A Pilot Study Investigating the Effect of the Supervision-Questioning-Feedback Model of Supervision on Stimulating Critical Thinking in Speech-Language Pathology Graduate Students

Samantha J. Dalessio (Procaccini)
Carlow University, sdalessio@carlow.edu

Nancy Carlino
California University of Pennsylvania, carlino@calu.edu

Mary G. Barnum
Springfield College, mbarnum@springfieldcollege.edu

A Pilot Study Investigating the Effect of the Supervision-Questioning-Feedback Model of Supervision on Stimulating Critical Thinking in Speech-Language Pathology Graduate Students

Cover Page Footnote

Authors

Samantha J. Dalessio (Procaccini), Nancy Carlino, Mary G. Barnum, Denise Joseph, and Melissa M. Sovak

Introduction

Discussions surrounding the significance of supervisory methodologies in speech-language pathology and audiology have dated as far back as the 1950s (Anderson, 1988; Cogan, 1972; Dudding et al., 2017; Goldhammer et al., 1980). Since that time, preliminary discussions regarding supervisory methodologies have evolved into formalized position statements (American Speech-Language-Hearing Association [ASHA], 1985; ASHA, 2008a), national committees (ASHA, 2013; ASHA, 2016), and acceptance that clinical supervision is a professional specialty that requires its own representative knowledge and skill set (ASHA, 2008c; ASHA, 2013; Council of Academic Programs in Communication Sciences and Disorders [CAPCSD], 2013; Dudding et al., 2017). Most recently, the growing recognition of the complexities surrounding efficacious supervisory practices has prompted formalized training requirements for speech-language pathologists and audiologists who engage in clinical teaching (ASHA, 2016; Council for Clinical Certification in Audiology and Speech-Language Pathology of the American Speech-Language-Hearing Association [CFCC], 2018; Procaccini, et al., 2017).

Recent developments within the scope of clinical supervision have not only highlighted the complexities involved with fostering a successful clinical teaching-learning environment but have also underscored a need for further research investigating effective clinical teaching practices (ASHA, 2016; Dudding et al., 2017). Current available research surrounding efficacious models of clinical teaching in speech-language pathology is limited. Most supervisory approaches implemented within the clinical teaching-learning context of speech-language pathology and audiology have been influenced primarily by one supervisory approach, Anderson's continuum model (Anderson, 1988; ASHA, 2008b; Dudding et al., 2017). Anderson's continuum model has been widely favored because its theoretical underpinnings strongly emphasize the importance of modifying the supervisor's style in response to the needs, knowledge, and skills of the supervisee at each stage of clinical development (ASHA, n.d.; Anderson, 1988).

Although research specific to clinical teaching models within speech-language pathology and audiology is limited, there is more available evidence in related clinical disciplines. A model of clinical teaching developed outside the scope of speech-language pathology and audiology that, like Anderson's (1988) continuum model, views clinical teaching on a continuum is the supervision-questioning-feedback (SQF) model. The SQF model was developed to provide athletic training preceptors with a practical framework to integrate appropriate supervision, questioning, and feedback to students (Barnum et al., 2009). The SQF model is grounded in the concept that clinical learning is experiential learning, and that clinical teaching requires a different approach to teaching than what is used in the traditional classroom setting. Students participating in clinical practicums or fieldwork are learning in rich, dynamic, complex work-like settings. The quality, complexity, and depth at which the experiences are cognitively processed cannot be guaranteed just because a student participates in the experience (Dewey, 1938; Wiedner et al., 1997). Therefore, the clinical educator needs to act as a guide, providing support, direction, challenges, and feedback as needed to move the student through the experience (Barnum, 2008; Mitchell & Poutiatine, 2001; Wiedner et al., 1997; Willeford et al., 2009).

When examining experiential learning models, three dimensions are constant:
1. The clinical educator serves as a facilitator and supervisor of learning, coming in and out of the learning experience as needed to assist the student in reaching the outcomes identified for that specific experience.
2. Clinical educators ask questions throughout the learning experience for the expressed purpose of stimulating the student to think or process information in a specific way or for a specific outcome.
3. Feedback is given at different times, in different ways, for different reasons (Barnum et al., 2009).

The SQF model hypothesizes that critical thinking and metacognitive knowledge may more likely be stimulated when the supervisor intentionally adjusts supervision, questioning and feedback styles to the needs and skill level of the student.

Within the SQF model, situational supervision is used to match the level of supervision provided by the clinical educator with the specific situation in which the student is engaged. The more complex, urgent, or novel the experience, the closer the supervision. The level of supervision is decreased when the situation is less complex, non-urgent, or if the student has had multiple exposures/interactions with the content (Levy et al., 2009). Supervision styles consists of S1 (directing and coaching), S2 (supporting and encouraging) and S3 (delegating) (Barnum & Guyer, 2016; Levy et al., 2009). The type, frequency, and timing of questions posed by the clinical educator also vary within the SQF model.

Clinical educators using the SQF model are encouraged to develop strategic questioning patterns utilizing Bloom's Taxonomy of Educational Objectives to generate and sequence questions (Anderson et al., 2001; Barnum, 2008; Bloom, 1956). Barnum (2008) defined strategic questioning as consciously adapting the timing, sequencing, and phrasing of questions in order to actively engage and stimulate student use of increasingly complex cognitive processing skills. Strategic questioning builds a foundation by first targeting basic declarative, comprehension, and application knowledge that allow both the student and the clinical educator to gain an awareness of the students' knowledge and skill base. Asking low level questions is also thought to help build student confidence (Barnum, 2008). Follow-up questions target higher-level cognitive processes appropriate for the students' academic knowledge, skill level, past experience, and competency. Questions that target higher-level cognitive processes are thought to assist students in developing a process for thinking and to facilitate critical analysis (Barnum, 2008). Barnum, Guyer, Levy and Graham (2009) identified a three level system to use when creating strategic questioning: Q1, Q2 and Q3 level questions.

Q1 level questions target the factual and conceptual dimension of knowledge. Q1 questions require the student to recall, recite, and explain basic, foundational information needed to engage in a specific discussion, activity, or interaction. The purpose of asking Q1 level questions is to confirm for both the student and the clinical educator that the student has the basic factual knowledge base needed to engage and safely continue (e.g., List the typical features of acquired adult apraxia of speech; List the distinguishing features of apraxia and aphasia when making a differential diagnosis.).

Q2 level questions target the conceptual and procedural dimension of knowledge. Q2 questions require the student to use and apply information appropriately. The purpose of asking Q2 level questions is to confirm for both the student and the preceptor that the student is making appropriate connections and correctly applying/utilizing information (e.g., Based on the objective data you have collected thus far, which features of apraxia of speech have you connected to your case? Also using your objective observations from your case, compare and contrast features of apraxia and aphasia.).

Q3 level questions target metacognition skills. Q3 level questions require the student to explain/support choices and actions and to think through their own thought process. It is important to inquiry about "how" and "why" the student arrived at a specific decision to stimulate reflective practice. The purpose of asking Q3 level questions is to provide opportunity for students to develop and practice cognitive processing skills vital for developing sound clinical reasoning abilities (e.g., After critically appraising the evidence, what are some of controversies behind definitions of apraxia and aphasia? What are some of the limitations of the current available evidence describing definitions of acquired apraxia? What are some of the factors that may have disrupted the integrity of the available evidence?).

The final component of the SQF model is feedback. Feedback is any information that is given to a student regarding their skills and knowledge and can be delivered via verbal, written, or behavioral transmission. Quality feedback is dependent upon the content, timing, specificity, form of the feedback and the arena (private or public) in which the feedback is delivered (Nottingham & Henning, 2014). Three basic modes of feedback are used within the SQF model: confirming (correct application of knowledge and skills), corrective (incorrect application of knowledge and skills), and guiding (refinement of knowledge and skills).

Correlation of SQF to Critical Thinking. Critical thinking is seen as essential for healthcare providers as the foundation for developing sound clinical and diagnostic reasoning abilities (Kicklighter et al., 2018; Papathanassiou et al., 2014; Sharp et al., 2013). The ability to think critically within the scope of speech-language pathology and audiology is no longer considered an advanced skill set but increasingly accepted as a required competency needed to provide the highest evidence-based quality of care services. The American-Speech-Language-Hearing Association has placed strong emphasis on the successful development of critical thinking skills in new learners (ASHA, 2015). Furthermore, successfully stimulating critical thinking within the clinical teaching-learning environment has been recognized as an increasingly important competency for clinical educators in speech-language pathology and audiology (ASHA, 2008c; ASHA, 2013; CAPCSD, 2013).

Unfortunately, attempts to universally define critical thinking have not resulted in a standardized cross-disciplinary accepted definition (Finn et al., 2016; Mulnix, 2012). Finn et al. (2016) cited Davies (2015) when they suggested that while "most definitions of critical thinking reported across the literature share a compelling family resemblance" it is likely presumptuous to believe that a universally accepted definition will be reached "given differences in disciplinary focus and theoretical orientation" (p. 44). Despite a lack of agreement in definition, Finn et al. (2016) state the importance of identifying an "instructional definition" that can be used consistently within a context such that expectations surrounding adequate acquisition of critical thinking are mutually

transparent and agreed upon (p. 44). Specific to the scope of speech-language pathology and audiology, Finn et al.(2016) suggested using the following definition by Wade et al. (2014), "Critical thinking is the ability and willingness to access claims and make objective judgments on the basis of well-supported reasons and evidence rather than emotion or anecdote…it includes the ability to be creative and constructive…" (pp. 6-7). Critical thinking involves evaluating presented information for comparison with already held knowledge in order to formulate well-supported solutions, and new or different perspectives and options. As an instructional strategy, promoting the use of critical thinking provides an opportunity for students to process information multiple times, on multiple levels, and supports the retrieval of information from long-term memory stores and rehearsal of information while in the working memory for comparison with newly acquired information (Clark & Harrelson, 2002).

Many researchers agree that in order to promote the development of clinical proficiency and critical thinking, the instructor needs to be adept at selecting and using a variety of questioning styles and teaching strategies to better assist the student in clarifying, identifying and evaluating information gained from experiences (Borton, 1970; Davies, 1995; Joplin, 1995; Mensch & Ennis, 2002; O'Conner, 2001). Given that the SQF model intends to tap into critical thinking by strategically structuring questions to stimulate higher level thinking (e.g. analysis, evaluation, metacognition), one may argue that the SQF model can viewed as a potential vehicle for stimulating critical thinking. Furthermore, the use of higher-level strategic questioning techniques as a means for stimulating critical thinking and other higher order thinking skills has been cited in the literature (Hausmann & Schwartzstein, 2019; Toledo, 2015). Additionally, the SQF model supports a developmental continuum in attaining knowledge and skills proficiency because it intentionally scaffolds supervision, questioning, and feedback according to the needs and skill level of the learner. Finn et al. (2016) recognized that attaining proficiency as a critical thinker requires practice, progression through a series of developmental stages, and often, a time commitment.

Purpose/Hypotheses. The SQF model was developed using a grounded theory approach to provide clinical instructors with a system for assisting students in developing a process for thinking and enhancing critical thinking and clinical reasoning. Although SQF model has been used successfully in athletic training and is growing in popularity amongst clinical educators in speech-language pathology, the benefits of this model have not been systematically studied. Furthermore, there are no existing studies substantiating the ability for the SQF model to stimulate critical thinking within a clinical teaching environment, including within the discipline of athletic training. The purpose of this study was to investigate the effects of the SQF model on students' critical thinking. The researchers hypothesized that students who received the SQF model of supervision would score higher than students who received the non-SQF style (NSQF) of supervision on the selected critical thinking measures.

Methods

Design. This mixed randomized control trial and prospective cohort study design was approved by the California University of Pennsylvania Institutional Review Board between June 2015 and June 2016. The data collection period was between September 2015 and December 2015, a

duration of 1 academic semester. All participants provided informed written consent prior to participation.

Participants. All 48 graduate students, 24 first semester graduate students and 24 fourth semester graduate students, enrolled full-time in the master's program in speech-language pathology at California University of Pennsylvania were considered for the study. Graduate students attending California University of Pennsylvania complete a clinical practicum within the on-site University Speech & Hearing Clinic for semesters 1, 2, and 3 and then complete off-site externship experiences in semesters 4 and 5. As such, all 24 first semester students considered for the study were completing their clinical practicums within the University Speech & Hearing Clinic. Fourth semester students were completing their first clinical practicums off-site within a school or medical based externship setting for a minimum of 28 hours per week. Fourth semester students had completed 36 credits of academic and clinical coursework at the beginning of the study period.

All 5 on-site clinical supervisors within the Speech & Hearing Clinic and 24 off-site clinical supervisors within school or medically based settings were considered for the study. All on-site and off-site clinical supervisors affiliated with California University of Pennsylvania were state licensed and/or possessed state teaching certification (depending on work setting), were certified by the American Speech-Language-Hearing Association, had a minimum of 5 years of clinical experience, and a minimum of 1 year of clinical supervisory experience.

All 5 clinical supervisors within the University Speech & Hearing Clinic (n=5) and 17/24 first semester students (n=17) consented to participate in the study. Of the 5 on-site supervisors, 3 were selected to be SQF supervisors and 2 were selected to be NSQF supervisors. Of the 17 participating first semester students, 9 were assigned to 1 of the 3 SQF-trained supervisors by simple randomization. The other 8 students were assigned to 1 of the 2 NSQF trained supervisors also by simple randomization. Mean age at the start of the study period for participating first semester SQF students and NSQF students was 22.22 (SD =.63) and 22.25 (SD =.83) years, respectively. Mean score on the Graduate Record Examination (GRE) for first semester SQF students and NSQF students was 290.44 (SD =9.44) and 292.38 (SD = 3.53), respectively. Mean grade point average for first semester SQF students and NSQF students was 3.87 (SD = .10) and 3.70 (SD = .16), respectively.

Selection criteria for determining on-site SQF and NSQF supervisors was based on previous knowledge and experience with the SQF model. Of the 5 on-site supervisors who consented to participate, 3 had previous knowledge and experience pertaining to the SQF model. Specifically, 1/3 on-site SQF selected supervisors had completed an additional SQF workshop prior to the study and had engaged in scholarship activities related to the SQF model. While the other 2 on-site SQF selected supervisors did not complete any previous continuing education or training related to the SQF model, they also engaged in previous scholarship activities related to the method. The 2 on-site NSQF selected supervisors had no previous continuing education or training in the SQF model and had limited knowledge of the method. Years of general supervisory experience for on-site SQF supervisors, on-site NSQF supervisors, and off-site SQF supervisors ranged from 7-30 years (M = 20.7, SD= 9.9), 14->30 years (M = 22, SD = 8), and >1->20 years (M = 7.5, SD = 7.4), respectively.

Out of the 24 fourth semester students, 3 students (n=3) and their off-site supervisors (n=4) consented to participate in the study. In order to be included in the study both off-site supervisor and student had to consent to participate. Given the small sample size, all 3 fourth semester students were assigned to an SQF-trained supervisor. Two of the 3 participating fourth semester students each had one SQF trained supervisor at their respective externship sites. Due to student and supervisor scheduling, one of the 3 participating students had two SQF trained supervisors at her assigned externship site. Other than completing the SQF training for this pilot study, all 4 participating off-site supervisors had no additional continuing education or training in the SQF model and had limited previous knowledge of the method. Mean age at the start of the study period, score on the Graduate Record Examination (GRE), and grade point average for participating fourth semester SQF students were 23.33 ($SD = .47$), 294 ($SD = 2.16$), and 3.72 ($SD = .15$), respectively.

Four additional study participants were selected and consented to serve as independent SQF raters. Selected SQF raters were charged with the task of analyzing video recorded student-supervisor conferences to determine whether or not the SQF model of supervision was being implemented. Selection criteria for 3/4 raters was based on clinical and supervisory knowledge and experience. Three out of the 4 consented raters had a minimum of 10 years of clinical and supervisory experience in speech-language pathology. Of the 3 selected raters, 2 had clinical supervisory experience within the academic setting, and the other rater had extensive clinical supervisory experience within the externship setting. One out of the 3 selected raters had limited prior knowledge and experience related to the SQF model while the other 2 raters had some prior knowledge and experience pertaining to the SQF model. The SQF developer who provided the SQF training workshop for this study served as an additional independent fourth rater in order to assess agreement in how the videos were analyzed by the other 3 independent raters.

SQF Training. Approximately 4 weeks prior to the intervention phase, a total of 3 on-site clinical supervisors, 4 off-site supervisors, and the 3 selected SQF raters completed a 4-hour SQF training workshop. The SQF workshop was led by one of the initial developers of SQF and a clinical educator within speech-language pathology who had experience with implementing SQF. The SQF workshop comprised of lecture-based content information and hands-on role play. At the completion of the workshop, all participants completed a 15 question post-SQF training assessment quiz. All participants were required to receive a minimum score of 80% on the quiz.

Selection of Critical Thinking Measures. Two measures of critical thinking were selected to evaluate students' baseline and post critical thinking skills:

- California Critical Thinking Skills Test (CCTST)
- Simucase® clinical simulation case studies.

The California Critical Thinking Skills Test (CCTST) is a valid and reliable discipline neutral instrument used to measure critical thinking for undergraduate and graduate level students (California Academic Press, Inc., 2019; Facione, 1990; Facione, 1991; Facione et al., 1994; Khallli & Hossein Zadeh, 2003; Pitt et al., 2015). A review of the literature indicated that the CCTST has been widely used as a measurement of critical thinking across studies investigating critical thinking in healthcare related disciplines (Bowles, 2000; Ross et al., 2016; Zygmont & Schaefer, 2006). The CCTST is designed to engage critical thinking skills required to succeed in educational or

workplace settings, where solving problems and making decisions by forming reasoned judgments are important. The instrument is intended to provide objective measurement of core reasoning skills associated with critical thinking and is comprised of 9 subsections:

- Overall reasoning skills
- Analysis
- Interpretation
- Evaluation
- Explanation
- Inference
- Deduction
- Induction
- Numeracy

Simucase® is an on-line learning platform created by Case Western Reserve University to assist speech-language pathology and audiology students with access to virtual patients, case studies, and clinical simulations. The program is designed to enhance clinical competency, build knowledge and professional judgement (Johnson et al., 2018). Professional judgement, or clinical reasoning/diagnostic thinking is thought to be developed through engagement with interactive modules and resources that require the user to analyze information from multiple sources, engage in reflective practice, and make clinical decisions regarding patient care. Simucase® also provides assessments for evaluating decision making abilities within each section of a case. The user is scored based on the percentage earned within each testing situation. A score of 90% or higher is designated as *mastering*. A score of 70-89% is designed as *developing*. And a score of 70% or lower is designated as *emerging* (Johnson et al., 2018).

Validation of the Simucase® ability to accurately measure critical thinking was limited in the literature. A study conducted by Carter in 2019 provided preliminary support for the use of computer-based simulations with graduate students in a Communication Sciences and Disorders program. In the study, graduate students enrolled in a "School-Aged Language Disorders" class were divided into 2 groups (Carter, 2019, p. 46). The traditional group was required to complete paper-based case studies, while the SimuCase® group was required to complete a case simulation. Students' performance in both groups was rated using the SimuCase® Clinical Skills Inventory (SCSI) and the Critical Thinking Test for Communication Sciences and Disorders (CTCSD) (Carter, 2019). The SCSI was created by the author of the study in conjunction with SimuCase® as a measure of student decision-making regarding clinical scenarios. The CTCSD was an unpublished test of critical thinking in CSD students (Carter, 2019). Based upon these two measures, students in the simulation group exhibited a significantly higher level of improvement on the SCSI and the CTCSD than the students in the traditional group. The simulation group exhibited improved ability to formulate appropriate questions, select assessments, formulate a diagnosis, and make recommendations (Carter, 2019). In addition, the higher scores on the CTCSD suggest improved critical thinking unrelated to content learned in academic coursework (Carter, 2019). Although there are limitations related to use of an author-created tool and an unpublished test of critical thinking, these preliminary findings lend credence to the use of case simulations in assessing critical thinking in CSD students.

The use virtual case studies and simulations to improve diagnostic thinking, clinical reasoning, and critical thinking have been studied extensively in the medical and healthcare fields. Studies from nursing, medicine, athletic training, physical therapy, and mental health support that participating in simulations assist students with improving clinical reasoning, diagnostic thinking and clinical decision making (Guise et al., 2011; Johnsen et al., 2016; Macauley, 2018; Palmer et al., 2014; Weller et al., 2012). Additionally, clinical simulation activities appear to align with the definition of critical thinking recommended by Finn et al. (2016) presented by Wade et al.(2014) that states, "Critical thinking is the ability and willingness to access claims and make objective judgments on the basis of well-supported reasons and evidence rather than emotion or anecdote" (pp. 6-7). Clinical simulation activities are typically comprised of discipline specific content that requires the learner to make objective judgments using evidence rather than anecdote. For example, Simucase® clinical simulations require the user to make appropriate referrals, ask appropriate questions, make clinical hypotheses, and support diagnostic decisions with appropriate recommendations.

Completion of Critical Thinking Measures and Selection of Simucase® Clinical Simulations.
All first and fourth semester students completed the CCTST and two selected Simucase® clinical simulations at the beginning of the academic semester, approximately 1-2 weeks prior to initial clinical contact at the assigned clinical practicum site. All critical thinking measures were repeated at the end of the academic semester, approximately 1 week post final clinical contact at the respective clinical practicum sites.

Simucase® clinical simulations were selected based on students' knowledge and experience level. First semester students completed a basic pediatric speech-sound disorder clinical simulation and a more complex pediatric language disorder clinical simulation. Adult clinical simulations were not selected due to the fact that clinical contact in the first semester is typically limited to pediatric cases. The same two selected Simucase® clinical simulations were repeated at the post-intervention phase.

Fourth semester students were assigned more complex Simucase® clinical simulations due to the fact that they had more clinical knowledge and experience. Fourth semester students completed a pediatric fluency clinical simulation and adult traumatic brain injury clinical simulation. Again, the same case studies completed at baseline were repeated at the post-intervention phase.

All Simucase® clinical simulations were completed in assessment mode, which is intended to be a summative method of evaluation. Assessment mode as opposed to learning mode was selected due to the fact that real-time feedback is not provided to the user in assessment mode. In addition, all participating students did not receive any supervisory feedback or debriefings at baseline or at the post-intervention phase on their Simucase® clinical simulation performance.

Video-Recordings and Analyses of Student-Supervisor Conferences. All 17 first semester students participating in the study engaged in weekly video-recorded conferences with their supervisors for the duration of 1 academic semester. All 3 fourth semester students also engaged in video-recorded conferences; however, due to time and scheduling constraints within the externship sites, the total number of video recorded conferences were substantially fewer than the total number of on-site video recorded conferences (see Table 4). Each clinical supervisor was

given an iPad for the sole purpose of recording supervisory conferences. Clinical supervisors were instructed to record the weekly conference using the camera app on the iPad. In an attempt to reduce bias during the video-analysis phase, clinical supervisors were also instructed to position the camera so that the clinical supervisor was not in view of the picture. There were no time requirements for each video-recorded conference. Clinical supervisors downloaded the videos to Mediasite, a centralized video-platform set up by university technology services, to be securely accessed by the independent raters using a unique log on and password.

At the conclusion of the semester, each weekly video-recording was randomly assigned to 1 of 3 SQF trained raters for analysis to determine if the video debriefing session was rated to be SQF or NSQF type of supervision. The purpose of the rater analyses was to determine the reliability and validity of the supervisors in delivering SQF supervision. This pilot study followed an "intention to treat" approach. According to Hollis and Campbell (1999) an "intention to treat" approach "is generally interpreted as including all patients, regardless of whether they actually satisfied the entry criteria, the treatment actually received, and subsequent withdrawal or deviation from the protocol" (p. 670). An intention to treat approach is used to more closely parallel real-world conditions and help to avoid overestimating the benefit of an intervention. As such, those SQF students whose videos were felt by the raters to not meet the criteria for the SQF model were still included in the SQF group for the purposes of the data analysis.

Each of the 3 SQF raters was instructed to watch and review the assigned videos until there was sufficient data to complete the analysis form. Each rater completed one analysis form for every assigned video and uploaded the completed analysis form to a centralized location on OneDrive (See Appendix A). The analysis form was comprised of 19 possible checkbox statements. Raters were instructed to check off each statement that applied to the video observation. Twelve of the 19 statements were designed to qualify the video as SQF, while the other 7 statements were NSQF qualified statements. One of the twelve statements was a repetitive SQF question related to stimulating critical thinking using higher level questions. Therefore, the total number of SQF statements were 11. A minimum number of 8 of the 11 SQF statements (70%) had to be checked in order for the video to meet SQF requirements. For the purposes of this paper, this will be referred to as *criterion X*. In addition, at the end of the form, the raters answered the question, "Do you feel SQF was implemented? Why?" For the purposes of this paper, this will be referred to as *criterion Y*. In order to ensure that the raters watched the videos, raters were also instructed to list the questions the supervisor asked during the conference. In order to assess SQF rating agreement, the SQF developer who provided the SQF training workshop for this study and who served as an additional independent fourth rater was also randomly assigned videos rated by the other 3 raters using the same procedures as the other 3 raters.

Post-Survey. After the completion of the study, all 20 participating students were asked to complete a 12-question survey, rating their supervision experiences from 1 to 5, using the following rating scale: 1= strongly, 2=disagree, 3=neutral, 4=agree, and 5=strongly agree (total score range 12 to 60). A copy of the post-survey is provided in Appendix B.

Data Analysis. All statistical analyses were completed using Minitab, Version 17 (Minitab, LLC) and SAS 9.04.01. Demographic data for study participants was completed using descriptive statistics. Two sample *t* tests were calculated to determine differences between SQF and NSQF

groups on critical thinking measures and the post-survey. Pre and post comparisons on the selected critical thinking measures for fourth semester students were completed using paired t tests. Descriptive statistics were conducted in order to assess for variability of ratings by supervisor. In order to evaluate rater validity, chi-square tests of independence were conducted to determine whether or not an association existed between the intention to use SQF on a given video and the raters' assessment of whether or not SQF was being implemented. Chi-square tests of independence were completed for each of the criteria separately: (a) subjective identification of SQF (criterion Y) and (b) meeting the 70% criteria (criterion X). Intra-rater reliability was assessed using descriptive statistics and Kappa values with regards to when videos were rated as SQF by criterion X versus criterion Y. The level of agreement was determined using criteria established by Landis and Koch (1977) (poor agreement = less than 0.20, slight agreement = 0.20 to 0.40, moderate agreement = 0.40 to 0.60, substantial agreement = 0.60 to 0.80, and almost perfect agreement = 0.80 to 1.00). Descriptive statistics were conducted in order to determine the percentage of true positives and true negatives with regards to assessment of the intention to use SQF. Kappa values were also calculated to determine inter-rater reliability between rater 4 and raters 1, 2, and 3, for scoring videos. For the sake of assessing agreement, a video was rated overall as a "yes" to SQF only if it satisfied both criteria (X and Y). A video was rated overall as a "no" if it failed either criterion or both of the two criteria (X and Y). Post-hoc power analyses were completed to determine sample size information needed to reach statistical significance.

Results

SQF and NSQF Video Analysis Assignments. Overall, there were a total of 133 SQF videos and 61 NSQF videos that were assigned to each of the 3 raters. The SQF developer who provided the SQF training workshop served as an additional independent fourth rater in order to assess agreement in how the videos were analyzed by the 3 independent raters. Please see Table 1 for total number of videos analyzed by each of the 4 raters.

Table 1.

Total number of videos analyzed according to rater.

	Videos Analyzed			
	SQF On-Site	NSQF 1st On-Site	SQF Off-Site	Total
Rater 1	38	20	5	63
Rater 2	35	20	8	63
Rater 3	39	21	8	68
Rater 4	9	6	8	23
Total	121	67	29	$n = 217$

Impact of SQF on Assessments of Critical Thinking. In terms of the participating first semester students, results from two sample t tests indicated that there were no statistically significant differences between SQF and NSQF groups of students based upon pre to post-intervention scores on both critical thinking measures. Simucases® results for the NSQF group showed improvement in pre to post-intervention scores ($M = 5.1250$, $SD = 13.6912$) which was slightly higher than the SQF group ($M=3.833$, $SD = 22.0434$), but was not statistically significant ($p = .8411$). CCTST results for both groups showed a slight worsening in pre to post intervention scores with the NSQF group ($M = -2.63$, $SD = 4.84$) with a slightly smaller decline than the SQF group ($M=-2.89$, $SD = 4.81$), but again with the difference not being statistically significant ($p = .544$). Further analysis of individual subsections on the CCTST and Simucases® for all participating first semester students also showed no statistically significant differences between NSQF and SQF groups. See Tables 2 and 3 for results.

Table 2

Results of Simucase® for First Semester On-Site Students

	NSQF (*n*=16)		SQF (*n*=18)		
Y1 Simucase® results	*M*	*SD*	*M*	*SD*	*p*-value
Pre-test Scores	24.375	14.8767	33.1111	16.8344	.1207
Post-test Scores	29.5000	15.9917	36.9444	17.6284	.2085
Overall Score Difference	5.1250	13.6912	3.8333	22.0434	.8411
Case History	3.0000	26.9592	-1.0556	29.1981	.6780
Collaborators	7.5000	24.0139	14.1667	35.4903	.5312
Hypothesis	14.2857	37.7964	12.5000	35.3553	.9261
Assessment/Treatment	-3.1250	22.8849	11.5556	36.2900	.1742
Diagnosis/Progress	10.2500	44.7966	-0.7222	49.0112	.5025
Recommendations	6.2500	57.3730	0	68.5994	.7767

Note: Results based upon two sample t tests and paired t test

Table 3.

Results of CCTST for First Semester On-Site Students

Y1 CCTST results	NSQF (n=8)		SQF (n=9)		
	M	*SD*	*M*	*SD*	*p*-value
Pre-test Scores	20.00	2.39	20.11	3.18	0.468
Post-test Scores	17.38	4.17	17.22	5.52	0.525
Overall Score Difference	-2.63	4.84	-2.89	4.81	0.544
Induction	-1.00	1.60	-1.67	1.58	0.798
Deduction	-1.38	1.19	-0.56	2.65	0.210
Analysis	0	1.20	-0.78	1.92	0.835
Inference	-0.75	1.83	-0.67	1.58	0.461
Evaluation	-1.13	1.25	-1.56	1.88	0.708

Note: Results based upon two sample *t* tests and paired *t* test

Using paired *t* tests, all 3 participating fourth semester students who received SQF, also did not demonstrate statistically significant differences on pre to post scores on both critical thinking measures. Simucases® results show a slight decline in pre to post intervention scores with pretest scores ($M = 66.33$, $SD = 8.76$) higher than posttest scores ($M = 62.67$, $SD = 9.77$) but not with any degree of statistical significance ($p = .879$). Similarly, CCTST results show a slight decline in pre to post intervention scores with pretest scores ($M = 18.67$, $SD = 3.51$) higher than posttest scores ($M = 16.00$, $SD = 4.58$), but again without any degree of statistical significance ($p = .827$).

Student Post-Survey Ratings. Seventeen out of 20 students (11/12 SQF students and 6/8 NSQF students) completed a post-survey rating their supervisory experience. One of the 6 participating NSQF students did not respond to one of the survey questions (question 9); therefore, total sample size for survey completion was 16 students. However, in the breakdown by question, all of the questions other than question 9 had a sample size of 17. Data collected from responses to post-survey ratings included both first and fourth semester students. Results from two sample *t* tests indicated that there were statistically significant differences between SQF and NSQF groups of students based upon post-intervention survey ratings with SQF students rating their experience higher ($M = 54.55$, $SD = 5.11$) than the NSQF students ($M = 45$, $SD = 9.49$) ($p = .04$). Table 4

shows that there was a statistically significant difference in all but 3 of the individual questions. Questions pertaining to self-reflection (Question 1), use of higher level supervisory questioning in stimulating critical thinking (Question 7), and overall effectiveness of the supervisory method used (Question 12) each showed the strongest statistically different ratings ($p = .017$). with the SQF students rating their experience higher than the NSQF group. Questions pertaining to strengthening clinical competence (Questions 5 and 8) and use of supervisory questions that required problem solving client performance (Question 4) were not statistically different between SQF and NSQF groups.

Ratings by Supervisor. Descriptive statistics summarized in Table 5 show the breakdown by rater and supervisor regarding the percentage of videos that met both criteria (X and Y).

Rater Validity. A chi-square test of independence was performed to examine the association between when raters 1, 2, and 3 identified a video as SQF by criterion Y and when it was intended to be SQF. The association between these variables was significant X^2 (1, $N = 194$) = 102.25, $p < .0001$. Chi square test of independence was performed to examine the association between when raters 1, 2, and 3 identified a video as SQF by criterion X and when it was intended to be SQF. The association between these variables was significant X^2 (1, $N = 194$) = 113.02, $p < .0001$.

The same chi-square tests of independence were performed specific to rater 4. Results revealed that the association between identifying a video as SQF by criterion Y and when it was intended to be SQF was significant X^2 (1, $N = 23$) = 6.66, $p = .0099$. The association between rater 4 identifying a video as SQF by criterion X and when it was intended to be SQF was also significant X^2 (1, $N = 23$) = 13.55, $p = .0001$.

Intra-Rater Reliability. Table 6 shows that there was agreement between coding videos as (1) "yes" for SQF by both criteria (X and Y) or (2) "no" for SQF by both criteria (X and Y) across each of the 4 raters.

A Kappa value was calculated to determine the agreement between when raters 1, 2, and 3 identified a video as SQF by criterion X and when by criterion Y. The value for Kappa was .7533, indicating a substantial level of agreement. The value was significantly different from zero ($p < .0001$). In terms of rater 4, the value for Kappa was .7416, also indicating a substantial level of agreement. Again, the value was significantly different from zero ($p = .0002$).

Table 4

Supervisory post-survey results.

Survey	NSQF $n=6$		SQF $n=11$		
	M	*SD*	*M*	*SD*	*p*-value
1	3.667	0.816	4.636	0.505	.017*
2	4.000	0.632	4.727	0.467	.019*
3	3.833	0.753	4.545	0.522	.039*
4	3.833	0.983	4.636	0.505	.055
5	3.833	0.983	4.182	0.751	.236
6	3.500	0.837	4.455	0.688	.022*
7	3.667	0.816	4.636	0.505	.017*
8	3.667	0.816	4.273	0.905	.093
9**	3.000	0.894	4.364	0.505	.037*
10	3.833	0.753	4.636	0.505	.026*
11	3.833	0.983	4.818	0.405	.033*
12	3.667	0.816	4.636	0.505	.017*
Total *M***	45	9.49	54.55	5.11	.044*
Total Median**	42		56		.060

Note: *indicates significant a *p*-value. *p*-value reflects results from two sample *t* tests with a significance set at .05.
**The NSQF *n = 5* for question 9. As such, the NSQF Total *M* and Median were only able to be conducted on 5 of the NSQF surveys, so *n = 5* for these values as well

Table 5.

Percentage of videos identified by raters as SQF by both criteria (X and Y)

	NSQF On-Site Supervisors (*n* = 2)			SQF On-Site Supervisors (*n* = 3)				SQF Off-Site Supervisors (*n* = 4)			
	A	B	Total	C	D	E	Total	F	G	H/I	Total
Rater 1	0% (0/2)	6% (1/18)	5% (1/20)	69% (9/13)	77% (10/13)	100% (12/12)	82% (31/38)	33% (1/3)	100% (2/2)	n/a (0/0)	60% (3/5)
Rater 2	0% (0/2)	0% (0/18)	0% (0/20)	75% (9/12)	100% (11/11)	100% (12/12)	91% (32/35)	75% (3/4)	100% (3/3)	0% (0/1)	75% (6/8)
Rater 3	0% (0/2)	0% (0/19)	0% (0/19)	38% (5/13)	86% (12/14)	92% (11/12)	72% (28/39)	33% (1/3)	100% (3/3)	50% (1/2)	63% (5/8)
Total 1,2,3	0% (0/6)	2% (1/55)	2% (1/61)	61% (23/38)	87% (33/38)	97% (35/36)	81% (91/112)	50% (5/10)	100% (8/8)	33% (1/3)	67% (14/21)

Note: Numbers in parentheses reflect the number of videos meeting both criteria (X and Y) divided by the total number of videos.

Composite of Rater Validity and Intra-Rater Reliability. Results from descriptive statistics in Table 7 show how often raters gave ratings in agreement with the SQF vs NSQF assignment for each video (by both criteria X and Y).

Table 6.

Intra-rater reliability as assessed by labeling a video as SQF or NSQF consistently by both criteria

	Video rated: Criterion X: No Criterion Y: No	Video rated: Criterion X: Yes Criterion Y: Yes	Total Agreed
Rater 1	30% (19/63)	56% (35/63)	86% (54/63)
Rater 2	37% (23/63)	60% (38/63)	97% (61/63)
Rater 3	34% (23/68)	49% (33/68)	82% (56/68)
Total 1,2,3	34% (65/194)	55% (106/194)	88% (171/194)

Table 7.

Composite of rater validity and intra-rater reliability

	NSQF Assignment Criterion X: No Criterion Y: No	SQF Assignment Criterion X: Yes Criterion Y: Yes	Total Matched Videos that Agree with SQF/NSQF Assignment
Rater 1	70% (14/20)	79% (34/43)	76% (48/63)
Rater 2	100% (20/20)	88% (38/43)	92% (58/63)
Rater 3	86% (18/21)	70% (33/47)	75% (51/68)
Total 1,2,3	85% (52/61)	79% (105/133)	81% (157/194)

Inter-Rater Reliability. A Kappa value was calculated to determine the agreement between rater 4 and raters 1, 2, and 3 for scoring videos. The value for Kappa was .4773, indicating a moderate level of agreement. The value was significantly different from zero (p = .0221).

Post-Hoc Power Analysis. A post hoc statistical power analysis was performed to determine what sample size would be needed to reach statistical significance. Simucases® results for an $N= 34$ with NSQF and SQF group $M = 5.1250$ and $M = 3.833$, respectively, and a $SD = 18.33$ revealed a power of .055. CCTST results for an $N= 17$ with NSQF and SQF group $M = -2.63$ and $M = -2.83$, respectively, and a $SD = 4.671$ revealed a power of .051 , With an alpha = .05 and power = 0.80, the projected sample size needed for Simucases® was $N = 6322$ and for CCTST was $N= 17128$.

Additionally, a standard power analysis was performed to determine what sample size, given a similar standard deviation in scores, would detect a difference between a group mean of 10 and 5 (one group scores 5 points higher than the other). Standard deviations similar to those revealed by the study results were used. For Simucases®, a standard deviation of 20 was used. For CCTST, a standard deviation of 5 was used for the power calculation. For Simucases®, the projected sample size was $N = 506$ and for CCTST was $N= 34$.

Discussion

Impact of SQF on Assessments of Critical Thinking. The ability to use critical thinking within the clinical setting is seen as necessary and essential to providing evidence-based diagnostic and treatment services (Barrett et al., 2018; Procaccini et al., 2016). As such, clinical educators are challenged with the responsibility of establishing a stimulating teaching-learning environment such that critical thinking is developed. Current available evidence within the scope of clinical education has underscored the complexities surrounding effective clinical teaching methodologies, specifically those that intend to develop critical thinking skills (Kavanagh & Szweda, 2017; Procaccini et al., 2016). Recent changes to training requirements for speech-language pathologists and audiologists engaging in clinical teaching have highlighted the need for systematic and evidence-based methods of clinical teaching. Although many different models of clinical teaching have been presented within the literature in disciplines outside of speech-language pathology and audiology, few have been systematically investigated within the clinical teaching-learning environment of speech-language pathology. Clinical educators are challenged with selecting a clinical teaching method but also determining its effectiveness.

The aim of this study was to determine whether or not the SQF model of supervision stimulated critical thinking in graduate level speech-language pathology students within the clinical teaching-learning environment. The researchers hypothesized that students who were provided with the SQF model of supervision would score higher than students who received the NSQF style of supervision on the selected critical thinking measures. Overall findings of the present study did not support improvements in critical thinking as measured by scores on the CTTST and Simucase® clinical simulations when comparing SQF and NSQF groups. Additionally, results also did not support improvements in critical thinking for fourth semester students as measured on pre to post scores on Simucase® clinical simulations and the CCTST. The failure to reach statistical significance could be interpreted that SQF does not have a statistically significant effect on critical thinking. However, as noted below in the limitations section, multiple confounding variables have likely limited the negative predictive value of this study. Unfortunately, the results of this study cannot be compared to previous existing studies investigating the use of SQF on critical thinking. To date, this is the first pilot study either within or outside the scope of communication disorders that has specifically investigated SQF and its effect on critical thinking. However, Nottingham and

Henning (2014) provides some preliminary support that SQF has the potential to assist students in several different areas of clinical practice (e.g. clinical reasoning, clinical skills etc.) based on feedback methods. The researchers of this study concluded that the SQF model provides guidelines for clinical instructors in athletic training on how to provide effective feedback to students based on adjusting feedback methodologies according to the level of the learner. For example, providing more corrective feedback to novice learners and more directive feedback to advanced learners. Theoretically, conscious adjustment of feedback methods according to the level of learner may, in part, support the development of critical thinking.

Student Preferences and Impact of SQF on Student Engagement. External evidence has supported that students prefer a clinical teaching-learning environment that is positive, encouraging, and thought-provoking. Thrasher, Walker, and Weidner (2018) found that newly credentialed athletic trainers preferred preceptors who were encouraging them to make clinical decisions and perform clinical skills. Burningham, Deru, and Berry (2010) emphasized the importance of responsiveness to students actively engaging students in the clinical teaching-learning environment. The results of this study showed that students preferred the SQF model over the NSQF model based on post-intervention survey ratings ($p= .04$), providing some support that the supervisory method used with the SQF group was not only different than that used with the NSQF group but also gauged to be superior. First semester graduate students who participated in the SQF supervision group related increased clinical confidence and the ability to self-reflect as compared to students in the NSQF group (see Table 4). Interestingly, the SQF group also rated the overall effectiveness of the supervisory method (Question 12) higher than the NSQF group ($p= .017$), again providing some support for not only a difference in supervisory method between the groups but also a preference for the SQF method. These results may provide evidence that some of the protocols within the design of the study, such as more conscious effort on the part of the clinical educator to ask stimulating questions and a routine, designated time to do so, may stimulate student engagement and self-efficacy. Similar to other studies that correlate perceptions of effective clinical teaching characteristics with stimulating student involvement (Smith et al., 2011), the SQF students may have perceived a higher level of engagement in the learning process and thus felt the SQF model to be more effective overall. This gives credence to the use of strategic questioning methodologies, such as the SQF model, in the scope of clinical education. Last, evidence suggests that the ability to self-reflect is integral to critical thinking processes (Ghanizadeh, 2017; Kuiper, 2002). The SQF group felt their supervisory method facilitated self-reflection to a greater degree than the NSQF group. Use of strategic questioning models may stimulate self-reflection (or at least perceived self-reflection) and by extension critical thinking. An additional clinical implication for future investigation may be to reinforce self-reflection by encouraging students to write down and reflect on strategic questions asked by their clinical educator.

Ratings by Supervisor. Results showed some variability in video ratings of SQF and NSQF styles across raters and supervisors which may provide some information on the consistency at which SQF was being implemented. While an "incorrect" rating of a video as being NSQF when it was intended to be SQF may be considered an inaccuracy on the part of the rater, it may also reflect how consistently the supervisor was actually implementing SQF. Another review of Table 5 will reveal that while rater 3 consistently rated all of the on-site SQF Supervisors lower than raters 1 and 2, it will also show that Supervisor C consistently was rated by all three raters lower than

Supervisors D and E. Consistent with these results, the lowest rating was of Supervisor C by rater 3.

In terms of off-site supervisors F, G, H/I, there was wide variability in video ratings across raters ranging from 33% to 100% with regards to whether or not SQF was being implemented. The overall number of video debriefings was substantially less for all off-site supervisors, which may have impacted the amount of SQF that the fourth semester students were receiving. In addition, supervisors H/I were rated as meeting criteria X and Y in only 33% of the videos; however, supervisors H/I submitted the lowest total number of videos at 3. This was still included in the data analysis based off of the "intention to treat" approach; however, this may have potentially skewed the results.

Number of years of supervisory experience did not appear to influence the accuracy at which SQF was being implemented, as supervisor E had the lowest number of years supervisory experience of the on-site supervisors (7 years) yet was rated highest by SQF raters. Similarly, supervisor G of the off-site supervisors (5 years), was rated highest by raters and had significantly less clinical supervisory experience than supervisor F (>20 years). Interestingly, out of the 3 on-site SQF trained supervisors, supervisor E had the most previous knowledge and experience of SQF prior to the study period. Supervisor E engaged in scholarship activities pertaining to SQF and completed a prior SQF training in advance of the study. Supervisors C and D did not complete any prior trainings pertaining to the SQF model, although they did engage in some scholarship activities pertaining to the model. This may provide some additional evidence that training beyond a 4-hour workshop, perhaps more hands-on, is needed in order to attain adequate proficiency in using the SQF model. Supervisor E may have also had more engagement and commitment to the SQF model given the previous experience and exposure to the model. These findings may corroborate existing available evidence that correlate effective supervision outcomes with emphasis on training supervisors in the quality of their supervision practices (e.g. feedback, appraisal, assessment methods etc.) (Kilminster et al., 2007).

SQF Rater Validity and Reliability. Both SQF criteria, (a) the subjective SQF assessment (criterion Y) and (b) the 70% criteria (criterion X), were independently found to have a statistically significant correlation to when SQF was intended to be implemented. This provides some evidence that the criteria designed on the SQF analysis form used by raters was able to identify when SQF was being implemented. Similarly, this also shows that raters that had been trained in SQF were able to accurately identify when SQF was being used.

Results from descriptive statistics in Table 6 show that the raters were largely consistent with regards to their global subjective assessment of whether SQF was being implemented matching with the assessment of whether SQF was being implemented based on the 70% criteria. This was consistent across all three raters. Furthermore, Kappa values indicated substantial agreement for all four raters between criterion X and criterion Y. This indicates that a quick yes/no question about whether SQF was being used consistently agreed with the more detailed 19 item questionnaire. This may have implications for future studies as a simpler rating system may facilitate studies with a larger sample size

Results from descriptive statistics in Table 7 combined the agreement factor discussed in Table 6 with whether those videos that matched both criteria (X and Y) also agreed with the SQF assignment that was intended. This again shows that not only was each rater internally consistent in their assessment of video's SQF vs NSQF status, but that this assessment was typically in agreement with the SQF assignment. With the exception of rater 1, it does appear that the raters identified the NSQF videos "correctly" slightly more consistently than the SQF videos. Essentially, raters 2 and 3 had a higher specificity (identifying true negatives) than sensitivity (identifying true positives).

Lastly, for the purposes of assessing the inter-rater reliability of raters 1, 2 and 3, rater 4 reviewed some of the videos that each of the other raters had previously rated. The Kappa value of .4773 indicates only a moderate level of agreement, although still statistically significant. Rater 4 is considered an expert in SQF with far more experience with SQF than raters 1, 2, and 3. It is certainly possible that as an expert and SQF developer, rater 4 may have had more stringent criteria when rating videos. The lack of a higher level of agreement does raise some concerns that raters in future studies may need more extensive training in SQF prior to assessing videos. The sample size for rater 4 was rather small (n=23), and it is certainly possible that a larger sample size may have given a better assessment of the level of agreement between rater 4 and the other raters.

Limitations

Several limitations to the overall study design may have contributed to the results of the study. Timing may have adversely impacted scores on critical thinking measures. Post-intervention completion of the CCTST and Simucase® clinical simulations occurred on the final day of the semester, just prior to winter break. Students may not have been fully engaged and did not take the time to thoughtfully complete post-treatment assessments due to eagerness to complete the semester. Additionally, proficient critical thinking has been cited in the literature to require time and practice. Given that the study was completed in the duration of only 1 semester, students may not have been given enough time to fully develop their critical thinking skills. Further studies investigating the effects of SQF when implemented for longer time periods (possibly at the beginning and end of graduate program) are needed. Due to difficulties finding dedicated time to record SQF debriefings, externship supervisors completed substantially less video debriefings than on-site supervisors. This may provide some insight into the need for further research on the length and frequency of SQF debriefings and also exposed some of challenges regarding allotting dedicated time to clinical teaching.

Results also showed some variability in video ratings of SQF and NSQF styles across raters and supervisors which provides some insight into the need for better standardization in training and inter-rater reliability measures to ensure SQF is actually being implemented. Despite participating in a 4-hour hands-on SQF workshop, supervisors had limited opportunity to engage in direct practice with students using the SQF model prior to participating in the study. More hands-on training over the course of several days with the achievement of a minimal level of proficiency in the SQF model may be necessary to ensure that supervisors are engaging in the appropriate use of the model. In addition, there was a delay of approximately 4 weeks between the SQF training and the opportunity to engage in the SQF model during the study that may have impacted the supervisors' ability to effectively use the SQF model when interacting with the students. Lastly,

the raters were charged with identifying whether or not SQF was being implemented; however, they were not charged with assessing the quality of the SQF (e.g. amount and variety of higher level questioning) being implemented which may provide further insight into the study's results.

Although the CCTST is designed to engage the test-taker's critical thinking skills and is intended to be discipline-neutral, it may not be sensitive to the critical thinking abilities necessary in the clinical setting with CSD students. Use of a test of critical thinking specifically designed for CSD students should be considered in future investigations. Similarly, although Simucase® clinical simulations address more discipline specific critical thinking skills, use of students' scores on the Simucase® clinical simulations may not correlate well with more general critical thinking abilities. Additionally, the selected Simucase® clinical simulations were limited to the specific disorder areas of speech-language pathology and for some students, may not have been specific to the clinical assignments at the respective clinical practicum sites. It may be possible that students did not show over-arching changes in critical thinking and may have acquired knowledge and skills in the specific clinical area practiced for that semester. Further research into selecting and measuring critical thinking within the clinical teaching-learning environment is necessary.

An additional limitation may have been the randomization procedures selected. First semester study participants were randomized to SQF and NSQF groups prior to pre-test completion of the critical thinking measures. Given that the results revealed that the SQF group was close to reaching statistical significance on the pre-test Simucase® clinical simulations when comparing to the NSQF group, randomization after the pre-test period may have assisted with improving overall randomization across both groups.

Last, unfortunately due to the size of the graduate cohort, the sample size of this study was small. Post-hoc power analyses provided confirmation of the limitations in sample size and study power. Recruitment of a larger number of students and supervisors across several university graduate programs would certainly yield a higher sample size, stronger study power, and greater external validity. The clinical settings were also limited in this study, which reduces the ability to generalize the implementation of SQF across a variety of clinical settings.

Conclusions

The SQF model of supervision was developed within the scope of athletic training in order to provide a structural framework for clinical teaching. The theoretical underpinnings of the SQF model align with a developmental view of acquiring critical thinking skills such that the level of supervision, questioning, and feedback provided is gradually scaffolded to the needs and skill level of the learner. The implementation of the SQF model in disciplines outside the scope of athletic training is gaining appeal among clinical educators due to its potential for stimulating critical thinking. However, the generalizability of the SQF model to clinical teaching-learning contexts within speech-language pathology and audiology has not been formally investigated.

To date, this is the first small pilot study investigating the effects of the SQF model on stimulating critical thinking in the clinical teaching-learning environment. Overall results from this preliminary study indicated that the SQF model of supervision did not appear to influence the outcomes on the CCTST or SimuCase® clinical simulations. However, critical appraisal of the

study's limitations has provided great insight into how clinical educators may improve the design of future research investigations involving the use of SQF and its effect on developing critical thinking in learning clinicians. Given the limited research available in the use of systematic clinical teaching methodologies such as SQF, it is important to provide information on how future research investigations may be designed to optimize external validity. Additional considerations for future directions of research may include investigating questioning specificity and frequency and its effects on discipline specific and discipline neutral critical thinking skills. The study also exposed some of the potential challenges of ensuring appropriate SQF training and implementation, effectively assessing critical thinking in the clinical setting, and time commitments needed for clinical teaching. In addition, students who received the SQF model preferred this model over those that did not receive SQF, which may provide evidence that students want to be actively engaged in the supervisory process through strategic questioning methods. Further investigation of the effects of SQF model on students' critical thinking is warranted using larger sample sizes across a variety of clinical teaching-learning settings.

Acknowledgements

Disclosure Statement

References

American Speech-Language-Hearing Association. (n.d.). *Practice portal: Clinical education and supervision*. ASHA Practice Portal. https://www.asha.org/Practice-Portal/Professional-Issues/Clinical-Education-and-Supervision/

American Speech-Language-Hearing Association. (1985). Clinical supervision in speech language pathology and audiology. *ASHA*, 27, 57–60.

American Speech-Language-Hearing Association. (2008a). *Clinical supervision in speech-language pathology* [Position statement]. ASHA. http://www.asha.org/policy/ PS2008-00295/

American Speech-Language-Hearing Association. (2008b). *Clinical supervision in speech-language pathology* [Technical report]. ASHA. https://www.asha.org/policy/TR2008-00296/

American Speech-Language-Hearing Association. (2008c). *Knowledge and skills needed by speech-language pathologists providing clinical supervision* [Knowledge and skills]. ASHA. https://www.asha.org/policy/KS2008-00294/

American Speech-Language-Hearing Association. (2013). *Knowledge, skills and training considerations for individuals serving as supervisors.* [Final report, Ad Hoc Committee on

Supervision]. ASHA. https://www.asha.org/uploadedFiles/Supervisors-Knowledge-Skills-Report.pdf

American Speech-Language-Hearing Association. (2015). *The role of undergraduate education in communication sciences and disorders.* ASHA. http://www.asha.org/uploadedFiles/ASHA/About/governance/Resolutions_and_Motions/2015/BOD-21-2015-AAB-Report-on-the-Role-of-Undergraduate-Education.pdf

American Speech-Language-Hearing Association. (2016). *A plan for developing resources and training opportunities in clinical supervision* [Final report of the ASHA Ad Hoc Committee on Supervision Training]. ASHA. https://www.asha.org/uploadedFiles/A-Plan-for-Developing-Resources-and-Training-Opportunities-in-Clinical-Supervision.pdf

Anderson, J. (1988). *The supervisory process in speech-language pathology and audiology.* College-Hill.

Anderson, L. W. (Ed.), Krathwohl, D. R. (Ed.), 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.* Longman.

Barnum, M. G. (2008). Questioning skills demonstrated by approved clinical instructors during clinical field experiences. *Journal of Athletic Training, 43*(3), 284-292.

Barnum, M., Guyer S., Levy L., & Graham C. (2009). Supervision, questioning, feedback model of clinical teaching: A practical approach. In T. G. Wiedner (Ed.), *The athletic trainers' pocket guide to clinical teaching* (pp. 85-99). SLACK Inc.

Barnum, M., & Guyer, S. (2016, April 21). *SQF model for clinical instruction* [Invited speakers]. The 2016 Annual Meeting of the Michigan State Hearing Association, Grand Rapids, MI.

Barrett, J. L., Denegar, C. R., & Mazerolle, S. M. (2018). Challenges facing new educators: Expanding teaching strategies for clinical reasoning and evidence-based medicine. *Athletic Training Education Journal, 13*(4), 359-366.

Bloom, B.S. (Ed.), Engelhart, M.D., Furst, E.J., Hill, W. H., & Krathwohl, D.R. (1956). *Taxonomy of educational objectives: The classification of educational goals, Handbook 1: The cognitive domain.* David McKay Company.

Borton, T. (1970) *Reach, touch and teach.* McGraw-Hill Paperbacks

Bowles, K., 2000. The relationship of critical-thinking skills and the clinical-judgment skills of baccalaureate nursing students. *Journal of Nursing Education, 39*(8), 373-376.

Burningham, D. S., Deru, L., & Berry, D. C. (2010). What traits make for an effective athletic training educator and mentor? *Athletic Training Education Journal, 5*(4), 183-186.

California Academic Press, Inc. (2019). *California Critical Thinking Skills Test (*CCTST). https://insightassessment.com/Products/Products-Summary/Critical-Thinking-Skills-Tests/California-Critical-Thinking-Skills-Test-CCTST/node_489

Carter, M. D. (2019). The effects of computer-based simulations on speech-language pathology student performance. *Journal of Communication Disorders, 77*, 44-55.

Clark, R., & Harrelson, G. (2002). Designing instruction that supports cognitive learning processes. *Journal of Athletic Training, 37*(4 Supplement), S-152-S-159.

Cogan, M. (1972). *Clinical supervision.* Houghton Mifflin.

Council for Clinical Certification in Audiology and Speech-Language Pathology of the American Speech-Language-Hearing Association. (2018). *2020 Standards for the Certificate of Clinical Competence in Speech-Language Pathology.* ASHA. https://www.asha.org/certification/2020-SLP-Certification-Standards

Council of Academic Programs in Communication Sciences and Disorders. (2013). *White paper: Preparation of speech-language pathology clinical educators.* https://wordpressstorageaccount.blob.core.windows.net/wp-media/wp-content/uploads/sites/1023/2019/06/White-Paper-Prep-of-SLP-Clinical-Education-Final-April-11-2013.pdf

Davies, M. (2015). A model of critical thinking in higher education. In M. B. Paulsen (Ed.), *Higher education: Handbook of theory and research* (Vol. 30, pp.41-92).Springer International.

Davies, E. (1995). Reflective practice: A focus for caring. *Journal of Nursing Education, 34*(4), 167-174.

Dewey, J. (1938). *Experience and education.* Touchstone Books.

Dudding, C., McCready, V., Nunez, L., & Procaccini, S. (2017). Clinical supervision in speech-language pathology and audiology in the United States: Development of a professional specialty. *The Clinical Supervisor*, *36*(2), 161-181.

Facione, P. (1990). *The California Critical Thinking Skills Test: College level: Interpreting the CCTST, group norms and sub-scores (Report No. 4)*. California Academic Press

Facione, P. A. (1991). *Using the California Critical Thinking Skills Test in research, evaluation and assessment.* https://eric.ed.gov/?id=ED337498.

Facione, N. C., Facione, P. A., & Sanchez, C. A. (1994). Critical thinking disposition as measure of competent clinical judgment: The development of the California Critical Thinking Disposition Inventory. *Journal of Nursing Education, 33*, 345-350.

Finn, P., Brundage, S. B., & DiLollo, A. (2016). Preparing our future helping professionals to become critical thinkers: A tutorial. *Perspectives of the ASHA Special Interest Groups, 1*(10), 43–68.

Ghanizadeh, A. (2017). The interplay between reflective thinking, critical thinking, self-monitoring, and academic achievement in higher education. *Higher Education, 74*, 101–114.

Goldhammer, R., Anderson, R. H., & Krajewski, R. (1980). *Clinical supervision: Special methods for the supervision of teachers (2nd Edition).* Holt, Rinehart and Winston Inc.

Guise, V., Chambers, M., & Valimaki, M. (2011). What can virtual patient simulation offer mental health nursing education? *Journal of Psychiatric and Mental Health Nursing, 19*(5), 410-418.

Hollis, S., & Campbell, F. (1999). What is meant by intention to treat analysis? Survey of published randomised controlled trials. *BMJ, 319* (7211), 670-674.

Hausmann, J. S., & Schwartzstein, R. M. (2019). Using questions to enhance rheumatology education. *Arthritis Care & Research*, *71*(10), 1304-1309.

Johnson, C., Jansen, L. J., Ondo, K., Williams, S. L., & Pantalone, B. (2018). *Simucase User's Guide: 3.0.* Ohio Department of Education-OMNIE Program. Case Western Reserve University. http://c2630440745ec56453fe-4b47e81df9184afd10797caf49eafabb.r47.cf2.rackcdn.com/7a22112084ff11e891945f0173284d22_SC_1117_UserGuide_19553-2018.pdf

Johnsen, H. M., Fossum, M., Vivekananda-Schmidt, P., Fruhling, A., & Slettebø, A. (2016). Teaching clinical reasoning and decision-making skills to nursing students: Design, development, and usability of a serious game. *International Journal of Medical Informatics, 94*, 39-48.

Joplin, L. (1995). On defining experiential education. In Warren, K., Sakofs, M. & Hunt, J.S. (Eds.), *The theory of experiential education* (3rd ed., pp. 17–22). Kendall/Hunt.

Kavanagh, J. M., & Szweda, C. (2017). A crisis in competency: The strategic and ethical imperative to assessing new graduate nurses' clinical reasoning. *Nursing Education Perspectives, 38*(2), 57-62.

Khallli, H., & Hossein Zadeh, M. (2003). Investigation of reliability, validity, and normality Persian version of the California Critical Thinking Skills Test; form B (CCTST). *Journal of Medical Education, 3*, 29-32.

Kicklighter, T. H., Geisler, P. R., Barnum, M. G., Heinerichs, S., & Martin, M. (2018) Exploration of factors perceived to influence development of diagnostic reasoning in athletic trainers and athletic training students. *Athletic Training Education Journal, 13*(2), 120-130.

Kilminster, S., Cottrell, D., Grant, J., & Jolly, B. (2007). AMEE Guide No. 27: Effective education and clinical supervision. *Medical Teacher, 29*, 2-19.

Kuiper, R. (2002). Enhancing metacognition through the reflective use of self-regulated learning strategies. *The Journal of Continuing Education in Nursing, 33*(2), 78–87.

Landis, R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159-174.

Levy, L., Gardner, G., Barnum, M**.**, Willeford, S., Sexton, P., Guyer, S., & Fincher, L. (2009). Situational supervision for athletic training clinical education. *Athletic Training Education Journal, 4*(1), 19-22.

Macauley, K. (2018). Evaluating changes in clinical decision making in physical therapy students after participating in simulation. *Health Professions Education, 4*(4), 278-276.

Mensch, J. M., & Ennis, C. D. (2002). Pedagogic strategies perceived to enhance student learning in athletic training education. *Journal of Athletic Training, 37*(4 Supplement),S-199-S-207.

Mitchell, M. M., & Poutiatine, M. I. (2001). Finding an experiential approach in graduate leadership curriculum. *The Journal of Experiential Education, 24*(3), 179-185.

Mulnix, J. W. (2012). Thinking critically about critical thinking. *Educational Philosophy and Theory, 44,* 464-479.

Nottingham, S., & Henning, J. M. (2014). Feedback in clinical education, part I: Characteristics of feedback provided by approved clinical instructors. *Journal of Athletic Training, 49*(1), 49-57.

O'Conner, A. B. (2001). *Clinical instruction and evaluation: A teaching resource*. Jones and Bartlett Publishers.

Palmer, E., Edwards, T., & Racchini, J. (2014). High-fidelity simulation meets athletic training education: An innovative collaborative teaching project. *Athletic Education Journal, 9*(2), 96-100.

Papathanassiou, I. V., Kleisiaris, C. F., Fradelos, E. C., Kakou, K., & Kourkouta, L. (2014). Critical thinking: The development of an essential skill for nursing students. *Acta informatica medica: AIM : Journal of the Society for Medical Informatics of Bosnia & Herzegovina : casopis Drustva za medicinsku informatiku BiH, 22*(4), 283–286.

Pitt, V., Powis, D., Levitt-Jones, T., & Hunter, S. (2015). The influence of critical thinking skills on performance and progression in a pre-registration nursing program. *Nurse Education Today, 35*, 125-131.

Procaccini, S. J, Carlino, N. J., & Joseph, D. M. (2016). Clinical teaching methods for stimulating students' critical thinking. *Perspectives of the ASHA Special Interest Groups, 1*(11), 3-7.

Procaccini, S. J., McNamara, K., & Lenzen, N. M. (2017). Leading the way with supervision training: Embracing change & transforming practice. *Perspectives of the ASHA Special Interest Groups, 2*(11), 42-46.

Ross, D., Schipper, S., Westbury, C., Banh, H. L., Loeffler, K., Allan, M., & Ross, S. (2016). Examining critical thinking in family medicine residents. *Family Medicine, 48*(2), 121-126.

Sharp, M. Y., Reynolds, R. B., & Brooks, K. N. (2013). Critical thinking skills of allied health science students: A structured inquiry. *Educational Perspectives in Health Informatics and Information Management, 4*, 1-13.

Smith, C., Swain, A., & Penprase, B. (2011). Congruence of perceived effective clinical teaching characteristics between students and preceptors of nurse anesthesia programs. *American Association of Nurse Anesthetists Journal, 79*(4), S62-S68.

Thrasher, A. B., Walker, S. E., & Weidner, T. G. (2018). Newly credentialed athletic trainers' perceptions of preceptor characteristics that facilitate learning and transition to practice [Abstract, p. 81] in Athletic Training Educator's Conference, February 17-19, 2017, Grapevine, TX. *Athletic Training Education Journal, 13*(1), 74-87.

Toledo, C. A. (2015). Dog bite reflections: Socratic questioning revisited. International Journal of Teaching and Learning in Higher Education, 27(2), 275–279.

Wade, C., Tavris, C., & Garry, M. (2014). *Psychology* (11th ed.). Prentice Hall.

Weidner, T., Trethewey, J., & August, J. (1997). Learning clinical skills in athletic therapy. *Athletic Therapy Today, 2*(3), 43-49.

Weller, J. M., Nestral, D., Brooks, P. M., & Conn, J. J. (2012). Medical education: Simulation in clinical teaching and learning. *Medical Journal of Australia, 196*(9), 1-5.

Willeford, S., Fincher, L., Barnum, M**.,** Gardner, G., Guyer, S., Levy, L., & Sexton P. (2009). Improving clinical education through proper supervision. *Athletic Training Education Journal, 4*(1), 6-7.

Zygmont, D. M., & Schaefer, K. M. (2006). Assessing the critical thinking skills of faculty: What do the findings mean for nursing education? *Nursing Education Perspectives, 27*(5), 260-268.

Appendix A

SQF analysis form

Instructions:

1. Log on to Mediasite with your Username and Password.
2. Complete the information regarding the SQF Rater #, Title of Video, and Date of Analysis.
3. Watch and review the video until you have achieved sufficient data to complete the SQF Analysis Form.
4. Complete the SQF Analysis Form.
   ***Please be certain to complete the SQF Rater Form during and immediately following the viewing of each video (Watch 1 video and then complete 1 form, then move to the next video).

   _____

SQF Rater #:
Title of Video :
Date of Analysis:

Please check the boxes that apply to your observation:

- ☐ **Supervisor asked most questions at the student's developmental level**
- ☐ **Supervisor intentionally structured the level of questioning to the knowledge, skill, and motivational level of the student**
- ☐ Supervisor focuses on providing mostly general comments to student about the session as a whole, such as, "pretty good", "I think things went well", " that was appropriate"
- ☐ **Supervisor provided feedback immediately or delayed, based on the situation and content**
- ☐ **Supervisor used a mixture of feedback that focuses on clinical skills/knowledge and clinical decision making as well as professional development.**
- ☐ Supervisor provided adequate pause time to permit the student to respond
- ☐ **Supervisor provided a balance of general and specific feedback**
- ☐ **Supervisor posed a series of questions that targeted differing types of knowledge so that the student is required to utilize progressively higher order thinking processes**
- ☐ Supervisor used the majority of the interaction to verbally teach concepts
- ☐ _Supervisor asked some questions above the student's developmental level to stimulate critical thinking_
- ☐ **Supervisor clearly set the "situation" by addressing the "task" discussed with the student**
- ☐ Supervisor's verbal feedback is consistently optimistic across all clinical interactions and situations

- ☐ Supervisor's verbal feedback is negative
- ☐ **Supervisor appropriately matched level of question with student's development level**
- ☐ *Supervisor asked higher level critical thinking questions such as "how?", "why?", and "what if…" appropriate to the situation and to the skill level of the student*
- ☐ **Supervisor facilitated student reflection on decisions made and actions taken and consequences or implications.**
- ☐ Supervisor implemented a mainly directive teaching style across most clinical interactions and situations
- ☐ Supervisory meeting time was adequate for the student's needs
- ☐ **Supervisor's responses and feedback are fluid to the situation**

> **Bolded: SQF**   *Italicized: Repetitive SQF*
> **Need 70% (8/11 SQF to fulfill SQF criteria)*

Do you feel SQF was implemented? Yes   No   Why?

_____
_____
_____
_____

Please write down the questions that the supervisor asked.

1._____
2._____
3._____
4._____
5._____
6._____
7._____
8._____
9._____
10._____

Appendix B

Supervisory post-survey

<div align="center">SUPERVISORY SURVEY</div>

Who was your supervisor for the Fall 2015 Semester?

_____

1. The supervisory model used by my supervisor facilitated self-reflection on decisions made, actions taken, and implications related to my clinical case.
      Strongly Agree   Agree   Neutral   Disagree   Strongly Disagree

2. The supervisory model used by my supervisor challenged me to think critically about my clinical case.
      Strongly Agree   Agree   Neutral   Disagree   Strongly Disagree

3. The supervisory model used by my supervisor challenged me to synthesize information from my coursework when developing my client's goals and objectives in therapy.
      Strongly Agree   Agree   Neutral   Disagree   Strongly Disagree

4. The supervisory model used by my supervisor included questions that required the use of problem solving skills to evaluate my client's performance in therapy.
      Strongly Agree   Agree   Neutral   Disagree   Strongly Disagree

5. The supervisory model used by my supervisor made me feel increasingly competent with my clinical skills as the semester progressed.
      Strongly Agree   Agree   Neutral   Disagree   Strongly Disagree

6. The supervisory model used by my supervisor required me to utilize progressively higher order thinking processes.
      Strongly Agree   Agree   Neutral   Disagree   Strongly Disagree

7. The supervisory model used by my supervisor incorporated higher level critical thinking questions, such as "how?", "why?", and "what if…" appropriate to the situation and to my skill level.
      Strongly Agree   Agree   Neutral   Disagree   Strongly Disagree

8. The supervisory model used by my supervisor was an effective in strengthening my clinical competence.
      Strongly Agree   Agree   Neutral   Disagree   Strongly Disagree

9. The supervisory model used by my supervisor increased my confidence as a clinician.
    Strongly Agree   Agree   Neutral   Disagree   Strongly Disagree

10. The supervisory model used by my supervisor increased my preparation for my therapy
    sessions.

Strongly Agree   Agree   Neutral   Disagree   Strongly Disagree

11. The supervisory model used by my supervisor helped me to effectively evaluate my clinical
    strengths and weaknesses.

Strongly Agree   Agree   Neutral   Disagree   Strongly Disagree

12. The supervisory model used by my supervisor was an overall effective method of
supervision.

Strongly Agree   Agree   Neutral   Disagree   Strongly Disagree

Why?

_____
_____

Please feel free to comment on your supervisor's model of supervision (e.g. what you liked,
disliked, felt was helpful etc.):

_____
_____
_____
_____
_____
_____

Appendix C

Demographic data for participating on-site and off-site clinical supervisors.

	On Site Clinical Supervisors $n=5$					Off-Site Clinical Supervisors $n=4$		
	NSQF $n=2$		SQF $n=3$			SQF $n=4$		
Supervisor	A	B	C	D	E	F	G	H/I
Clinical Supervision Experience (years)	>30	14	30	25	7	>20	5	4 / >1
Licensed & Certified (years)	40	17	37	27	9	29	14	11/6
License (L) & Teaching Certification (T) Status	L&T	L&T	L&T	L&T	L	L	T	T/T

Appendix D

Demographic data for participating first and fourth semester graduate students.

	First Semester Students *n*=17				Fourth Semester Students *n*=3	
Clinical Practicum Setting	On-Site University Speech & Hearing Clinic *n*=17				Off-Site Externship School *n*= 2 Medical *n*=1	
	SQF *n*=9		NSQF *n*=8		SQF *n*=3	
	M	*SD*	*M*	*SD*	*M*	*SD*
Age (years)	22.22	.63	22.25	.83	23.33	.47
GPA	3.87	.10	3.70	.16	3.72	.15
GRE	290.44	9.44	292.38	3.53	294	2.16

Appendix E

Results of the CCTST and Simucase® for Fourth Semester Off-Site Students

	Pretest		Posttest		
Y2 CCTST and Simucase®	*M*	*SD*	*M*	*SD*	*p*-value
Simucase® (*n*=6)	66.33	8.76	62.67	9.77	0.879
CCTST (*n*=3)	18.67	3.51	16.00	4.58	0.827