Gender Differences in Medical Students: A Mixed Method Examination into the Potential Driver of Referral Patterns

David L. Thomas[2], W. Alex Edmonds[2], Tom D. Kennedy[2]

Corresponding author: Dr W. Alex Edmonds edmonds@nova.edu
Institution: 2. Nova Southeastern University
Categories: Medical Education (General)

Received: 27/06/2016
Published: 01/07/2016

Abstract

The fiscal impact of the referral patterns of physicians known to increase the overall cost of medical care, however the complex relationship between these variables is less understood (Stratton, McLaughlin, Witte, Fosson, & Nora, 2005; Lau, Alpert, Huckfeldt, Hussey, Auerbach et al., 2014). More specifically, when procedures that can be completed by the primary care physician are referred out it increases the cost of medical care at both a micro and macro level. However, it is difficult to isolate the exact variables and context that would precisely predict this pattern. This examination is an attempt to piece together the preliminary factors that may be exhibited medical-school training that would help to explain the referral-pattern process. A sequential multi-phase mixed method approach was utilized including the first two quantitative strands and the follow-up qualitative component. The first phase included an evaluation of male-female teams and who took the lead on the medical procedure. Results indicated that males took the majority of the lead on these procedures (38 for males and 26 for females). The next phase included an assessment of self-efficacy which revealed that on average females reported lower levels of efficacy, which was further elaborated by the performance evaluation. The final qualitative phase asked to elaborate on their self-efficacy while doing the required procedures. Results revealed that those students who did not take the lead did not feel as competent as their peers to perform the procedures. The findings can shed light on the gender-shift phenomenon, which is important to the receivers and the providers of medical services and potentially on referral-pattern processes.

Keywords: gender differences; self-efficacy; medical education; mixed methods; referral patterns

Introduction

Concerns about the cost of medical care have increased since the advent of the widespread public payment for
health-care services of the Medicare program beginning in 1963. Political contests frequently highlight the cost of medical care and strive to limit public dollars flowing into this economic sector. Most of these political debates underscore the ever increasing costs of health care. The cost of medical care, the reasons for that cost (e.g., the overuse of patient referrals), and its continuing increase are multifaceted and quite complex. Because of the social importance of effective health care, its rising costs warrant further academic inquiry.

The fiscal impact of referral patterns of physicians is a major determinant of the cost of medical care, however there is not enough research on these variables to fully understand their complex relationship (Stratton, McLaughlin, Witte, Fosson, & Nora, 2005; Lau, Alpert, Huckfeldt, Hussey, Auerbach et al., 2014). For example, referrals can superfluously increase the cost of medical care when they lead to patients having to seek out a different physician for a procedure that is well within the training and purview of the referring doctor. Interventions that reduce these types of unnecessary referrals are likely to lower the overall cost of health care, but is rarely addressed in medical-school education.

Virtually all American-medical students are trained in specific clinical procedures with the expectation that these procedures will be performed by a primary-care practitioner in a routine and effective fashion (Lockyer, Hodgson, Lee, Faremo, & Fisher et al., 2015). Once training is sufficiently advanced during residency, these procedures are reinforced, and physicians in-training are expected to demonstrate competence in their performance. To date, gender differences in relation to competency during medical-school training are rarely examined and could have an impact on referral patterns (Zhuge, Kaufman, Simeone, Chen, & Velazquez, 2011). Referring common medical procedures from one primary care practitioner to another primary care practitioner or to a specialist increases the overall cost of medical care (Lau et al., 2014). Interestingly, there is evidence of physicians who are trained in performing specific procedures but still refer the cases out because they don’t perceive they possess adequate training (Ringdahl, Delzell, & Kruse, 2006; Tucker, Diaz, Carek, & Geesey, 2007), while, furthermore, there is some evidence to suggest that female physicians are more likely to refer out (Chaytors, Szafran, & Crutcher, 2001). Research examining the gender differences may illuminate any disparities in these patterns.

Gender Discrepancies in Physicians

One way to reveal the potential cause of these unnecessary referrals, and the associated gender differences related to these referral patterns is to examine medical-school education trends between males and females. The current study may provide a platform from which to initiate the discussion. The objective is to start the process of determining whether or not the procedural referral patterns for female physicians were developed within the first few years of their medical school education. Along with these gender differences in referral patterns, there are other well-documented discrepancies between male and female physicians. It is important to examine some of these additional gender differences in order to provide some contextual clues to what may be occurring with the different referral patterns for male and female physicians.

One well known gender disparity is that female physicians tend to have lower total incomes than males even when adjusted for work hours (Esteves-Sorenson & Snyder, 2012; Jagsi, Griffith, Stewart, Sambuco, DeCastro, & Ubel, 2012; Rose, Sanghani, Schmidt, Karafa, Kodish, & Chisolm, 2015). This trend has been in place for some time as past research indicated that female physicians are remunerated less compared to their male counterparts (Croasdale, 2007; Jagsi, Griffith, Stewart, Sambuco, DeCastro, & Ubel, 2012). This income discrepancy begins at the earliest part of their career and increases with years of practice (Lo Sasso, Richards, Chou, & Gerber, 2011). Lo Sasso et al. indicated that contract negotiations at the beginning of a new physician’s career may account for initial gender gap differences, although the trend toward widening the gap is not easily explained. Rose et al. (2015) furthered the discussion by detailing the gender disparity stating that salaries and funded research are typically tilted in favor of
male workers. They also revealed that the financial ties to pharmaceutical and biomaterial companies favor male physicians while female physicians receive far lower compensation for similar exchanges. The gender disparity is a concern and perhaps these disparities can be targeted for further investigation during training (Zhuge et al., 2011), while longitudinally linking those findings to the career path that females tend to follow (Jagsi, DeCastro, Griffith, Rangarajan, Churchill et al., 2011). These financial and career-based gender disparities may impact the aforementioned differences in referral patterns and deserves further scientific inquiry.

If female physicians make referrals at a greater rate than male physicians do (Chaytors et al., 2001), then the rate is relevant to both overall referrals and referrals for procedures. It therefore follows that if a provider is referring well-compensated procedures rather than performing them herself, the total overall reimbursement to that provider will be less. The simple act of not performing heavily reimbursed events will lower the total income of the physician electing not to perform the procedures.

Research has also indicated that female physicians performed fewer highly-reimbursed procedures on their patients and made more referrals for procedural interventions than their male colleagues (Stratton et al., 2005; Weeks, 2006). These claims are still speculative and no etiologic study has been conducted to determine the root cause of this, although Rose et al. (2015) indicated there is indeed an issue. Likewise, no study has been conducted to determine when this begins in the career of a female physician. Neither has there been any work to determine if this is reversible during the course of medical education. This issue is socially relevant in the larger context because of the association between referrals and the increased cost of medical care.

Faculty are influential over medical students and can play a role in determining how young physicians will practice medicine and how referrals tendencies will manifest (Sethi, Schofield, Ajjawi, & McAleer, 2016; Zhaniga, Leeac, Gruppenad & Baae, 2013). As mentioned, observation of the educational process of medical students could lead to understanding of where and how referral patterns are conducted later in the students' independent practice. For example, if during the medical school experience, females tend to perform, on average, fewer procedures than their male counterparts, then medical school educators could theoretically have an influence on decreasing the cost of care and remediating the current escalation in costs due to excessive referrals. Research has demonstrated that medical educators can engage female medical students to become more active within educational environments (Wright et al., 2003), which is of interest to the framework of the current investigation. Consequently, a description of a conceptual framework that posits the association between gender dynamics and engagement in the medical-school setting is warranted.

Social Learning Theory in Medicine

As noted, we are not aware of any studies designed to determine the etiologic underpinnings regarding gender differences in physicians. Lo Sasso et al. (2011) pointed out that young physicians begin their career with a significant gender-based income gap and attributed this to negotiations for compensation in which women desire more intangibles. Although that concept is developed well in the article, the authors were at a loss to explain why the gender-based income gap becomes increasingly wider with time.

One theory that may explain or offer insight into this phenomenon specifically is Bandura's (1986) social learning theory—better known through the measure of self-efficacy. Bandura (1997) defined self-efficacy as the belief an individual maintains about his or her abilities in a domain-specific task. With this in mind, it is essential, therefore, to measure the relationship between the alacrity of performance and the individual's self-efficacy. Bandura acknowledged the fact that environments are not static, rather, certain individual personality traits alter not only the perception of the environment but also interaction with the environment. The central theme of his theory was that people are not simple passive-reactive organisms shaped and molded by their environment; there is a dynamic
interaction involving personal, behavioral, and environmental components, all of which modify the response. Flowing naturally from this is that the student who does not feel confident doing procedures will tend to not do them (i.e., perceptions of self-efficacy decrease, resulting in loss of motivation to conduct the procedure), leading to a cycle of feelings of inability (Gender Verger et al., 2009; Pelaccia et al., 2010; Skajkovic & Luthans, 1998).

The concept of self-efficacy has been widely researched in the context of medical education, specifically in the areas of interprofessional clinical training (Nørgaarda, Draborgbf, Vestergaardag, Odgaardch, Jensenbd & Jan Sørensen, 2013); alternative consultation formats (Apera, Reniersac, Koolead, Valckeae & Dereseaf, 2012); student evaluation of observers and patients (Ammentorp, Thomsen, Jarbøl, René Holst, Øvrehus & Poul-Erik Kofoed, 2013); academic achievement (Burgoon, Meece, & Granger, 2012); and family-centered care during bedside rounds (Young, Schumacher, Megan, Moreno, Brown et al., 2012).

Though, not a primary emphasis during medical education, it is important for graduates to develop anxiety-coping mechanisms and a high sense of self-efficacy (Braeckman, De Clercq, Janssens, Gehanno, Bulat et al., 2013). However, the issue is multifactorial, and only a small portion of those factors can be explored in any single investigation. For instance, no attempt has been made to discern if there are qualitative differences in the selection process of female medical students over male medical students that may screen out those individuals who would perform procedures with alacrity (Powis, 2015). Similarly, though not necessarily feasible, no attempt has been made to measure the impact of the extensive mandatory postgraduate training requirements and how they impact female physicians differently from their male counterparts (Mercer & Puddy, 2011). Further research could elucidate all of these proposed limitations as well as a variety of other areas that impact the disparity.

The second year of medical education is when students are initially exposed to and expected to participate in certain medical procedures. A program analysis of the second-year aspect of medical education could be instrumental in eventually determining whether procedures are performed by primary care providers or referred to others.

The purpose of current investigation was to explore the gender differences in M-2s' training and, based on the findings, discuss the potential associated long-term implications on medical referral patterns. Based on this premise, the following objectives guided the development of the methodology.

**Objectives**

1. Determine the proportion of complex procedures (specifically, chest tube insertion and central line placement) performed by
2. Examine the perceptions of self-efficacy in male and female M-2 students and associate the levels of self-efficacy with their performance evaluations.
3. Determine if the medical education of females is deficient in procedural training such that female students feel inadequate to perform typical procedures, (b) If the medical education in procedures is sufficient, ascertain if there a cultural or gender bias in the providers (faculty) of the training, and (c) establish if it is possible that the provision of education is bias free but the female recipient of the technical training imposes self-restrictions and limitations.

**Methods**

**Participants**
Participants were medical students in their second year of training, otherwise known as M-2s. A nonprobability sampling approach and convenience sampling strategy was used and was predicated upon teams randomly made up of one male member and one female member. The design of the study was broken down into three stages: (a) examine students in male-female teams, (b) administer a quantitative survey of the male-female teams, and (c) conduct a qualitative follow-up strand. A total of sixty four male-female teams participated in the first phase (32 male, 32 female), fifty students participated in the survey stage (22 male, 28 female), and 13 students participated in the qualitative phase (6 male, 9 female). The age breakdown across the three stages were: $M_{age} = 25.48$ years, $SD = 1.52$.

**Instruments**

A dichotomous scale was used for the initial stage. A simple assessment of which partner of the team actually performed the procedure was used, adding the line "Who performed the procedure? Male/Female (circle one)" to the existing assessment tool. No separate observational instrument was proposed for this study, only the insertion of the single observational inquiry as to which member of the male-female team performed the procedure. The response to the observational directive was recorded in frequencies. The medical course was designed so one member of the team performed the procedure and the other did important but nonprocedural performance duties. For instance, in placing a chest tube, the procedure’s performing member would make an incision in the faux skin, use a surgical spreading device called a Kelly clamp to dissect just above the rib, and insert the tube into the chest, guiding the tube into proper position. The nonprocedural team member would communicate with the faux patient, gain consent, sterilize and drape the area, and provide support to the procedure performing member of the team. The gender of the individual actually inserting and placing the tube was recorded on the observation sheet. Table 1 details the list activities executed in these sessions.

**Self-efficacy.** For the second component of quantitative phase, a self-efficacy instrument, originally constructed by Mavis (2001), was employed. Self-efficacy was determined by a Likert-scale survey instrument using 1 as a rating for no self-efficacy and 5 as a rating for complete self-efficacy. The scale used the term *confidence* in lieu of *efficacy* to make it clearer for the students. Specific items asked inquired if the student had confidence in several areas such as "manage therapeutic interventions safely, prepare clients for diagnostic procedures and treatments, and complete your assessment in a timely manner following agency protocol."

**Qualitative data.** The final phase of the study included a qualitative strand. The students who originally participated as M-2s were now considered M-3s. The M-3 students received a debriefing and critique as a normal part of the M-3 curriculum. This was performed by the same instructor who taught these students in their initial Introduction to Clinical Procedures course during the second year. This debriefing and critique is a normal part of the educational process and is an open-ended series of inquiries determined to assist the student in becoming more competent in clinical procedures.

In the past, questions concerning perception of bias by the teaching faculty were asked, but particular attention was paid to the concept of gender bias. During the required debriefing, attention was paid to responses that impacted the qualitative research questions attempting to ascertain if there was inadequate training of females, if there was bias during the training, if there was some internalization of roles that led to documented referral patterns as well as their performance in relationship to their belief in their self-efficacy. See table 2 for the qualitative sub-questions.

**Procedures**

**Design.** A multi-phase mixed method approach was utilized for this examination. The three phases of the study were conducted sequentially. The initial two phases were considered the quantitative strands and the follow-up third
phase was the qualitative component, which served as a means to confirm or provide further elaboration on the quantitative results. Initially, the quantitative phase had to demonstrate that the hypothesis was correct before the qualitative phase was implemented. The results from the quantitative strand would not reveal the root cause of the phenomenon alone; however, analysis of the observations proposed and a qualitative component were expected to provide further insight into the results.

**Phase I.** During the second year, students were introduced to a formal hands-on course in clinical procedures most commonly performed in a hospital setting. The course was designed to be accomplished just before the students began hospital clinical rotations. The procedures course involved presenting specific procedures to students in a demonstration fashion and then having students perform the procedure on purpose-built, computer-enhanced mannequins. The procedures included insertion of a chest tube, tapping a chest for an effusion, injecting a knee, insertion of central venous pressure lines, and other procedures most hospital-based physicians have familiarity with and that are performed in the private offices of many physicians.

The M-2s were graded during practicum exercises by senior faculty. Students were divided into small two-person groups for training and evaluation. Two-person teams were evaluated by the faculty member for competence in performing the procedure. Many groups, because of the demography of the students, were male-female teams. For this study, an additional observation was made by the evaluating faculty members. A determination was made as to whether the male or female member of the team actually performed the procedure.

The initial determination of how the teams were established was done by the student-services office. The M-2s were previously assigned to small units of 12-14 students by the academical society. The assignment process was randomized after assuring that different level undergraduate grade-point averages were represented within each academical society. Because of the random distribution, it was unlikely that any society was biased with predominantly previous high performers or predominantly previous low performers. A random distribution of previous performance levels should have been distributed throughout the societies. Random distribution of students in the academical societies permitted the ready employment of convenience sampling.

**Phase II.** During the clinical (third) year, students must return from their hospital rotations to the college and demonstrate at least basic competence in the clinical procedures to which they were exposed in their second year. At that time, the self-efficacy instrument was implemented. The self-efficacy instrument was used to determine if there was a correlation between self-efficacy and performance of the procedure, and if there were gender differences in terms of self-efficacy.

**Assessment of performance.** Performance assessment was a subjective evaluation by a senior faculty member very familiar with procedures and the developer of the procedures program. The students were evaluated as poor, which meant they had to repeat the procedure until at least an "OK" rating was achieved. The next higher level was OK, meaning the procedure was accomplished, but it was possible that speed, dexterity, patient comfort, sterility, or another major component was compromised. Good was the next highest level and indicated that the procedure was performed with competence in all its component parts without unduly traumatizing the patient or the patient's tissue. Excellent was the highest rating, indicating that the procedure was performed at the level of an experienced physician.

**Phase III.** The qualitative strand followed the quantitative phase by approximately 1 year, but not necessarily a calendar year. At this point, the M-2s were now classified as M-3s. This phase addressed the two main issues of (a) whether the medical education was adequately preparing the students to perform the selected clinical procedures, and (b) the thoughts and feelings associated with the preparedness. For this phase, the same faculty member used an oral survey asking team members open-ended questions. These M-3s had a debriefing and critique as a normal part
of the third-year curriculum. This was performed by the same instructor who taught these students in their initial Introduction to Clinical Procedures course during the second year. The debriefing and critique were a normal part of the educational process and were an open-ended series of inquiries to assist the students in becoming more competent in clinical procedures. The questions are seen in Table 2.

Results

For the quantitative part of this study, sixty four medical students in male-female teams were evaluated by a single observer over a 4-month period to determine which member of the team (male or female) would take the lead and actually perform the procedure. Students were observed during practice on seven cases (see Table 3).

Procedures for the burn cases that were expected to be performed and evaluated included patient evaluation; identification of certain breath sounds made in the chest by the lungs; endotracheal intubation, which involved placing a tube through the mock patient's mouth into the trachea, requiring a moderate amount of skill for the purpose of assisting the mock patient in breathing or controlling or overriding the patient's own breathing process; and removing potentially infected dead or nearly dead skin with a forceps scalpel and scissors.

Procedures for the congestive heart failure cases and myocardial infarction (heart attack) cases that were to be performed and evaluated were as follows:

1. Assessment of sounds the lungs made in response to too much fluid in the lungs.
2. Tilting the head of the bed up to at least 30 degrees to cause the fluid to sink by gravitational force and make exchange of air easier.
3. Identification of a specific type of heart sound associated with this condition. insertion of a special type of intravenous catheter to measure pressure impinging on the heart from the venous side.
4. Insertion of a catheter in a sterile fashion to drain urine away from the patient in such a way that the amount can be accurately measured.
5. Possible endotracheal intubation and cardiac resuscitation, which may include administration of medications intravenously or directly into the heart.
6. Closed chest cardiac massage; defibrillation; or other techniques.

Sixty-four students divided into male-female teams were evaluated on these procedures by a single senior faculty observer. During the observations, 38 males took the lead and 26 females took the lead. A statistical analysis of these results is presented in Tables 4 and 5. The data for these specific medical procedures were analyzed using a difference in proportions test.

Students were then asked to respond to the self-efficacy survey in addition to receiving an assessment of their performance. Data for specific medical procedures were analyzed using a difference in proportions test. The self-efficacy survey was administered to 22 male and 28 female students. The key statement on the 14-item self-efficacy study was the last one, "I feel quite anxious about performing these scenarios today." A score 4 or 5 on the 5-point scale was considered to be low or questionable self-efficacy. A score of 1 or 2 was considered to be high self-efficacy. A score of 3 was considered not revealing of self-efficacy. Three male and 10 female students rated their
feelings of self-efficacy as low (4 or 5), whereas 15 male and five female students rated their feelings as high (1 or 2). Of these, one female and two male students chose 1, representing very high. Only one female student selected 5 for all parameters, indicating very low self-efficacy feelings. Three male and 11 female students rated their feelings of self-efficacy in the neutral range (3).

Following the completion of the self-efficacy survey was the subjective assessment of performance. The faculty member evaluating the performance of all 50 students rated none as *excellent*, five male and eight female students as *good*, 17 male and 17 female students as *OK*, and three female students as *poor*; these three students had to repeat the procedure until they reached at least the OK rating. Of the students who rated themselves at the next highest extremes (4) or (2), 12 males rated themselves as next highest in self-efficacy and received a rating of OK (one was rated as good). Two female students rated themselves next highest in self-efficacy and received a rating of OK.

The assessment of performance and self-efficacy was analyzed using Fisher's Exact Test. Data were collapsed into two measures: high self-efficacy (a score of 1 or 2 on the Likert scale) and low self-efficacy (a score of 4 or 5 on the Likert scale). Results showed that females reported lower perceptions of self-efficacy (*p* < 0.05), whereas males reported higher levels of self-efficacy (*p* < 0.05; see Table 6).

**Qualitative phase.** Thirteen M-3s (7 male and 6 female students) were specifically asked to elaborate on their self-efficacy while doing the required procedures. Three female and three male students indicated that they did not take the lead in any procedures. All of these six students indicated that they did not feel as competent as their peers to perform the procedures.

All students who responded that they took the lead at least once indicated that they did not take the lead in other events. Overwhelmingly, 11 of the 13 the responses were attributed to a lack of self-efficacy. Comments included lack of preparedness, lack of confidence, and others "knew it better." Females reported to have lower self-efficacy than males. Further, female students tended to use the term "not sure [of their ability]" or "uncomfortable" more frequently than males. In fact, a theme that emerged from the qualitative data, indicated that men attributed their hesitancy to take the lead for more pragmatic reasons ("I need to observe a few more procedures" and "I wanted to discuss some specific elements of the procedure first") while females often expressed their tentativeness utilizing emotive terminology ("I feel uncomfortable", "I am unsure of myself" and "I don't want to make a mistake in front of my colleagues"). Although there were some gender differences regarding the specific way the discussed there caution with taking the lead, the overall theme was "lack of preparedness" from all respondents who took the lead at least but failed to take the lead in other procedures.

**Discussion**

There is some evidence in the literature that female physicians perform fewer procedures on patients than do their male counterparts. Specifically, female physicians are more likely to refer a patient requiring a procedural intervention than their male counterparts (Ringdahl et al., 2006; Zhuge et al., 2011). This referral pattern increases the cost of medical care because it engages another physician in the patient’s care and reduces the reimbursement to the referring physician as procedural interventions are reimbursed at much higher rates than contemplative interventions (Chaytors et al., 2001; Tucker et al., 2007).

Understanding the gender shift phenomenon is important to both the receivers and the providers of medical services as well as the cost impacts on medical care delivery. We assumed that the gender shift will be sustained and intended to determine if the shift was responsible for peripheral manifestations such as increases of referrals for services.
Although, we do understand that making the leap between medical-school training and gender discrepancies in referral patterns within practice is purely speculative—however, it does provide a foundation from which researchers can base their theoretical postulation and ideally utilize longitudinal approaches.

A medical education is considered an essential societal resource, and therefore, the majority of medical colleges are cognizant of both the value of the service provided and the need to accept only candidates capable of full matriculation and success (Bakke vs. Board of Regents, 1978). Most recently, nearly half of all graduates from American medical schools were female. In fewer than three generations, the typical American medical school class moved from an acceptance rate of 16% and a graduation rate of 11% for women to an acceptance and graduation rate of approximately 50%. Such a shift in demographics has changed and continues to change the face of American medicine. The shift in gender proportions within American medicine delivery seem neither a transient phenomenon nor an unusual aberration (Goldin, 2014; Hoffman & Tarzian, 2003).

It may be important to see if this pattern of behavior begins in or can be modified by formal educational techniques in a medical-school program. In the past, questions concerning perception of bias by the teaching faculty were asked, but for this study, particular attention was given to that issue. The findings suggest a call to develop a comprehensive list of core procedures while being sensitive to gender bias as a means to standardize procedural training within the medical education context—similar to what The Society of Teachers of Family Medicine (STFM) Group on Hospital Medicine and Procedural Training developed for the practice of family medicine (see Kelly, Sicilia, Forman, Ellert, & Nothnagle, 2009).

**Study Synopsis**

All teams were observed by the same instructor to determine teamwork, appropriateness of the procedure selected, and execution of the procedure. Participants in the procedures course were observed in the same way that they are usually observed in the course when no research is being conducted. The only difference was to note whether the female member of the team permitted or encouraged the male member of the team to execute the procedure.

This part of the study developed the research into a mixed-methods modality to explicate the expected findings in the first part of the study. The same instructor who facilitated the Introduction to Clinical Procedures in the 2nd year and collected the quantitative data was the instructor for the 3rd year for the demonstration of competence and collection of the qualitative data. Data analysis of the qualitative portion of this research suggested little gender or cultural bias in the teaching of the students, but real cultural bias in the perceptions of the students. This qualitative data analysis was initially performed with descriptive summaries of the M-3’s comments to define categories within those descriptions. Because the goal of the data interpretation was to find a theme or concept without forcing the verbal responses into arbitrary non-narrative coding, this part of the research did not lend itself well to computer-based analyses. Rather, an attempt was made to extract a theme analysis by immersing the researcher into the data.

The participants for the qualitative phase were M-3s during their mandatory requirement to return to the medical school campus and demonstrate competence in specific medical procedures. As in the quantitative phase, only the most difficult procedures were used for data collection, because these are these are most commonly referred to another physician by female physicians. Their characteristics and demography were identical to the M-2s except they were a year older and were successful through the 2nd year.

**Quantitative Phase I**

The first phase was an evaluation of which member of the male-female team performed the procedure. The first objective was to determine of the male-female team who would perform the burn case procedure with more
frequency. On average, the male member of the team performed the procedure 68% of the time. The second source of data was the evaluation and treatment of a sophisticated mannequin presenting with congestive heart failure because of an underlying myocardial infarction. The students initially had to recognize the congestive heart failure and treat that and recognize that although the patient was significantly improving, there was an underlying myocardial infarction that was going to lead to cardiac arrest and require more sophisticated interventions. Specific procedures involved endotracheal intubation and insertion of a central line. In this scenario, males took the lead 57% of the time and females took the lead 43% of the time.

The difference in males performing the procedures between the first and second scenario was clear and required exploration. The second phase was 2 months later in the M-2 curriculum. In the interim, students had been exposed to much more material in their classrooms but had also been involved in many small-group learning sessions with actual or actor human patients. Although there was no procedural component to these interactions, the students were far more conversant with how to verbally interact with a patient and gather important data from the history. Also, the second set of cases had fewer auditory and visual distracters such as the moaning, crying, and disrupted appearance. Although the second scenario was also a very critical medical situation, it was a much calmer environment with the ability to have verbal interaction with the sophisticated mannequin more like other non-interventional scenarios to which the students had been exposed.

Predicated on multiple years of experience and clinical impressions, the senior faculty member believed that males would perform procedures at a more frequent rate than females but that the difference would be in the 45 - 55% range. Two thirds to one third seemed excessive and may reflect overall evaluations stretching through many years and cycles of training. It is entirely possible as the year went on the male-to-female disproportion would reduce. Later in the 2nd year, it may have reflected only a small distinction.

Quantitative phase II

This phase was marked by the self-efficacy survey that self-identified males or females and then consisted of 14 statements for which the Likert scale was used. The first 13 questions were designed to give the instructor feedback concerning the course and its acceptance and are considered distracters. The last question was, "I am quite anxious about performing these scenarios today."

We anticipated that female medical students would permit male medical students to perform procedures at a greater rate than would occur by chance. This would tend to support the primary hypothesis of the study, that female M-2s permitted male M-2 counterparts to perform procedures when they were paired together. Additionally, a review of their self-efficacy demonstrated gender differences did exist. Same-gender comparisons were found to be correlated to performance as well which allowed us to rule out the reports of self-efficacy as a factor in gender differences in performing procedures, which reflects similar findings from past research (Dorya, Beaulieub, Pestiauxa, Pouchainc, Gayd et al., 2009). These data suggested that males have stronger positive feelings of self-efficacy than females with regard to procedural events in medicine. Consistent with Bandura (1997), one would expect greater motivation, greater persistence, but possibly less preparation (Artino, La Rochelle, & Durning, 2010).

To round out phase II of this study, a performance evaluation was implemented to assess how each team and each team member performed the procedure. The critical part of the analysis was matching the performance evaluation to the feelings of self-efficacy. Of the 15 male students who rated themselves with high efficacy, two received an evaluation of good (13%) with 13 receiving an evaluation of OK (87%). Of the five female students who rated themselves as high, three were evaluated as good (60%) and two were evaluated as OK (40%) in their performance of the procedures.
An analysis of the data suggested that males have very strong feelings of self-efficacy compared to females by a margin of 3:1. In spite of these strong feelings of self-efficacy, the performance evaluations indicated that two of the three females (66%) and two of 15 males were good (7%). This is a remarkable difference and may indicate that males have much higher feelings of self-efficacy than was reflected in their performance (Bierer, Prayson, & Dannefer, 2015). It is yet to be determined the reason for this large difference.

The trend identified in the data suggested that males may have an inflated sense of self-efficacy and females may undervalue their self-efficacy in the area of performing medical procedures. Far more males rated themselves as high self-efficacy but received the relatively modest evaluation of OK on their performance. Similarly, many females rated themselves as low in self-efficacy but performed adequately on the evaluation. Interestingly, the only poor evaluations came from females who had self-identified as low in self-efficacy. It may be that that the evaluation of self-efficacy is most accurate in the performance of female medical students presented with procedural events, however, gender bias is clearly an issue in the medical profession (Jagsi et al., 2012; Zhuge et al., 2011).

In terms of the question of bias, no M-3 student (male or female) perceived a gender bias. All students were asked about the value of the program, the teaching methods, the competency of the faculty, and gender or other bias. There may have been gender-driven expectations in the performance of the procedures, but no attempt in this study was made to assess this other than the observation that males had a much higher sense of self-efficacy in the ability to perform medical procedures than in the actual performance. As speculated, this may be a reflection of a perception of a societal gender role (Babaria, Abedin, & Nunez-Smith, 2009; Goldin, 2014). Despite the inquiry, we believe examinations need to be conducted in the form of a program evaluation particularly focused on biases that exist in curriculum. Although a bias may be present, a student may not be aware of such discrepancies. To circumvent this, the evaluation would need to include objective and subjective assessments from both internal and external reviewers.

Qualitative Phase III

Overall, the qualitative data provided additional support for the quantitative findings. There was a positive relationship between increased self-efficacy and taking the lead in procedures. While those individuals who indicated the least self-efficacy, also expressed the most fear and apprehension with taking the lead on procedures, which is not unusual for students to experience anxiety related to clinical training (Sarikaya, Civaner, & Kalaca, 2006). Some gender differences were found, however they were more related to the different ways that the male and female participants discussed why the felt they were not prepared to take the lead. Females used more emotive language, while their male counterparts tended to use more pragmatic responses to why they chose not to take lead. This is reminiscent to the imposter syndrome, which posits that high-achieving individuals ironically can struggle with self-efficacy and esteem thus harboring a fear of being exposed as "fraudulent" (Vergauwe, Wille, Feys, De Fruyt, & Anseel, 2015).

Recommendations for Future Research

Future research in this domain would be better suited with a larger number of participants and a longitudinal design would result in more useful data as would a greater variety of procedures. There does seem to be a gender difference in performing medical procedures by beginning M-2s, however this study did not examine this phenomenon in depth, thus negating the predictive of value of subsequent behaviors. Whether this would be reflected later in their career could have been determined in this study by simply noting which member of the third-year male-female teams performed the procedures. Certainly, linking the students to their residency years and noting their performance of procedures would be very revealing of whether this gender trend continues after the degree is conferred.
A longitudinal study beginning in the second year of medical school and following the students through graduation and into their residency programs could demonstrate that the findings of this study do not follow the fledgling physician throughout their career. The possibility is, rather, with more training, the conferring of the degree, and experiences with patient care self-efficacy in all areas including procedures is improved and the seeming gender difference elucidated in this study disappears.

Second, this research did not indicate whether or not there was a cultural bias within the training. Further research into this realm is needed if indeed the training is truly not culture neutral (Greer, Park, Green, Green, Betancourt, & Weissman, 2007). Third, similar to the second result, this research did not indicate the training was gender neutral or whether the performance was gender biased based on inherent perceptions of gender roles by the students. It is important to continue study designs in this area to attempt to determine if the findings of this study are continued throughout a physician’s lifetime of practice.

### Tables

#### Table 1

**List of Medical Procedures**

<table>
<thead>
<tr>
<th>#</th>
<th>Task</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Recognized the problem that the simulated mannequin patient presented.</td>
</tr>
<tr>
<td>2</td>
<td>Identified the procedure needed to assist in the correction of the problem.</td>
</tr>
<tr>
<td>3</td>
<td>Performed the procedure or procedures identified as necessary.</td>
</tr>
<tr>
<td>4</td>
<td>Reevaluated the status of the simulated mannequin patient after the procedure is performed.</td>
</tr>
<tr>
<td>5</td>
<td>Evaluated their own performance of the procedure.</td>
</tr>
<tr>
<td>6</td>
<td>Received the communications of the observing students and faculty after performing the procedure.</td>
</tr>
<tr>
<td>7</td>
<td>Translated this scenario into a real-life situation where they are the responsible party.</td>
</tr>
<tr>
<td>8</td>
<td>Established rapport with a patient in an extreme situation.</td>
</tr>
<tr>
<td>9</td>
<td>Evaluated their own problem solving skills.</td>
</tr>
</tbody>
</table>

#### Table 2

**Qualitative sub questions**

<table>
<thead>
<tr>
<th>#</th>
<th>Question</th>
</tr>
</thead>
</table>
1. Out of ____ cases presented, how many did you take the lead on and actually perform the procedure?

2. Explain in detail what your thoughts and feelings related to why you did NOT take the lead on specific procedures.

3. For the procedures you did take the lead on, did you feel prepared to perform the procedure? (elaborate on your response)

4. For the procedures you did NOT take the lead on, did you feel prepared to perform the procedure? (elaborate on your response)

### Table 3

**Procedures and Lead Team Member (by Gender)**

<table>
<thead>
<tr>
<th>Case</th>
<th>Male/female teams</th>
<th>Male leads</th>
<th>Female leads</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burn case 1</td>
<td>10</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>Burn case 2</td>
<td>8</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Burn case 3</td>
<td>9</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>Congestive heart failure and myocardial infarction</td>
<td>7</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Assessment of pulmonary crackles 1</td>
<td>10</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>Assessment of pulmonary crackles 2</td>
<td>10</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>Assessment of pulmonary crackles 3</td>
<td>10</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

**Note.** Burn case = Identification of respiratory problems, intubation, fluid resuscitation via central line, and skin debridement; congestive heart failure = assessment of pulmonary crackles, head of bed to 30 degrees, identification of an S-4 heart sound, insertion of an intravenous catheter then later a central venous catheter, insertion of a Foley catheter, intubation and cardiac resuscitation; assessment of pulmonary crackles = head of bed to 30 degrees, identification of an S-4 heart sound, insertion of an intravenous catheter then later a central venous catheter, insertion of a Foley catheter, intubation, and cardiac resuscitation.

### Table 4

**Statistical Analysis Using a Difference in Proportions Test**
<table>
<thead>
<tr>
<th>Case</th>
<th>Male (%)</th>
<th>Female (%)</th>
<th>Difference (%)</th>
<th>Lower 95% CI (%)</th>
<th>Upper 95% CI (%)</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burn Case 1</td>
<td>70</td>
<td>30</td>
<td>40</td>
<td>31</td>
<td>49</td>
<td>p &lt; 0.05</td>
</tr>
<tr>
<td>Burn Case 2</td>
<td>63</td>
<td>38</td>
<td>25</td>
<td>13</td>
<td>37</td>
<td>p &lt; 0.05</td>
</tr>
<tr>
<td>Burn Case 3</td>
<td>67</td>
<td>33</td>
<td>33</td>
<td>23</td>
<td>44</td>
<td>p &lt; 0.05</td>
</tr>
<tr>
<td>Congestive heart failure and myocardial infarction</td>
<td>57</td>
<td>43</td>
<td>14</td>
<td>0</td>
<td>28</td>
<td>NS</td>
</tr>
<tr>
<td>Assessment of pulmonary crackles 1</td>
<td>60</td>
<td>40</td>
<td>20</td>
<td>10</td>
<td>30</td>
<td>p &lt; 0.05</td>
</tr>
<tr>
<td>Assessment of pulmonary crackles 1</td>
<td>60</td>
<td>40</td>
<td>20</td>
<td>10</td>
<td>30</td>
<td>p &lt; 0.05</td>
</tr>
<tr>
<td>Assessment of pulmonary crackles 2</td>
<td>50</td>
<td>50</td>
<td>0</td>
<td>-10</td>
<td>10</td>
<td>NS</td>
</tr>
</tbody>
</table>

Table 5

*Self-Efficacy Categorized by Gender*

<table>
<thead>
<tr>
<th></th>
<th>High self-efficacy</th>
<th>Low self-efficacy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>No.</td>
<td>%</td>
</tr>
<tr>
<td>Female</td>
<td>6</td>
<td>35</td>
</tr>
<tr>
<td>Male</td>
<td>16</td>
<td>84</td>
</tr>
</tbody>
</table>

Table 6

*Statistical Analysis of Self-Efficacy Evaluation (in Percentages)*
<table>
<thead>
<tr>
<th></th>
<th>70</th>
<th>30</th>
<th>40</th>
<th>31</th>
<th>49</th>
<th>p &lt; 0.05</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burn Case 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Burn Case 2</td>
<td>63</td>
<td>38</td>
<td>25</td>
<td>13</td>
<td>37</td>
<td>p &lt; 0.05</td>
</tr>
<tr>
<td>Burn Case 3</td>
<td>67</td>
<td>33</td>
<td>33</td>
<td>23</td>
<td>44</td>
<td>p &lt; 0.05</td>
</tr>
<tr>
<td>Congestive heart failure and myocardial infarction</td>
<td>57</td>
<td>43</td>
<td>14</td>
<td>0</td>
<td>28</td>
<td>NS</td>
</tr>
<tr>
<td>Assessment of pulmonary crackles I</td>
<td>60</td>
<td>40</td>
<td>20</td>
<td>10</td>
<td>30</td>
<td>p &lt; 0.05</td>
</tr>
</tbody>
</table>

**Take Home Messages**

The fiscal impact of referral patterns of physicians is a major determinant of the cost of medical care

There may be a gender disparity in referral patterns and initial factors that explain this phenomenon can be revealed during medical-school training

During medical procedures tests, males on average take the lead more than females

Females reported lower levels of self-efficacy than males

Males and Females reported different reasons as to why they didn't take the lead on a procedure

**Notes On Contributors**

David Thomas, M.D., J.D., Ed.D., chair of Surgery and of the Correctional Medicine, College of Osteopathic Medicine. Served in the military in Vietnam as an U.S. Army Major. He also served for a decade in the Florida House of Representatives. Thomas was trained in general and chest surgery at the University of Miami and ophthalmology at the University of Florida. He was involved with NSU’s Center for Bioterrorism and All-Hazards Preparedness curriculum development and was chair of the NSU institutional review board. Thomas is currently a practicing attorney.

W. Alex Edmonds, PhD, BCB, is currently an associate professor of research at Nova Southeastern University. He graduated from Florida State University and received his doctoral degree in Educational Psychology with a minor in Statistics and Measurement. He also has over 10 years of experience in applying biofeedback in field and research applications. He has utilized biofeedback extensively with various types of athletes for performance enhancement, as well as stress-regulation techniques for individuals with type 2 diabetes and pain management for patients suffering from chronic pain. Dr. Edmonds is certified through the Biofeedback Certification International Alliance in general biofeedback.
Tom D. Kennedy, PhD, BCB, is currently an associate professor of research at Nova Southeastern University. He received his PhD in Counseling Psychology from the University of Miami and his MA in Clinical Psychology from Southern Methodist University. His clinical experience consists of providing neuropsychological assessments, behavioral medicine interventions, and group therapy in inpatient and outpatient settings, including the following: the University of Texas Southwestern Medical Center, University of Miami Mailman Center for Child Development, Jackson Memorial Hospital, University of Miami Institute for Individual and Family Therapy, and the Dallas County Jail. Dr. Kennedy is a licensed psychologist PY 8307 and is certified through the Biofeedback Certification International Alliance in general biofeedback and maintains a small private practice for children and adolescence.

Acknowledgements

Bibliography/References


http://dx.doi.org/10.3109/0142159X.2012.668627


http://dx.doi.org/10.1186/1472-6920-13-49


http://dx.doi.org/10.1111/j.1365-2923.2010.03712.x


http://dx.doi.org/10.1097/ACM.0b013e3181a8130c


http://dx.doi.org/10.1007/s10459-014-9531-7

http://dx.doi.org/10.1097/JOM.0b013e3182a50016


http://dx.doi.org/10.1002/ase.1283


http://dx.doi.org/10.1016/j.econlet.2011.12.133


http://dx.doi.org/10.1257/aer.104.4.1091


http://dx.doi.org/10.1007/s11606-007-0229-x


http://dx.doi.org/10.2147/AMEP.S13877


http://dx.doi.org/10.1111/j.1748-720X.2003.tb00057.x


http://dx.doi.org/10.3109/0142159X.2013.789495


http://dx.doi.org/10.1097/SLA.0b013e3182111120

Appendices

Declarations

The author has declared that there are no conflicts of interest.

This has been published under Creative Commons "CC BY 4.0" (https://creativecommons.org/licenses/by-sa/4.0/)

AMEE MedEdPublish: rapid, post-publication, peer-reviewed papers on healthcare professions’ education. For more information please visit www.mededpublish.org or contact mededpublish@dundee.ac.uk.