objectives, the Classification of Educational Goals. Handbook II: Affective Domain. New York: David McKay Co., Inc.
22. Miller GE. The assessment of clinical skills/competence/performance. Acad Med 1990:S63-7.

23. McKinley RK, Fraser RC, Baker R. Model for directly assessing and improving competence and performance in revalidation of clinicians. BMJ 2001;322:712-5
24. Mauchly, J. W. (1940). Significance Test for Sphericity of a Normal n-Variate Distribution. The Annals of Mathematical Statistics 11 (2): 204–209.
25. Strassburger, K.; Bretz, Frank (2008). Compatible simultaneous lower confidence bounds for the Holm procedure and other Bonferroni-based closed tests. Statistics in Medicine 27 (24): 4914–4927.
26. Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. Biometrics, 33, 159-174.
27. Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. Psychometrika, 24, 95-112.
28. Reem Haj-Ali and Philip Feil. Rater Reliability: Short- and Long-Term Effects of Calibration Training. J Dent Educ 2006 70:428-433
29. Jay A. Morrow, M. Teresa Pulido, P. Bradford Smith, Thomas F. McDaniel, and Austin B. Willcox. Effective Use of e-Grading in the Dental Simulation Clinic. J Dent Educ 2014 78:829-837
30. Vanchit John, Seung-Jun Lee, Sivaraman Prakasam, George J. Eckert, and Gerardo Maupome. Consensus Training: An Effective Tool to Minimize Variations in Periodontal Diagnosis and Treatment Planning Among Dental Faculty and Students. J Dent Educ 2013 77:1022-1032
31. Peter Rechmann and John D.B. Featherstone. Quality Assurance Study of Caries Risk Assessment Performance by Clinical Faculty Members in a School of Dentistry. J Dent Educ 2014 78:1331-1338
32. Colthart I, Bagnall G, Evans A, et al. The effectiveness of self-assessment on the identification of learner needs, learner activity, and impact on clinical practice: BEME Guide no. 10. Med Teach. 2008;30(2):124-45.
33. Henzi D, Davis E, Jasinevicius R, Hendricson W. North American dental students' perspectives about their clinical education. J Dent Educ. 2006;70(4):361-77.

34. Kwon SR, Restrepo-kennedy N, Dawson DV, et al. Dental anatomy grading: comparison between conventional visual and a novel digital assessment technique. J Dent Educ. 2014;78(12):1655-62
35. Lane BA, Luepke P, Chaves E, et al. Assessment of the calibration of periodontal diagnosis and treatment planning among dental students at three dental schools. J Dent Educ. 2015;79(1):16-24.
36. 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):545-57.
37. McAndrew R, Chadwick B, Treasure ET. The influence of a short training program on the clinical examination of dental restorations. Oper Dent. 2011;36(2):143-52.
38. Partido BB, Jones AA, English DL, Nguyen CA, Jacks ME. Calculus detection calibration among dental hygiene faculty members utilizing dental endoscopy: a pilot study. J Dent Educ. 2015;79(2):124-32.

TABLES

Table1: Descriptive Statistics for Faculty Departmental Grader i<clicker Calibration

Faculty Departmental i<clicker Calibration Scores			
	Mean Scores	Standard Deviations	N
Quarter 1 (Control Value)	79.60 (a)	5.49	43
Quarter 2	81.98 (b)	4.80	43
Quarter 3	86.06 (c)	5.90	43
Quarter 4	88.46 (d)	6.10	43

*Different lower case letters represent significant different mean values using a pairwise comparison with Bonferroni correction ($p < 0.05$).

Table 2: Descriptive Statistics for Faculty Competency Grader i<clicker Calibration

Faculty Competency i<clicker Calibration Scores			
	Mean Scores	Standard Deviations	N
Quarter 1 (Control Value)	83.90 (a)	6.38	10
Quarter 2	87.60 (b)	6.60	10
Quarter 3	90.90 (c)	5.87	10
Quarter 4	93.80 (d)	6.05	10

*Different lower case letters represent significant different mean values using a pairwise comparison with Bonferroni correction ($p < 0.05$).

Table 3: Departmental Interrater Reliability Scores (n=15)

Pre-Calibration Departmental Faculty Kappa Scores				**Post- Calibration Departmental Faculty Kappa Scores**			
	Director Grader	**Agreement Value**	**95% CI p<0.05 Significant**		**Director Grader**	**Agreement Value**	**95% CI p<0.05 Significant**
Grader 1	0.85	Outstanding	Yes	**Grader 1**	0.92	Outstanding	Yes
Grader 2	0.43	Moderate	No	**Grader 2**	0.74	Substantial	Yes
Grader 3	0.38	Moderate	No	**Grader 3**	0.79	Substantial	Yes
Grader 4	0.25	Fair	No	**Grader 4**	0.65	Substantial	Yes
Grader 5	0.65	Substantial	Yes	**Grader 5**	0.78	Substantial	Yes
Grader 6	0.71	Substantial	Yes	**Grader 6**	0.81	Outstanding	Yes
Grader 7	0.21	Fair	No	**Grader 7**	0.69	Substantial	Yes
Grader 8	0.35	Fair	No	**Grader 8**	0.62	Substantial	Yes
Grader 9	0.88	Outstanding	Yes	**Grader 9**	0.95	Outstanding	Yes
Grader 10	0.74	Substantial	Yes	**Grader 10**	0.82	Outstanding	Yes
Grader 11	0.43	Moderate	No	**Grader 11**	0.76	Substantial	Yes
Grader 12	0.29	Fair	No	**Grader 12**	0.58	Moderate	No
Grader 13	0.15	Poor	No	**Grader 13**	0.21	Fair	No
Grader 14	0.29	Fair	No	**Grader 14**	0.72	Substantial	Yes
Grader 15	0.53	Moderate	No	**Grader 15**	0.81	Outstanding	Yes

Table 4: Competency Grader Interrater Reliability Scores (n=10)

Pre-Calibration Competency Graders				**Post- Calibration Competency Graders**			
	Director Grader	**Agreement Value**	**95% CI p<0.05 Significant**		**Director Grader**	**Agreement Value**	**95% CI p<0.05 Significant**
Grader 1	0.79	Substantial	Yes	**Grader 1**	0.85	Outstanding	Yes
Grader 2	0.82	Outstanding	Yes	**Grader 2**	0.92	Outstanding	Yes
Grader 3	0.75	Substantial	Yes	**Grader 3**	0.88	Outstanding	Yes
Grader 4	0.83	Outstanding	Yes	**Grader 4**	0.94	Outstanding	Yes
Grader 5	0.91	Outstanding	Yes	**Grader 5**	0.97	Outstanding	Yes
Grader 6	0.87	Outstanding	Yes	**Grader 6**	0.91	Outstanding	Yes
Grader 7	0.71	Substantial	Yes	**Grader 7**	0.88	Outstanding	Yes
Grader 8	0.86	Outstanding	Yes	**Grader 8**	0.95	Outstanding	Yes
Grader 9	0.74	Substantial	Yes	**Grader 9**	0.91	Outstanding	Yes
Grader 10	0.59	Moderate	No	**Grader 10**	0.79	Substantial	Yes

Table 5: Descriptive Statistics for Likert Student Questionnaire (n=100)

Paired Samples T-Test Student Perception Likert Questions				
Questions		**Mean**	**SD**	**N**
1. During daily dental operative dentistry procedures, the covering faculty are consistent in their understanding and instruction of direct dental materials.	Pre-Calibration	1.70 (a)	.46	100
	Post-Calibration	3.50 (b)	.50	100
2. During daily dental operative dentistry procedures, the covering faculty are consistent in their understanding and instruction of pulpal protection (liners and sealers).	Pre-Calibration	1.90 (a)	.70	100
	Post-Calibration	3.50 (b)	.50	100
3. During daily dental operative dentistry procedures, the covering faculty are consistent in their understanding and instruction of preparation design (retention and resistance forms).	Pre-Calibration	1.90 (a)	.70	100
	Post-Calibration	3.50 (b)	.50	100
4. During daily dental operative dentistry procedures, the covering faculty are consistent in their understanding and instruction of auxiliary retention (pins, slots and pots).	Pre-Calibration	1.90 (a)	.70	100
	Post-Calibration	3.50 (b)	.50	100
5. During daily dental operative dentistry procedures, the covering faculty are consistent in their understanding and instruction of final restoration design (anatomy, contours and contacts).	Pre-Calibration	1.90 (a)	.70	100
	Post-Calibration	4.50 (b)	.81	100
6. During operative competency examinations, the covering graders are consistent with qualification criteria (radiographic and clinical indications).	Pre-Calibration	2.60 (a)	.49	100
	Post-Calibration	4.40 (b)	.80	100
7. During operative competency examinations, the covering graders are consistent with terminology used for preparation modifications.	Pre-Calibration	2.50 (a)	.50	100
	Post-Calibration	4.30 (b)	.90	100
8. During operative competency examinations, the covering graders are consistent with their expectations of pulpal protection.	Pre-Calibration	2.50 (a)	.50	100
	Post-Calibration	4.20 (b)	.87	100
9. During operative competency examinations, the covering graders are consistent with their expectations of final restoration design (anatomy, contours and contacts).	Pre-Calibration	2.70 (a)	.78	100
	Post-Calibration	3.80 (b)	.75	100
10. During operative competency examinations, the covering graders are consistent with overall grading and outcomes assessment.	Pre-Calibration	2.80 (a)	.98	100
	Post-Calibration	4.10 (b)	.54	100

*Different lower case letters within each pair represent significant different mean values using a paired samples t-test ($p < 0.05$).

CURRICULUM VITA

NAME: Sean Alexander Aiken

ADDRESS: University of Louisville School of Dentistry
501 South Preston St
Louisville, KY 40205

DOB: Atlanta, Georgia - May 3, 1990

EDUCATION
& TRAINING: B.S., Kinesiology & Health Sciences
The College of William & Mary
2008-12

M.S., Physiology & Biophysics
University of Louisville School of Medicine
2012-13

MSOB Candidate, Oral Biology
University of Louisville School of Dentistry
2014-16

AWARDS: James C. Moore Scholarship for excellence in Physiology (2015)

PROFESSIONAL SOCIETIES:

ASDA Representative to the 2016 ADA House of Delegates – District 7
Student Member of the Kentucky Dental Association Executive Board
LASDA Legislative Liaison - University of Louisville
Student Member of the American Academy of General Dentistry
ASDA Board Member - District 7 Trustee

PUBLICATIONS:

1) Miller CJ, Aiken S, Metz MJ. Perceptions of D.M.D. Student Readiness for Basic Science Courses in the United States: Can Online Review Modules Help? European Journal of Dental Education. (May 2013)
2) Metz MJ, Durski MT, Metz CJ, Aiken SA, Mayfield TG and Lin WS: Using an Audience Response System to Calibrate Dental Faculty Assessing Student Clinical Competence. *Journal of Dental Education*.