QUANTITATIVE GENDER AND ETHNICITY DIFFERENCES IN EMERGENCY

MEDICINE RESIDENCY APPLICATIONS

by

Steven Jay Warrington

Liberty University

A Dissertation Presented in Partial Fulfillment

Of the Requirements for the Degree

Doctor of Philosophy

Liberty University

2023

QUANTITATIVE GENDER AND ETHNICITY DIFFERENCES IN EMERGENCY

MEDICINE RESIDENCY APPLICATIONS

by Steven Jay Warrington

A Dissertation Presented in Partial Fulfillment

Of the Requirements for the Degree

Doctor of Philosophy

Liberty University, Lynchburg, VA

2023

APPROVED BY:

Eric Lovik, Ph.D, Committee Chair

Darren Wu, Ed.D., Committee Member

ABSTRACT

The purpose of this quantitative causal-comparative study is to examine if gender affects a medical student's electronic Standardized Letters of Evaluation, or the characteristics of language used, in their application materials to an emergency medicine residency program through the Electronic Residency Application Service. This study is important as research is needed to determine if there is a difference in ratings or language used in emergency residency applications based on a medical student's gender. The participants include 173 applicants from the 2022-2023 cycle that submitted material to a single emergency medicine residency program through the Electronic Residency Application Service. Instruments used include the electronic Standardized Letter of Evaluation and Linguistic Inquiry and Word Count 2015. Convenience sampling was used with data analysis involving descriptive statistics, a two-way ANOVA, and an ANCOVA. No difference was found between the aggregate eSLOE ratings based on gender, ethnicity, or gender and ethnicity. There was also no difference in use of communal language based on gender of students when controlling for total word count. These results contradict prior literature, though it is difficult to determine the reasoning for such findings or if it may be due to one of the studies limitations. Future research should evaluate these findings as a potential temporal or regional trend, address limitations found within this study, and seek to identify reasons for such contradictory findings.

Keywords: emergency medicine, gender, ethnicity, residency

Dedication

This work is dedicated to my family, without which none of this would be possible or worthwhile.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Analysis of Covariance (ANCOVA)

Analysis of Variance (ANOVA)

Association of American Medical Colleges (AAMC)

Electronic Residency Application Service (ERAS)

Electronic Standardized Letter of Evaluation (eSLOE)

Institutional Review Board (IRB)

Linguistic Inquiry and Word Count 2015 (LIWC)

National Residency Matching Program (NRMP)

Research Question (RQ)

Statistical Package for the Social Sciences (SPSS)

CHAPTER ONE: INTRODUCTION

Overview

The purpose of this quantitative causal-comparative study is to examine if an individual's gender affects the materials contained within a medical student's application to an emergency medicine residency. Specifically, this study will examine if an individual's gender and ethnicity affects the ratings found on their electronic Standardized Letters of Evaluation (eSLOE), or the characteristics of language used in their application materials, that are submitted through the Electronic Residency Application Service (ERAS). This chapter will provide a background on the application process to residency training along with a historical perspective to graduate medical education and residency training. Additionally, the theoretical frameworks for the proposed study are provided. The problem statement then details recent publications on the topic and is followed by the purpose of the study. After the purpose of the study, the significance of the study is discussed, and the research questions (RQs) are listed. The chapter then closes by defining key terms.

Background

Without residency training, a doctor is unable to practice medicine in the United States, placing their career and dreams in jeopardy while potentially leaving them over $200,000 in debt without a means to earn money (Phillips et al., 2019; Zavlin et al., 2017). Failing to attain a spot in a residency program occurs to approximately 20% of applicants and is a devastating situation that has led individuals to commit suicide (Ezimora, 2020; The National Residency Matching Program, 2021-a). The process of obtaining a residency training position is primarily done using ERAS and the National Residency Matching Program (NRMP) (Gruppuso & Adashi, 2017; The National Resident Matching Program, 2018; Wetz et al., 2010). Individuals submit application

materials to selected programs using ERAS, which also includes demographic information. Application materials are specified by the programs they are applying to, and generally include letters of recommendation, evaluations, and personal statements. After submission of materials, residency programs review the materials and offer interviews, which candidates are able to accept or decline, through ERAS. Following interviews, applicants rank the programs that they are interested in and would like to attend. Residency programs similarly rank applicants they would like to have in the residency program. Following the deadline of submission of rank lists by the applicants and residency programs, the NRMP uses an algorithm and system to match applicants with residency programs based on their ranked preference list. The NRMP's algorithm and system has been studied and resulted in a Nobel prize (Leopold, 2021; Roth, 2003; Roth & Peranson, 1997). Unfortunately, there is some evidence to suggest that the application process as a whole contains underlying biases based on gender, ethnicity, age, training background, and the presence of a disability (Craig et al., 2021; De Oliveira et al., 2012; Meeks et al., 2018; Ross et al., 2017). Additionally, the literature from some specialties shows that such biases lead to difficulty matching, though there is limited evidence within emergency medicine (Craig et al., 2021; Lakoff et al., 2020; Nagarajan et al., 2020).

Historical Overview

Medical education in the United States, and the North American colonies prior to that, predominately began as extended apprenticeships in the 18th century with few medical schools and no residency programs (Custers & ten Cate, 2018). Standards for residency programs began to be established in the early 20th century, along with states beginning to require some form of graduate medical education to obtain a medical license (Custers & ten Cate, 2018; Dugdale et al., 2019). As graduate medical education programs developed, the process of obtaining and

committing to one became earlier and more difficult, leading to the development of the origins of the NRMP forming in the 1950s (Roth, 2003).

Historically, biases have been present throughout the development of medical education, with clear documentation of ethnic and gender-associated biases in medical education. There were essentially only two medical colleges for black trainees until around 1970, with black physicians also facing difficulty obtaining privileges at white hospitals during the same time (Byrd & Clayton, 2001). Females also faced similar challenges, with only two percent of individuals willing to recommend a career in medicine to females in the 1950s (Boulis & Jacobs, 2011). With the rise of social justice and awareness, these issues have become more prominent with attempts to draw attention to the problem and identify methods to improve the situation (Ansell & McDonald, 2015; Boatright et al., 2018; Girod et al., 2016). Despite the progress made, it seems there are still biases present in the application process in at least some specialties (Grimm et al., 2020; Powers et al., 2020; Ross et al., 2017).

Society-at-Large

While biases within an application process are certainly impactful to the applicants and programs the applications are used for, biases within the application process to a residency program have the potential to impact society-at-large through multiple means. The first potential impact on society is that the initial and ongoing accreditation status of a residency training program requires residency programs and their sponsoring institutions to recruit and retain a diverse workforce, including residents, faculty, and other support staff (Accreditation Council for Graduate Medical Education, 2020). Should a program not work towards ensuring a diverse group of residents through the residency application process, they may have their accreditation status revoked. Further, the most recent Accreditation Council for Graduate Medical Education

(2020) Common Program Requirements includes the expectation that a residency program's annual program evaluation includes an assessment of its efforts to attract, obtain, and retain a diverse workforce. Should programs be unable to attain or maintain accreditation, the physician workforce shortage may worsen, leading to difficulty for individuals to attain healthcare (Zhang et al., 2020).

In addition to the potential for worsening the physician workforce shortage, a lack of diversity among physicians may lead to higher mortality and worse health outcomes (Greenwood et al., 2018, 2020). When looking specifically at gender, the physician's gender has been shown to influence practice patterns, especially when there is a concordance with the patient's gender (Eggermont et al., 2018; Greenwood et al., 2018). Such influence has been demonstrated to lead to differences in prescribing patterns of antibiotics, which may worsen drug resistance patterns of bacteria (Eggermont et al., 2018). It has also been demonstrated to impact more critical clinical situations, such as patients suffering from heart attacks (Greenwood et al., 2018). In patients suffering from heart attacks, females that are treated by male physicians that have less exposure to female physicians and female patients had higher mortality rates. Not only do female physicians then help by providing that exposure to improve outcomes, but they also were not found to have disparities in survival outcomes of their male and female patients suffering from heart attacks.

Theoretical Background

This study focuses on feminist and homosocial theories to explore underlying biases and their associated advanced discrimination. Within feminist theories the theories of postmodern feminist and inequality regimes are considered. Afterwards, homosocial reproduction and homophily are discussed.

Feminist theories, specifically postmodern feminist theory and inequality regimes, are directly applicable. The importance of feminist theories in considering biases comes from the feminist theories focus on the nature of social constructs and ending oppression and exploitation (Sharma, 2019). The specific feminist theories applied are postmodern feminist theory and the theory of inequality regimes. Frug's (1992) postmodern feminist theory highlights the importance of language and the interpretation of such language in the social construct of gender. Intersectionalism falls within the postmodern feminist theory according to Davis (2008), allowing other biases being explored (e.g., ethnicity, age, disability) to also be evaluated from a linguistic and socially constructed perspective (Frug, 1992). The theory of inequality regimes was developed by Acker (2006), which was also built out of feminism to address intersectionality. Inequality regimes are the practices and processes in place that maintain the inequalities within an organization (Acker, 2006). These inequality regimes are linked to the organization and to the environment the organization exists within, with gender being a focus, though intersectionality allows it to be applied more broadly (Koivunen et al., 2015; Whitehead, 2013).

The second set of theories applied to the study include homosocial reproduction and homophily. The ongoing replacement of the dominant majority (e.g., white males) with individuals that are similar (i.e., white males) has been described as homosocial reproduction (Dressel et al., 1994). While not explicitly linked by Dressel et al. (1994), homophily is similar, referring to individuals often interacting with those that are similar to themselves (Kleinbaum et al., 2013). Such phenomena have been demonstrated in a variety of professional settings and occur through multiple mechanisms (Clifton et al., 2019; Fu et al., 2012; Kleinbaum et al., 2013). These theories are useful to apply to the study due to their applicability to biases against various

groups, such as groupings based on gender. They may provide insight as to why a male dominated program, institution, or field remains male dominated throughout years and attempted reforms.

Despite the process of securing residency training through The Match being more standardized than it was a century ago, there are still suggestions that biases are present. These biases may be socially constructed and language-focused, which would potentially lead to ongoing inequalities based upon categorization of individuals (i.e., gender) and be best evaluated through the theories discussed earlier. Unfortunately, there is a dearth of recent literature related specifically to The Match for residency training positions in emergency medicine.

Problem Statement

Unintentional and intentional biases have been documented previously within The Match across multiple specialties such as radiology, orthopedics, and pediatrics (Babal et al., 2019; Grimm et al., 2020; Maxfield et al., 2019; Powers et al., 2020). There is limited research available on the topic in emergency medicine, though what is available suggests that it may also suffer from similar biases (Hopson et al., 2019; Li et al., 2017;). Chapman et al. (2020) suggest that both author and applicant genders were associated with bias within letters of recommendation. Another recent study identified both gender and ethnicity of an applicant influencing the process of The Match in emergency medicine (Hopson et al., 2019).

Recently there has been an update to the application process of The Match in Emergency Medicine as the Standardized Letter of Recommendation changed to the Standardized Letter of Evaluation and then to the eSLOE (Jackson et al., 2019). Jackson et al. (2019) noted that with the most recent modification to the process there is an improved spread of ratings, though the authors did not study biases or aspects outside the eSLOE such as gender or ethnicity. In addition

to the updated processes in the match, attention has been drawn to gender and ethnicity issues within the match and leadership in emergency medicine (Boatright et al., 2018; Madsen et al., 2017). The problem is that more research is needed to determine if there is a difference in evaluations or language used in residency applications based on a medical student's gender and ethnicity (Boatright et al., 2018; Chapman et al., 2020; Hopson et al., 2019; Li et al., 2017; Madsen et al., 2017; Meeks et al., 2018).

Purpose Statement

The purpose of this quantitative causal-comparative study is to examine if gender affects a medical student's evaluations, or the characteristics of language used, in their application materials to an emergency medicine residency program through the ERAS. Gender serves as the first independent variable and represents the socially constructed differences between men and women (Risberg et al., 2009; Verdonk et al., 2009). Self-identified ethnicity serves as the second independent variable and refers to the multi-faceted identity of an individual (Santos et al., 2010). Options within ERAS for ethnicity include American Indian / Alaska Native, Asian, Black / African American, Hispanic / Latino / of Spanish Origin, Native Hawaiian / Other Pacific Islander, White, Other, and unreported / unknown. The first dependent variable is the eSLOE aggregate rating, which is the mean of ratings provided (Hopson et al., 2019). The second dependent variable is the communal language used within the narrative section of an emergency medicine residency application measured by Linguistic Inquiry and Word Count 2015 (LIWC). Communal language is the capturing of communion in natural language, with communion or communality referring to establishing close relationships, becoming part of the community, and having a focus on being kind, sympathetic, and helpful (Grimm et al., 2020; Pietraszkiewicz et al., 2019). The number of words within narrative portions of an application

serve as a covariate when evaluating the characteristics of language used in this study. The population is graduating medical students applying for an emergency medicine residency program using ERAS during the 2022-2023 application cycle, with a convenience sample using individuals applying to a single institution's emergency medicine residency program.

Significance of the Study

Obtaining a residency training position is an important milestone in the medical career of physicians whose goals are to practice in the United States, with decisions on who to interview and match based significantly upon application materials (Dugdale et al., 2019; The National Resident Matching Program, 2020). Despite emerging literature in some medical and surgical fields on biases within the application materials, there is no current evidence on the presence or absence of biases in applications to emergency medicine residency programs in the context of the current application process (Babal et al., 2019; Durham et al., 2018; Turrentine et al., 2019). This research will generate evidence to determine if there are still biases in emergency medicine residency applications based upon gender as identified prior by Girzadas Jr et al. (2004) and Li et al. (2017). The study has the potential to be generalizable to other residency programs across the United States considering they use some of the same application materials (e.g., medical student performance evaluation) authored by the same individuals (i.e., medical schools designate author(s) of medical student performance evaluations for the entire student body). There is also the potential that this study may be generalizable to domains outside medical education such as application processes to higher education systems and human resources considering the foundation in homosocial reproduction and feminist theories that contain a focus on intersectionality.

Only once biases and barriers are identified are methods to overcome them able to be

developed. This project represents the first step in overcoming such biases in applications to emergency medicine residency programs, which may lead to a more diverse pool of trainees. Ensuring a diverse pool of trainees is of the utmost importance both from a regulatory standpoint and for the outcomes of our population's health (Accreditation Council for Graduate Medical Education, 2020; Eggermont et al., 2018; Greenwood et al., 2018; McDade, 2019).

Research Questions

RQ1: Is there a difference among the electronic Standardized Letter of Evaluation aggregate ratings of male and female medical students applying to an emergency medicine residency based on their self-identified ethnicity?

RQ2: Is there a difference in communal language used in narrative portions of applications to an emergency medicine residency found by Linguistic Inquiry and Word Count 2015 among male and female medical students when controlling for total word count?

Definitions

1. *Common Program Requirements* – The Accreditation Council for Graduate Medical Education (n.d.) set basic standards and expectations for the training and preparation of physicians (residents and fellows). Such requirements are used to ensure clinical and learning environments are developed and maintained to promote the meaningful interactions with faculty and patients required to develop the knowledge and skills required to be successful following training (Accreditation Council for Graduate Medical Education, n.d.).

2. *Communal Language* – Language and wording that is relationship-oriented such as kind, sympathetic, reliable, and friendly (Grimm et al., 2020; Pietraszkiewicz et al., 2019).

3. *Disability* – Kiuhara and Huefner (2008) consider a disability to be a mental or physical impairment that limits at least one major life activity.

4. *eSLOE Aggregate Rating* – An eSLOE aggregate rating can be found by calculating the simple average of scores on the eSLOE for an individual (Hopson et al., 2019).

5. *Ethnicity* – Race and ethnicity are often used interchangeably, however ethnicity is the more accurate scientific term that refers to the identity of an individual and is multi-faceted (e.g., appearance, nationality, language, etc…) (Santos et al., 2010). The American Psychological Association (2019) note that ethnicity is a social construction that has changed over time and can be difficult to define with varying inclusion of physical traits as a component of ethnicity.

6. *Gender* – While sex and gender are often used interchangeably, gender refers to the socially constructed differences between men and women (Risberg et al., 2009; Verdonk et al., 2009).

7. *Race* – Race is a non-scientific term generally applied to individuals based on general physical characteristics that has been used somewhat interchangeably with ethnicity, though ethnicity is the scientifically appropriate term (Santos et al., 2010).

8. *Sex* – Verdonk et al. (2009) note that sex is based on biological characteristics such as chromosomes and organs.

9. *Sponsoring Institution* – The Accreditation Council for Graduate Medical Education (2020) defines a sponsoring institution as the overarching entity that is responsible for a graduate medical education training program academically and financially.

10. *Word count* – The number of words within a certain text is the word count (Turrentine et al., 2019).

CHAPTER TWO: LITERATURE REVIEW

Overview

The Match, also known as The National Resident Matching Program's (NRMP) Main Residency Match, is the culmination of medical school, serving as the process by which individuals may enter graduate medical education training programs (i.e., residency programs) (Almarzooq et al., 2021; Wetz et al., 2010). A thorough review of the literature was performed to explore the problem of potential biases within the application process of The Match in emergency medicine. This chapter will review current literature associated with the topic of study. Initially, the theories relevant to biases within the application process of The Match, specifically postmodern feminist theory, the theory of inequality regimes, and homosocial reproduction will be reviewed. Following the discussion on the identified theories, a synthesis of the recent literature available on biases within the application process of The Match surrounding gender, ethnicity, and training background will be provided. Ramifications surrounding the short and long-term potential impact of such biases in the application process of The Match will be addressed. Finally, a gap in the literature will be identified, demonstrating a need for the current study.

Theoretical Framework

This study has foundations within feminist theories and the theory of homosocial reproduction. While numerous feminist and gender-based theories may be applied to medical education, this study primarily utilizes Frug's (1992) postmodern approach and the theory of inequality regimes that was developed by Acker (2006) (Risberg et al., 2009; Sharma, 2019). In alignment with those, the theory of homosocial reproduction is also incorporated.

Feminist Theory

Despite feminism existing for centuries, identifying an all-encompassing definition is a difficult process, leading to a near impossibility of identifying one unifying theory (Bennett, 1989; Sharma, 2019). Further, as movements and focuses have changed, ranging from suffrage through workplace equality and acknowledging the radicalized and queer, numerous feminist theories evolved to address such issues (Sharma, 2019). With that in mind, one can instead approach the idea of feminism and feminist theory instead as a family of feminist theories that share similar concepts, such as gender being a social construct and a focus on social change (Acker, 2006; Ebert, 1991; Frug, 1992; Risberg et al., 2009).

Considering gender as a social construct refers to the idea that male and female differences are not engendered from a biological standpoint, but instead due to socially constructed differences (Frug, 1992; Risberg et al., 2009). Even as early as the writings of de Pizan in the 15th century this was seen when she focused on disproving the notion that being female was inherently associated with characteristics of weakness and evil (Bennett, 1989). The topic of gender as a social construct has become more refined over the centuries, with more recent literature delving into better understanding and combating the labels, characteristics, and stereotypes of one's sex (Frug, 1992; Risberg et al., 2009).

Postmodern Feminist Theory

Frug (1992) helped to develop the postmodern feminist theory in her manifesto that was published as an unfinished draft due to her murder. Within that manifesto, Frug founds her discussion upon two principles. The first principle is that human experience is "inescapably within language" (p. 1046), highlighting the importance of language's constructive function. She notes the impetus for the principle is not for a fixed definition on differences between sexes, but instead to focus on the ongoing interpretive difficulties surrounding sex-related differences. This

understanding of an ongoing interpretive struggle over meaning aligns with the resistance subset of postmodernism, recognizing language's meaning comes as a result of social struggles (Ebert, 1991). The second principle that Frug identifies is that male and female identity differences are semiotic, and thus based within signs and symbols that must then be interpreted. This importance of signs within postmodernism is also detailed by Ebert (1991), who notes the signified of the sign is developed out of social conflict and struggle.

Considering this focus on understanding gender through the context of associated social struggles, it leads somewhat naturally to intersectionality. The concept of intersectionality referring to the interaction of race, gender, culture, and other components of individual's lives and society at large (Davis, 2008). While Sharma (2019) maintained intersectional feminism as its own theory, Davis (2008) believed intersectionality fit within postmodern feminist theory. One can even see intersectionalism when reviewing Frug's (1992) writing as she notes the effects of age, race, class, religion, and a host of other aspects of a female's life.

It is evident that these underlying principles of postmodern feminism have had a role in shaping research evaluating gender differences. Sharma (2019) notes the importance of utilizing feminist theories in working towards gender and ethnicity-based equity within medical education. Such focus on inequities is necessary considering recent literature in undergraduate medical education that while gender inequities have improved, the ethnic equity gap has continued to grow (Morris et al., 2021; Talamantes et al., 2019). Graduate medical education has also examined gender and ethnic inequities in medical education using principles underlying postmodern feminism.

Within graduate medical education, many specialties have recognized the importance of language that postmodern feminism highlights, with general surgery, ophthalmology, and

pediatrics, being some of the specialties with recent literature on the subject (Babal et al., 2019; French et al., 2019; Lin et al., 2019). Grimm et al. (2020) found language differences in letters of recommendation used to apply to diagnostic radiology residency based creating biases based on gender and ethnicity. Multiple groups have demonstrated language differences in letters of recommendation used to apply to general surgery residency programs as well (French et al., 2019; Turrentine et al., 2019).

Recent work continues to advance using Frug's (1992) underlying principles of postmodern feminism. Within orthopedics, for instance, not only have the differences in language used to describe candidates been brought to light, but also the potential variation in the interpretation of terms and traits when they are used to describe males as compared to when they are being used to describe females (Kobayashi et al., 2020; Leopold, 2020). In general surgery, Ostapenko et al. (2018) studied the gender-based language differences in how applicants conveyed their rationale for such a career choice in personal statements submitted with residency applications.

Unfortunately, not all recent literature has been as progressive to consider the importance of such linguistic or socially constructed differences. Aisen et al. (2018) recently published on gender differences within urology, however limited evaluation to national board examination scores and achievements. Gong et al. (2019) evaluated gender differences in ophthalmology training, however limited it rates of procedure performance.

The proposed research questions relate to postmodern feminism in evaluating language differences, specifically communal terminology which has been associated with females and female-dominated roles (Pietraszkiewicz et al., 2019). Specifically, the proposed research seeks to evaluate language differences used in applications and recommendations, which Sharma

(2019) points out as an example of postmodern feminism in medical education. The study may help to advance the theory of postmodern feminism by evaluating ethnicity with gender as Frug (1992) pointed out in her original publication.

Inequality Regimes

Arising from feminism, though appearing more intersectional in nature, the theory of inequality regimes serves as the second feminist theory to build the study on (Acker, 2006). The theory of inequality regimes was developed by Acker (2006) to specifically address intersectionality while providing a methodology to analyze organizations for barriers to equality. Inequality regimes themselves are "defined as loosely interrelated practices, processes, actions, and meanings that result in and maintain class, gender, and racial inequalities within particular organizations" (p. 443). Such regimes are not static, nor are they isolated, as they are linked to an organization's surrounding society. With that noted, inequality regimes vary greatly considering they develop and evolve over time with an organization. Importantly, despite Acker proclaiming the development of inequality regimes from a feminist perspective, the theory's intersectionality component ensured it built support outside of just gender-related struggles (Koivunen et al., 2015; Whitehead, 2013).

The barriers identified by Acker (2006) have since been applied with the analysis of organizations as intended and include (a) control and compliance; (b) bases of inequality; (c) legitimacy of inequalities; (d) visibility of inequalities; (e) practices and processes of an organization that produce inequality; (f) shape and degree of inequality, also considered as the organizational structure (Gustafsson, 2018). Interestingly, Acker viewed these barriers as a means to reproduce inequality, though they could just as easily be viewed as barriers to propagate homosocial reproduction. Such barriers are widely applicable, with potential to

reinforce social divisions such as gender, ethnicity, or age (Koivunen et al., 2015).

Koivunen et al. (2015) point to the impact of inequality regimes on application processes from a general standpoint, with Reimann and Alfermann (2018) focusing on the impact on application processes within medicine. It is not just the application processes in medicine that have been informed by inequality regimes though. Masood (2019) published on gender and ethnic disparities based on inequality regimes for Pakistani female physicians, as did Mohsin and Syed (2020). Reimann and Alfermann (2018) highlighted discriminatory processes in the setting of Germanic female physicians with consideration of the existing inequality regime and discrimination. The discrimination is not limited to specific subgroups though, as Hu et al. (2019) found approximately 32% of general surgery residents faced discrimination due to gender and approximately 17% facing it due to ethnicity. Dimant et al. (2019) found some individuals censured themselves, such as with language used, in medical education due to barriers within their training.

This research study has the potential to advance the theory of inequality regimes by investigating Acker's (2006) barriers in the process of transitioning into graduate medical education that are responsible for continuation of gender and ethnic-based disparities. The research alone works to combat Acker's barrier associated with the inequality's invisibility by nature of focus on the potential problem. It also will examine additional components of inequality regimes, specifically the bases of inequality within emergency medicine applicants, along with the organization processes component (Acker, 2006).

Homosociality

Dressel et al. (1994) wrote on homosocial reproduction within academic institutions as a form of advanced discrimination. To them, homosocial reproduction was the replacement of

dominant group members, often white males in academia, with new members from the same dominant group (i.e., white males). As compared to Acker's (2006) barriers leading to inequality regimes, Dressel et al. describe homosocial reproduction as a subtle and insidious process that has a cumulative effect of inequity. They refer to these seemingly innocuous processes and rules, which at face value appear neutral yet have different effects based on belonging to a majority or minority group, as advanced discrimination. Interestingly, Dressel et al. believed their philosophy of advanced discrimination built upon some of Acker's (1990) earlier work on gender stratified organizations. Dressel et al. published on it prior to Acker's (2006) inclusion of intersectionality in her theory of inequality regimes. Homosociality has been considered as a barrier for females within Acker's (2006) framework of inequality regimes though (Carvalho et al., 2019).

While Dressel et al. (1994) did not explicitly link their theory of homosocial reproduction with homophily, the two are similar (Kleinbaum et al., 2013). Homophily, the tendency for individuals to interact with those that are similar to them, has been modelled both as a general phenomenon and specifically in professional hierarchies (Clifton et al., 2019; Fu et al., 2012). The literature suggests it occurs through two mechanisms, including choice and induced homophily (Kleinbaum et al., 2013). Choice homophily is the preference to associate with those that are similar, while induced homophily refers to individuals finding themselves in situations where they are surrounded by similar others to a disproportionate degree (Kleinbaum et al., 2013).

Homosociality has advanced the understanding of disparities within professional hierarchies, and specifically within medicine and academia (Clifton et al., 2019). It has also been used to advance the understanding of success within the continuum of medical education (Grace,

2018). Douglas and Hendrix (2021) specifically note challenges of homophily within the context of interviewing for residency programs for black individuals. Simons et al. (2021) consider issues surrounding homophily within the context of interviewing for residency programs for all under-represented individuals, inclusive of ethnic, gender, and socioeconomic perspective. In addition to advancing understanding surrounding the interview process, homophily has also advanced understanding of the selection process for residency (Gennissen et al., 2021).

The proposed research relates to homosociality as it explores the potential of advanced discrimination, as described by Dressel et al. (1994), occurring within the application process to an emergency medicine residency. The study may also elucidate additional rationale for prior published literature, such as additional mechanisms not identified by Gennissen et al. (2021). Further, it may provide validity evidence for prior published literature such as by Douglas and Hendrix (2021) or Simons et al. (2021).

Within Medical Education

Homosocial reproduction has been directly linked with inequality regimes and the socially constructed differences that Frug (1992) highlighted as part of her postmodern feminist manifesto (Carvalho et al., 2019). As these three theories are closely aligned and overlap at points, they can be used together when examining issues such as gender and ethnic-based disparities within medical education. Each individual theory included discussion of how it has helped to advance or inform the existing literature, in addition to how the proposed study may help to advance understanding of the theory.

These concepts can be seen in recent medical education literature as impacts of gender and ethnicity are studied (Aggarwal et al., 2018; Grimm et al., 2020). Frug's (1992) principles are evidenced by recent literature examining language variations attributable to gender or

ethnicity (Aggarwal et al., 2018; Grimm et al., 2020). As the studies and literature focus on biases that impact the transition from medical school into the workforce representing potential disruptions of existing inequalities of organizations, acknowledging inequality regimes and the encompassing potential barriers is important (Acker, 2006). Such barriers and examinations are not limited to applicants however, with Li et al. (2017) and Girzadas et al. (2004) considering the impact of instructor or leader gender and potential impact of homophily.

Gender and ethnic-associated disparities are well documented in undergraduate and graduate medical education (Grimm et al., 2020; Morris et al., 2021; Talamantes et al., 2019). Clifton et al. (2019) found through modelling that resolution of disparities within a profession is not inevitable, and that deliberate intervention may be required. This study utilizes the discussed theories to identify a potential source of bias within the application process to emergency medicine residency programs, which could prompt development of a deliberate intervention for use in closing the gap to equality.

Related Literature

The related literature explains topics relevant to the study and provides a review of recent research that has been done within the associated foci. First, a discussion on the psychological domains of words is important due to the proposed research's focus on communality in language. A discussion of The Match follows due to the importance of understanding the process as the study focuses on materials used for The Match. Next, gender and ethnic biases that are known to occur within The Match are discussed while providing recent research that has been done on those topics. Following that, other areas of bias such as from the nature of one's medical training (e.g., allopathic, osteopathic, international, etc…), presence of a disability, or an individual's age are covered. Finally, the short and long-term impacts of inequity in the application process of

The Match in emergency medicine are discussed.

Psychological Domains of Words

Linguistic analysis has been used in research for over a century with uses ranging from the diagnosis of patients to social relationships (Pennebaker et al., 2015; Tausczik & Pennebaker, 2010). Developing validity evidence on the psychometric properties of words has been noted to be more difficult than a standard instrument though due to multiple factors (Tausczik & Pennebaker, 2010). LIWC is one instrument with numerous word categories having been developed using many studies of psychological domains (Figueiredo, 2020; Tausczik & Pennebaker, 2010). Examples of categories of words with focused research done on them include agency and communal language (Pietraszkiewicz et al., 2019).

Agency and communion represent the most essential aims pursued throughout life including goal achievement and relationship forming (Pietraszkiewicz et al., 2019). Agency refers to an individual's independence and control of the environment, and has been aligned with male stereotypes of decisiveness and an achievement-oriented focus (Grimm et al., 2020; Heilman, 2001; Pietraszkiewicz et al., 2019). Communality has been aligned with female stereotypes of relationship focused individuals, with focuses on helpful, belonging, and subordinating their own needs (Heilman, 2001; Li et al., 2017; Pietraszkiewicz et al., 2019).

Medical School

Individuals applying to The National Resident Matching Program's Main Residency Match, which is used to attain a residency position in emergency medicine, must have attained admissions to, and be graduating from or completed, medical school or a fifth pathway program (The National Resident Matching Program, 2016). The National Resident Matching Program (2016) categorizes medical schools to include United States allopathic medical schools,

osteopathic medical schools, Canadian medical schools, and international medical schools. There is little difference between the education in United States allopathic medical schools and osteopathic medical schools, a form of medical education established and maintained in the United States (Chen & Mullan, 2009). The primary training difference is in the underlying philosophy in osteopathic medical schools that the musculoskeletal system has a heightened role in health and disease (Chen & Mullan, 2009; Gevitz, 1994). International medical education is more variable, with different processes to determine the number of medical students annually through post graduate training processes (Miani et al., 2015). The fifth pathway program was stopped in 2009, but was a program which issued degrees through a mixture of medical school training outside the United States followed by clinical training supervised by an American medical school for one year (Stimmel et al., 1981; The National Resident Matching Program, 2016). Most individuals applying in The National Resident Matching Program are from United States allopathic medical schools or osteopathic medical schools (The National Residency Matching Program, 2020, 2021-b).

The process of admissions to medical school involves evaluating an individual from a cognitive and non-cognitive standpoint (Benbassat & Baumal, 2007). There are a number of different methods used in the selection process by medical schools with varying goals (ten Cate & Smal, 2002; Urlings-Strop et al., 2009). Unfortunately, the admission process to medical school in the United States has been critiqued to potentially be biased from multiple perspectives such as gender and ethnicity (Capers et al., 2017; Simmenroth-Nayda & Görlich, 2015; Wilkinson et al., 2014). Morris et al. (2021) found that over the past few decades the gender inequity in medical school admissions has significantly improved, but that the same cannot be said for ethnic-based inequity. The proportion of medical students in the United States from an

ethnically underrepresented minority group has decreased by approximately 20% (Talamantes et al., 2019). Teherani et al. (2018) point out that increasing the number of underrepresented minority medical students does not translate to an equitable increase of their presence in all residency programs, specifically not an increase in more competitive specialties. Examples of more competitive specialties per the Association of American Medical Colleges (AAMC) include emergency medicine and dermatology, while examples of less competitive specialties include family medicine and pediatrics (Mitsouras et al., 2019).

The Match

Currently, there are a handful of routes that a medical student or physician may use to enter a specialty training program in graduate medical education (i.e., residency or fellowship program) including the NRMP's Main Residency Match, alternative smaller but similar matches (e.g., San Francisco Match), accepting a position that was in a match but went unfilled, accepting a position that previously had a trainee but is now vacant (e.g., dismissal from a program, death, etc…), or accepting a pre-match position (Dharssi et al., 2020; The National Resident Matching Program, 2018; Wetz et al., 2010). The most common though is the Main Residency Match, also known as The Match, which is the primary route into an accredited training position in over 80 specialties or subspecialties (Almarzooq et al., 2021; The National Resident Matching Program, 2018; Wetz et al., 2010). Considering that at least some residency training is required to obtain a medical license, which is required to practice as a physician in the United States, The Match is an extremely important process in the country's medical education and healthcare structure (Freeman, 2016).

Development of The Match started in 1952 after years of worsening issues surrounding medical students obtaining internships (Roth, 2003). In the early 20th century, hospitals would

compete to hire medical students to fill internship positions earlier than competitors. Medical students were faced with a choice of accepting an internship position at a hospital they did not prefer or turning down the offer in hopes of a better one that may never come. As the situation grew more problematic, multiple stakeholders introduced and agreed to a centralized clearinghouse model that became The Match. An algorithm was developed, that has over the years had some minor modifications, to ensure best possible outcomes for both hospitals and medical students (Roth, 2003; Roth & Peranson, 1997). The complexity and importance of the algorithm are likely best conveyed by noting that results of such an algorithm and associated work resulted in a Nobel Prize in Economics (Leopold, 2021).

Generally speaking, individuals first register in early fall with both the NRMP and the AAMC's ERAS (Association of American Medical Colleges, n.d.-a; Schmitt et al., 2019; The National Resident Matching Program, n.d.-b). Registration with the NRMP requires entering an agreement that outlines terms, conditions, and expectations ranging from ethical behavior to required disclosures (The National Resident Matching Program, n.d.-a). The process involves going through the NRMP in a multi-step system that involves demographic information, various identification methods, certain exam scores (i.e., United States Medical Licensing Examination, Educational Commission for Foreign Medical Graduates, or National Board of Osteopathic Medical Examiners), work experience, volunteer experience, contact information, an informed consent for research purposes, and a registration fee (The National Resident Matching Program, n.d.-b).

Registration to ERAS requires a token, which individuals generally obtain from their dean's office (The National Resident Matching Program, n.d.-b). Following registration to ERAS, a common electronic application must be completed in ERAS which involves filling out

demographic information, a personal statement, a curriculum vitae, and a photo (Association of American Medical Colleges, n.d.-b; Schmitt et al., 2019). In addition to filling out the application, the individual must also request relevant transcripts and await additional supporting files such as the Medical School Performance Evaluation and letters of recommendation, or in some specialties a standardized letter of recommendation or evaluation (Association of American Medical Colleges, n.d.-b; Kaffenberger et al., 2016; Keim et al., 1999; Schmitt et al., 2019).

Medical School Performance Evaluation

The Medical School Performance Evaluation, formerly known as the dean's letter, is a document provided by a medical school intended to convey an objective summary of a student's attributes, experiences, and accomplishments based on verified information and summative evaluations (Association of American Medical Colleges, 2017). In addition, the evaluation is intended to contain a comparative assessment to the rest of the individual's medical school class. Over the past few decades, the AAMC (2017) has made a number of changes while providing guidelines and instructions. Unfortunately, medical schools may have a conflict of interest as they aim to have their students match into the best programs possible (Naidich et al., 2014). Whether for that reason or others, there is low compliance to including transparent comparative assessment of students in Medical School Performance Evaluations (Boysen-Osborn et al., 2017; Naidich et al., 2014). Even when comparative rankings are included of students, there is extensive variation in the systems and reporting, leading to calls for further standardization and anchoring (Andolsek, 2016; Catalanotti et al., 2017; Osborn et al., 2016). It is difficult to determine the utility of the Medical School Performance Evaluation considering the varying results of studies on its ability to correlate to, or predict, future performance (Harfmann & Zirwas, 2011; Kenny et al., 2013; Swide et al., 2009). As it stands, approximately 2/3 did not

find it significantly influential, and less than half trust the information contained within it (Bird et al., 2021). Despite all these problems, a recent study done by The National Resident Matching Program (2020) showed that 76% of respondent programs used it to some degree in determining which applicants to interview, though only 58% used it after interviews to assist in determining rank lists.

Letters of Recommendation

As compared to the Medical School Performance Evaluation, letters of recommendation were involved in determining interviewees by 84% of respondent programs, with 70% using them in rank list determination (The National Resident Matching Program, 2020). Letters of recommendation vary broadly as they may come in narrative or standardized format from an individual or group of authors (Schmitt et al., 2019). Traditional letters of recommendation have been found to have poor evidence of validity and have slight to no predictive utility (Harfmann & Zirwas, 2011; Kenny et al., 2013; Lee et al., 2008). Standardized letters of recommendation have been found to have higher interrater reliability, though have been found to have issues surrounding differentiating candidates and grade inflation (Girzadas et al., 1998, 2004; Kominsky et al., 2016; Love et al., 2014). There are also standardized formats that include narrative portions, such as in emergency medicine or otolaryngology (Jackson et al., 2019; Kominsky et al., 2016). Emergency Medicine has continued to refine their standardized approach with the most recent iteration, called the called the electronic Standardized Letter of Evaluation, having improved on prior issues of rank designation (Jackson et al., 2019).

Standardized Tests

United States medical students are required to take multiple high stakes examinations during medical school that are used in the application process to residency programs (The

National Residency Matching Program, 2020). Allopathic medical students take and report the

United States Medical Licensing Examination step 1 and step 2 clinical knowledge that each

report a numeric score, along with the clinical skills portion of step 2 that is pass or fail. Some

validity evidence has been published in the literature regarding the United States Medical

Licensing Examinations, such as association with later practice patterns and cut offs for passing

(Cuddy et al., 2004; Margolis et al., 2010). Osteopathic medical students instead take a similar

series of exams entitled Comprehensive Osteopathic Medical Licensing Examination of the

United States, which also have validity literature published (O'Neill et al., 2016; Sandella et al.,

2016). The National Residency Matching Program (2020) documented that programs vary in use

of such standardized test results ranging from a low of 40% of programs using the result of an

osteopathic student passing their level two performance evaluation to a high of 90% of

programs using an allopathic student's step one score.

Post Application Phase

After completing the application process an individual may then apply to residency

programs through ERAS, which involves sending their application materials to those programs

(Association of American Medical Colleges., n.d.-a). Starting in September, residency programs

are able to use ERAS to review applications and offer interviews, which generally occur from

November through January (Association of American Medical Colleges, n.d.-c; Schmitt et al.,

2019). Interviews traditionally occurred in person, with applicants traveling to the location of the

program to interview (Blackshaw et al., 2017). In light of the coronavirus disease (COVID-19)

pandemic, the Council of Residency Directors in Emergency Medicine supported the transition

to an entirely virtual interview process recently (Pelletier-Bui et al., 2020). In addition to in-

person or virtual interviews, some individuals may do second looks, which involves spending

additional time in the emergency department at a program of interest (Klammer et al., 2019).

Following interviews, candidates and programs must upload a rank order list of preferences into

the NRMP in February, with results released in mid-March (Schmitt et al., 2019). The agreement

of The Match, if an individual matches into a program, lasts until 45 days into the training

program (The National Resident Matching Program, n.d.-a).

Gender Biases

Application processes have been fraught with gender bias for years, with literature

suggesting The Match's application process has also long suffered from it (Durham et al., 2018;

Fay & Williams, 1993; Hewett et al., 2016; Leonard & Jiang, 1999). Fay and Williams' (1993)

appear to demonstrate, at face value, processes are biased due to inequality regimes, with

suggestions of advanced discrimination occurring due to the social construct of gender.

These biases can be seen within academia and specifically within medicine. Madera et al.

(2009) found hiring processes in academia to be biased, finding females to be described as

communal more often than male counterparts and that having communal characteristics had a

negative association with being hired. Within medicine, authors have identified issues with

aspects of the application process such as language differences in letters of recommendation

written, language differences in evaluations, rating differences in evaluations, and standardized

exam scores being under-predictive in females (Axelson et al., 2010; Girzadas et al., 2004;

Leonard & Jiang, 1999; Turrentine et al., 2019).

Applicant's Gender

Gender of the applicant appears to be one source of gender-based bias within application

materials to The Match (Aggarwal et al., 2018; Babal et al., 2019; Ross et al., 2017). Gender-

based differences in language show up in multiple places including personal statements, Medical

Student Performance Evaluations, letters of recommendation, and standardized letters of evaluation (Aggarwal et al., 2018; Babal et al., 2019; Li et al., 2017; Ross et al., 2017). Examples include using communal adjectives (e.g., sympathetic, caring, kind) more often for female candidates, which Grimm et al. (2020) noted to be perceived as incompatible with traditional professional characteristics of agency (e.g., competent, independent, aggressive) (Aggarwal et al., 2018; Lin et al., 2019; Turrentine et al., 2019). While standardized letters of recommendation were found to reduce such gender descriptor disparities, they were still guilty of it to a degree (Friedman et al., 2017; Li et al., 2017). These differences in descriptors may be why prior work has shown that cues may be used with anonymized applications to infer gender identity (Foley & Williamson, 2018). These gender-based differences appear to be socially constructed stereotypes considering the linguistic differences that highlight Frug's (1992) postmodern feminist theory's principles.

Babal et al. (2019) point out another potential hidden issue that may also revolve around the adjectives used to describe females in letters of recommendation and Medical Student Performance Evaluations. Prior literature has demonstrated that females are more likely to obtain unfavorable evaluations if they behave in a manner that is incongruent with the typical role and perception of a female gendered individual (Eagly & Karau, 2002; Heilman, 2001). With that consideration, females who were described in a communal fashion may have not demonstrated those characteristics and in turn received lower evaluations due to the incongruity between that behavior and the perception of a typical female gender role.

Many groups studying gender biases based on the applicant's gender in the residency application process have used LIWC and evaluated the presence of various categories of word usage in application materials among male and female applicants (Filippou et al., 2019; Grimm

et al., 2020; Lin et al., 2019). Filippou et al. (2019) found materials for male applicants more often referenced words within the domains of power, work, and personal drive, finding an association between the usage of power words and candidates able to match into urology. In ophthalmology, Aggarwal et al. (2018) also identified differences in the use of words based on gender, and also found an association of specific words with ability to match into ophthalmology. While each author group identified an association of word usage to matching into a specific specialty, it should be noted that not matching into a specific specialty is not the same as not matching into in individual's first choice or not matching at all, and so while the association with an outcome is important, outcomes focused research on these situations are difficult to interpret (Aggarwal et al., 2018; Filippou et al., 2019).

Other authors have instead limited their focus to investigating word usage in narrative portions of residency application materials, which is important considering the importance of language within the framework of postmodern feminism (Chapman et al., 2020; Grimm et al., 2020; Li et al., 2017). Some groups have used numerous domains of words, such as Turrentine et al. (2019) who studied over 30 different categories, or Miller et al. (2019) who evaluated 16 categories. Other authors have appeared to focus on agency and communality (Chapman et al., 2020; Grimm et al., 2020; Li et al., 2017). Regardless of number of domains of words used, many authors have identified differences among male and female applicants (Aggarwal et al., 2018; Chapman et al., 2020; French et al., 2019; Grimm et al., 2020).

The results are mixed when focusing on words within the communal dictionary among male and female applicant materials. Ross et al. (2017) studied specific terms rather than the word dictionaries as whole domains, but found females more often described using communal terminology (Pietraszkiewicz et al., 2019). French et al. (2019) found significantly more

communal terms in letters of recommendation for female applicants, though could account for it after controlling for word length as a covariate. Powers et al. (2020) and Turrentine et al. (2019) found nonsignificant trends towards increased use of communal language in letters of recommendation for female applicants, but did not control for length of letters, with letter lengths found to be significantly different based on gender by each group. Hoffman et al. (2019) did not find a significant difference in use of communal words as an overall category, but identified specific differences within terms that were gender biased.

As noted, not all authors have found differences or trends in the use of communal terms among materials for male and female applicants. Babal et al. (2019), Chapman et al. (2020), French et al. (2019), and Grimm et al. (2020) found there to be no significant difference in communal language among males and females in their studies focusing on letters of recommendation in residency application materials in pediatrics, radiation oncology, general surgery, or diagnostic radiology. Polanco-Santana et al. (2021) focused on the Medical Student Performance Evaluation but found no significant difference in communal terminology among male and female students going into general surgery. In emergency medicine, Li et al. (2017) did not find a significant difference in frequency of communal language use in narratives of Standardized Letters of Evaluation, however they did not control for word count variations. Miller et al. (2019) did not find a significant difference in communal language among male and female applicant's Standardized Letter of Evaluation narratives in emergency medicine after adjusting for word count, however focused on the Standardized Letter of Evaluation instead of the updated eSLOE.

There is also potential bias in The Match process outside of just application materials. Multiple studies have demonstrated individuals being asked potentially illegal questions about

their gender during the interview process for residency programs (Hern et al., 2013, 2016). Further, female interviewees were more likely to have been asked potentially illegal questions than their male counterparts (Berriochoa et al., 2018; Hern et al., 2013, 2016). Berriochoa et al. (2018) not only identified such a bias with females being more likely to be asked about having children, but also demonstrated an increase in distress due to such questions. In addition to females being more likely to receive potentially illegal questions, females experienced unprofessional behavior or inappropriate conduct at approximately 2.4 times the rate of male counterparts during interviews (Lee et al., 2019). Examples of this unprofessionalism included perceived demeaning behavior, having language or behavior that was considered brash, aggressive, or offensive, and asking specifics about relationships and family status.

Gender of Other Stakeholders

While an applicant's gender may create a gender bias in an application, the gender of the author of a letter of recommendation and the gender of the individual(s) responsible for hiring also create a gender bias (Friedman et al., 2017; Gorman, 2005). Such biases in other stakeholders demonstrate potential importance of inequality regimes and of homosocial reproduction as frameworks. One example of differences identified are that letters of recommendation authored by a female have been found to be more likely to be missing at least one key attribute for an applicant (Friedman et al., 2017). Other examples include finding differences among letters authored by females and males in conveying specific gendered applicants' potential to match and the authenticity of letters of recommendation (Filippou et al., 2019; Friedman et al., 2017; Girzadas et al., 2004). With consideration of homosociality and inequality regimes, male authors have been found to write significantly longer letters of

recommendation for male applicants, while female authors have been found to give female applicants significantly greater chances of successfully matching (Girzadas et al., 2004; Turrentine et al., 2019).

Specific language used is also important, considering the postmodern feminist approach, and the meanings conveyed such as with communality. Grimm et al. (2020) found female authors to use agentic and communal language more frequently than their male counterparts. Chapman et al. (2020) found similar findings on female authors using significantly more agentic language, but did not find significance in the increased use of communal language by female authors. In contradiction to Grimm et al. and Chapman et al., Babal et al. (2019) found male authors to use agentic language significantly more frequently, and no difference in use of communal language, when evaluating by author of letters of recommendation to pediatric residency training. In emergency medicine, Li et al. (2017) and Miller et al. (2019) intended to evaluate potential gender bias due to author of materials, but each focused only on Standardized Letters of Evaluation, and each found most of them to be composed by group authors, being unable to evaluate for specific relationships or biases.

Aside from writing letters of recommendation, the gender of the application reviewer, and the faculty present in a training program, may also play a role. Gorman (2005) found that when a decision maker was female, they were more likely to hire a female applicant. Chapman et al. (2019) demonstrated that the percentage of female faculty within a graduate medical education program was associated with inequality of gender representation. Each of these findings demonstrate potential underlying homosociality and inequality regimes. Finding such an effect of gender concordance between applicant and decision maker is similar to the concordance or gender diad of applicant and author that was found by others (Filippou et al., 2019; Friedman

et al., 2017; Girzadas et al., 2004; Turrentine et al., 2019).

      Unconscious bias has also been demonstrated to impact graduate medical education training programs and can be found in students, residents, faculty, and decision-making committees (Backhus et al., 2019; Capers Q, 2020; Hanson et al., 2019). Such unconscious bias may be, at least in part, responsible for gender biases seen within graduate medical education training programs. Backhus et al., (2019) identified unconscious bias within faculty of graduate medical education programs, specifically within cardiothoracic surgery. Unconscious bias has also been found within emergency medicine and obstetrics and gynecology training programs (Hanson et al. (2019). Belanger (2018) notes unconscious bias plays a large role in homosocial behavior.

Ethnic Biases

      While race and ethnicity are often used interchangeably, it is important to consider the nature of each for the purposes of this study. Santos et al. (2010) consider race an "unscientific term" due to the basis in genetics and variants seen, noting race usually refers to phenotypic differences (e.g., skin color). They note ethnicity refers to an individual's identity built out of both visual and non-visual (e.g., religion, nationality, language) features. From their standpoint, it appears that race and ethnicity are both social constructs, though race to a larger degree. Despite the differences, since both race and ethnicity contain a foundation as a social construct and both are discussed relatively interchangeably in the published literature of the topic's focus area, they will continue to be used in conjunction for this study (Boatright et al., 2018).

      Many disparities exist in higher education surrounding an individual's race or ethnicity and have been recently highlighted by the American Council on Education's Status Report (Espinosa et al., 2019). Studies have also demonstrated that approximately 10% of individuals

interviewed for a residency position were asked potentially illegal questions surrounding their ethnicity (Hern et al., 2013, 2016). Training in graduate medical education appears to be no different, with multiple reasons identified as potential sources of bias that place non-white applicants at a disadvantage (Aggarwal et al., 2018; Chapman et al., 2020). Primary issues identified include language used in letters of recommendation and performance on standardized tests. Such issues have at least begun to be recognized within graduate medical education, and specifically within emergency medicine, with Boatright et al. (2018) having recently published on the matter including recommendations and suggestions.

Application Narratives

As with gender, language in letters of recommendation is a potential source of ethnic bias (Grimm et al., 2020). Similar to the comparison of females to males, Grimm et al. (2020) found that Latinx and black applicants were described as less agentic when compared to Asian or white applicants. Other authors have found descriptive differences as well, such as letters for white applicants including more standout words, while letters for an individual from an underrepresented race and ethnicity contained more grindstone descriptors (Powers et al., 2020). The use of standardized letters of recommendation over traditional letters of recommendation appears to improve the disparity of language used between underrepresented and non-underrepresented race and ethnicity candidates (Chapman et al., 2020; Powers et al., 2020).

Aside from letters of recommendation, Medical Student Performance Evaluations and personal statements have also had their narrative portions evaluated based on applicant ethnicity. Ross et al. (2017) found significant differences in categories of words used among different ethnicities when evaluating for linguistic differences in Medical Student Performance Evaluations. Demzik et al. (2021) found differences in personal statements based on the ethnicity

of the applicant.

Linguistic differences in application materials based on an applicant's ethnicity are important to consider within a postmodern feminist framework and inequality regimes. Unfortunately, there do not appear to be studies that evaluate narratives of applications to residency training as complete products (i.e., all application narratives as compared to just portions of the application). There also does not appear to be literature on potential homosocial relationships, as no published studies were found evaluating interactions of concordance or discordance among the ethnicity of applicants and authors.

Academic Performance

While linguistic differences have been found even when controlling for academic performance, it is important to note disparities in academic performance (Ross et al., 2017). Individuals from an underrepresented minority score significantly lower on standardized exams such as the United States Medical Licensing Examination, which has been shown to be associated with ability to match (Aggarwal et al., 2018). Woolf et al. (2011) found a more generalized academic underperformance in those from an ethnically underrepresented background. Multiple authors have also commented on reducing attention to scores in applications as a method to reduce racial and ethnic bias (Edmond et al., 2001; Spector & Railey, 2019).

Other Sources of Ethnic Bias

Similar to gender bias, unconscious bias also impacts ethnic bias within graduate medical education (Capers, 2020). Shah and Ahluwalia (2019) note the difficulty in truly identifying the impact of unconscious bias as it relates to ethnic biases in achievement within graduate medical education due to a general reluctance to discuss the topic. Despite that, implicit bias associated

with ethnicity has been found within emergency medicine (Zeidan et al., 2019).

No studies were found that evaluated the potential importance of the race or ethnicity of a letter of recommendation, or the potential for a racial concordance between an author and applicant. Despite literature available having been focused on the race or ethnicity of an applicant, other existing literature highlights the impact of race and ethnicity aside from the applicant. Kassam et al. (2021) found that approximately 80% of certain leadership were white, Chapman et al. (2020) noted only 3.3% of their workforce were black, and Madsen et al. (2017) identified disparities specifically in emergency medicine leadership positions, demonstrating the potential homosocial reproduction discussed by Dressel et al. (1994).

Medical Education Background Bias

An individual may complete their medical training in a variety of manner such as within or outside of the United States, and as allopathic (i.e., leading to an M.D.) or osteopathic (i.e., leading to a D.O.). While those training routes all lead to ability to apply for a training program in graduate medical education through The Match, some biases have been found based upon that background of training including likelihood of matching into certain specialties or the ability to match at all (Craig et al., 2021; Nagarajan et al., 2020). Craig et al. (2021) highlighted the disparity among osteopathic and allopathic graduates into certain specialties, such as dermatology. A similar disparity was also noted by Jenkins et al. (2019), along with finding a difference in programs that were filled with allopathic graduates from schools within the United States compared to those from outside the United States. Such disparities have been noted by other authors as well, including some claiming outright discrimination (Desbiens & Vidaillet, 2010; Jolly et al., 2011). These findings of programs that have clusters based on training background again highlight the importance of homophily as noted by Clifton et al. (2019) and Fu

et al. (2012).

Other Biases

Bias surrounding an individual's gender, along with race and ethnicity, appear to be the most predominant in the literature, though other sources must still be considered. Other potential sources of bias include the presence of a disability, an individual's age, sexual orientation or appearance (de Hughes et al., 2020; De Oliveira et al., 2012; Maxfield et al., 2019; Meeks et al., 2019). It is important to consider those other potential sources of bias as they may also be pervasive and impact numerous training programs. For instance, Sapp et al. (2021) found 26% of emergency medicine residency programs have at least one resident with a disability and 13.5% of emergency medicine residency programs have at least one faculty with a disability. Consideration of all biases is important as despite being illegal or unethical, multiple studies have demonstrated a large number of applicants being asked about topics such as religion, sexual orientation, and ethnicity (Hern et al., 2013, 2016; Hughes et al., 2020; Santen et al., 2010).

Disability Bias

Meeks et al. (2018, 2019) have noted the systemic barriers in graduate medical education, similar in nature to the inequality regimes described by Acker (2006), to individuals with disability. Such barriers may be why Schwarz and Zetkulic (2019) have discussed the notion of belonging, and individuals with disabilities misgivings about belonging in medical education settings. Pheister et al. (2020) found that disclosing depression, a mental illness per the Equal Employment Opportunity Commission, left residency applicants at a disadvantage (Kiuhara & Huefner, 2008). Some of this may be a result of unconscious bias of this minority group (Marcelin et al., 2019).

Age Bias

Another factor that appears to serve as a potential bias is the age of an applicant. Lakoff et al. (2020) found that being younger than 26 placed an individual at a disadvantage when matching into a residency program. Interestingly, De Oliveira et al. (2012) compared individuals equal to or under 28 and those 29 or older and found being 29 or older placed individuals at a disadvantage. It is difficult to tell if these are conflicting results, or if there is an ideal range in the mid-late 20s for an individual to apply to a residency program. Some of the age bias found may be related to unconscious bias, as can be seen with gender and ethnic biases (Backhus et al., 2019). Within emergency medicine, a recent study demonstrated that despite not being allowed to ask about an applicant's age, it still occurred approximately 8% of the time (Hughes et al., 2020; The National Resident Matching Program, n.d.-c).

Short-Term Impact

While identifying and working to improve biases and disparities may be a moral imperative, there are also more tangible implications in the short and long-term to consider. From a short-term perspective, these biases may lead to a positive feedback loop of ongoing disparity, which may also threaten accreditation of training programs, and may represent a significant commitment of effort and time to address (Craig et al., 2021; Dressel et al., 1994; Kleinbaum et al., 2013; McDade, 2019; Nagarajan et al., 2020; Spottswood et al., 2019).

Diversity of a program may be negatively affected by the biases discussed previously for multiple reasons, which may in turn lead to accreditation issues for a training program. Multiple studies have found a training program's diversity to impact decision-making of a medical student's rank list (Fairmont et al., 2020; Love et al., 2012; Phitayakorn et al., 2015). Fairmont et al. (2020) note that female and non-white applicants consider diversity of a program they are ranking more than their male and white counterparts though. Ruzycki et al. (2020) found an

association between gender and likelihood of matching into a first specialty. Those factors, along with issues discussed earlier, may lead to decreased diversity in a training program. As discussed by McDade (2019) the Accreditation Council for Graduate Medical Education (2020) has recently implemented new common program requirements that require training programs and sponsoring institutions to address diversity and inclusion.

Constant attempts to increase diversity of applicants without addressing the underlying bias and disparities may also take a significant amount of resources from training programs. Spottswood et al. (2019) published on a multi-pronged approach to address diversity at a single program requiring faculty time commitments and financial resources of the department. Garrick et al. (2019) noted it was not just resources at the department level for increasing diversity, but instead pulled on resources from the institution at large, including time commitments from chief officers.

Long-Term Impact

Such disparities in The Match and residency programs may also have a long-term impact as well. The primary long-term issue is that if there are less female and underrepresented minorities that enter training programs, then there will be less female and underrepresented minorities practicing as physicians. Such a disparity can lead to worse outcomes for those populations ranging from birthing mortality to antibiotic prescriptions or mortality following a heart attack (Eggermont et al., 2018; Greenwood et al., 2018, 2020). Shen et al. (2018) noted that racial discordance between a patient and provider was associated with worse communication between the two, while Takeshita et al. (2020) noted that there were improved patient satisfaction scores when there was racial/ethnic concordance between a physician and patient.

Summary

The Match is a multifaceted process that serves as the primary mechanism for medical students and physicians to obtain a training spot in a graduate medical education training program (i.e., residency or fellowship program) in the United States (The National Resident Matching Program, 2018; Wetz et al., 2010). Considering the importance of The Match, it is concerning that there are suggestions the application process may be biased with regards to gender, race or ethnicity, disabilities, age, and educational background (Craig et al., 2021; De Oliveira et al., 2012; Meeks et al., 2018; Ross et al., 2017). Such biases within the match could lead to negative short and long-term consequences such as difficulty for minorities or females to obtain a position within a residency program, accreditation issues for residency programs, and poorer health outcomes for minorities or females in the United States (Greenwood et al., 2018, 2020; Madsen et al., 2017; McDade, 2019). Should the application process of The Match be biased, it would be advanced discrimination, and would lead to homosocial reproduction and serving as an inequality regime (Acker, 2006; Dressel et al., 1994). There is a current gap in the literature regarding potential biases and variance attributable within the application process of The Match to emergency medicine residency programs in the United States.

Some specialties have begun examining the application process for biases, though the limited evidence available has yet to identify variance attributable or what changes may be occurring with time (Aggarwal et al., 2018; Grimm et al., 2020). Recognizing these biases as social constructs, along with the importance of language used in documents such as letters of recommendation and the Medical Student Performance Evaluation, points towards the use of postmodern feminist theory and intersectionality in addition to inequality regimes and homosocial reproduction. By identifying biases within emergency medicine applications in The

Match, and their associated variance and change over time, stakeholders may begin to identify

and implement measures to combat them.

CHAPTER THREE: METHODS

Overview

The purpose of this quantitative, causal-comparative study is to examine if certain variables (i.e., self-reported gender and self-reported ethnicity) affect a medical student's rating or amount of communal language used in their application materials to an emergency medicine residency program through ERAS. This chapter details the design of the research to determine if there was a difference found using the Electronic Standardized Letter of Evaluation or LIWC. The chapter provides details on the research questions, hypotheses, participants, setting, instruments, procedures, and data analysis.

Design

This section contains a descriptive summary of causal-comparative research designs, and identifies the appropriateness of using a causal-comparative design for the proposed study. The section concludes with a discussion of data collection and descriptions of variables used.

This will be a quantitative, causal-comparative design, which is an approach that seeks to identify the degree of differences between two or more groups without manipulation of the groups and without determining causality with certainty (Gall et al., 2007; Schenker & Rumrill, 2004). These designs primarily involve two types of variables, the independent and dependent variables (Schenker & Rumrill, 2004). This is an appropriate design for the purpose of my study, which is to investigate the difference in medical students' application materials to an emergency medicine residency based on gender and self-identified ethnicities. As a causal-comparative design, it uses independent categorical variables that are not manipulated and dependent variables that are continuous in nature. A limitation of the design is that cause and effect relationships are not able to be established.

The first independent variable for research question one is gender, which is defined as the socially constructed differences between men and women (Risberg et al., 2009; Verdonk et al., 2009). The second independent variable for research question one is self-identified ethnicity, which is defined as the scientific term referring to the socially constructed multi-faceted identify of an individual (American Psychological Association, 2019; Santos et al., 2010). The dependent variable for research question one is the eSLOE aggregate rating, which is the simple average of ratings on the eSLOE for an individual (Hopson et al., 2019).

The independent variable for research question two is gender, which is the socially constructed differences found between men and women (Risberg et al., 2009; Verdonk et al., 2009). The dependent variable for research question two is communal language and is defined as relationship-oriented language such as kind, sympathetic, reliable, and friendly (Grimm et al., 2020; Pietraszkiewicz et al., 2019). The covariate for research question two is word count which is defined as the number of words in a certain text (Turrentine et al., 2019).

Research Questions

RQ1: Is there a difference among the electronic Standardized Letter of Evaluation aggregate ratings of male and female medical students applying to an emergency medicine residency based on their self-identified ethnicity?

RQ2: Is there a difference in communal language used in narrative portions of applications to an emergency medicine residency found by Linguistic Inquiry and Word Count 2015 among male and female medical students when controlling for total word count?

Hypotheses

H_01: There is no significant difference among electronic Standardized Letter of Evaluation aggregate ratings of medical students between male and female medical students.

H$_0$2: There is no significant difference between electronic Standardized Letter of Evaluation aggregate ratings among applicants based on their self-identified ethnicity.

H$_0$3: There is no significant interaction among the electronic Standardized Letter of Evaluation aggregate ratings of medical students among male and female medical students based on their self-identified ethnicity.

H$_0$4: There is no difference in communal language used in narrative portions of emergency medicine residency applications, as measured by LIWC, among male and female medical students when controlling for total word count.

Participants and Setting

Within this section a description of the population of interest and participants is provided, followed by discussion of the sampling technique and sample size. The section closes with a description of the setting of the study.

Population

The participants for the study were drawn from a convenience sample of medical students that applied to emergency medicine residency programs during the 2022-2023 application cycle through ERAS. The population of applicants is detailed and updated annually by the Association of American Medical Colleges (2021a). During the 2022-2023 cycle, there were 2,846 male, 1,936 female, and 5 other / unknown reporting individuals (see Table 1) (Association of American Medical Colleges, 2023). There were 47 American Indian or Alaska Native, 893 Asian, 373 Black or African American, 510 Hispanic, Latino, or of Spanish Origin, 13 Native Hawaiian or Other Pacific Islander, 2,811 White, 241 Other Race/Ethnicity, and 155 that had unreported or unknown ethnicity (see Table 2). There were 2,256 United States allopathic medical college applicants, 1,218 United States osteopathic medical college applicants, and

1,313 international medical college applicants.

Table 1

Gender Frequency of Sample and of The Match during 2022-2023

Gender	Sample (n = 173)	Population (n = 4787)
Male	95 (54.9%)	2846 (59.5%)
Female	78 (45.1%)	1936 (40.4%)
Other / Not identified	0 (0%)	5 (0.1%)

Table 2

Ethnicity Frequencies of Sample Population Provided and Total Population

Ethnicity	Sample Primary Ethnicity (n = 173)	Sample Secondary Ethnicity (n = 173)	Population Ethnicity (n = 5043)
American Indian or Alaska Native	2 (1.2%)	0 (0%)	47 (0.9%)
Asian	56 (32.4%)	0 (0%)	893 (17.7%)
Black or African American	6 (3.5%)	1 (0.6%)	373 (7.4%)
Hispanic, Latino, or of Spanish Origin	21 (12.1%)	1 (0.6%)	510 (10.1%)
Native Hawaiian or Other Pacific Islander	0 (0%)	0 (0%)	13 (0.3%)
White	69 (39.9%)	22 (12.7%)	2811 (55.7%)
Other Race/Ethnicity	15 (8.7%)	1 (0.6%)	241 (4.8%)
Unreported	4 (2.3%)	148 (85.5%)	155 (30.7%)

Participants

For this study, the number of participants sampled was 173 which exceeded the required minimum when assuming a medium effect size. Convenience sampling was used by using a single participating residency site. According to Gall et al. (2007), 126 students is the required minimum for a two-way ANOVA assuming a medium effect size with statistical power of .7 at

the .05 alpha level. The minimum required sample is 166 for an analysis of covariance

(ANCOVA) with a power of .07 at the .05 alpha level.

The naturally occurring sample consisted of 95 males and 78 females (see Table 1) aged

25-49, with 4 individuals not providing an age. According to the primary self-reported ethnicity,

there were 2 American Indian or Alaska Native, 56 Asian, 6 Black or African American, 21

Hispanic, Latino, or of Spanish Origin, 0 Native Hawaiian or Other Pacific Islander, 69 White,

15 Other Race/Ethnicity, and 4 that did not report a primary ethnicity (see Table 2). There were

25 individuals that included a secondary self-reported ethnicity, this included 1 Black or African

American, 1 Hispanic, Latino, or of Spanish Origin, 22 White, and 1 Other Race/Ethnicity (see

Table 2). There were 67 United States allopathic medical college applicants, 75 United States

osteopathic medical college applicants, 26 Caribbean medical college applicants, and 5

international non-Caribbean medical college applicants. The sample population had 442

eSLOEs and 169 non eSLOE letters of recommendation with 237 of the eSLOEs being from

group authors and 205 of the eSLOEs being from individual authors. Of eSLOEs from group

authors, 36 groups were comprised only male authors and 15 groups were only female authors,

with 186 being groups of mixed gender. Of eSLOEs from individual authors, 158 were from

male authors and 47 were from female authors.

Setting

The institution is an urban, community-teaching hospital located in a Western U.S. state,

and is part of a larger non-profit system. It consists of multiple clinical sites that residents rotate

through. The institution hosts multiple residency programs including a three-year emergency

medicine residency that is accredited for 36 residents.

Instrumentation

This study is evaluating material used in the application process of the ERAS, which has approval granted through the Association of American Medical Colleges (2021a) in advance as it is for research purposes. This study will specifically evaluate the dependent variable aggregate rating using the instrument developed by the Council of Residency Directors in Emergency Medicine (2016), the eSLOE. See Appendix A for instrument. The study will use LIWC to determine the dependent variable of communal language (Pennebaker et al., 2015).

Electronic Standardized Letter of Evaluation

The purpose of this instrument is to evaluate the dependent variable of aggregate rating of an applicant to an emergency medicine residency program. The eSLOE is an appropriate instrument to use as it is considered "the gold standard used by emergency medicine (EM) program directors during the residency application process" (Jackson et al., 2019, p. 182). It is an instrument that was developed by the Council of Residency Directors in Emergency Medicine (2016) to serve as an assessment tool for applicants and is the product of multiple iterations of refinement that started in the mid 1990s (Krzyzaniak & Lin, 2022; Love et al., 2014, 2020; Martin & McNamara, 2014). Studies have demonstrated improved evidence of validity and an interrater reliability of 0.97 for Kendall's coefficient of concordance (Girzadas et al., 1998; Jackson et al., 2019; Kukulski & Ahn, 2021; Love et al., 2020). There are multiple areas of validity evidence to consider such as content, response process, internal structure, relations to other variables, and consequences (Cook & Beckman, 2006). Kukulski and Ahn (2021) provided a scoping review of literature on validity evidence for the emergency medicine standardized letter of evaluation, demonstrating studies in each of the categories of validity evidence. While the amount of literature available to support each aspect of validity evidence on the instrument varied, there was supporting evidence of validity for each category of validity evidence.

The instrument consists of multiple sections including demographic data, background information, evaluation of student, written comments, and institutional information. The evaluation of the student includes three sections part A, part B, and part C, with the written comments also being information about the student in a narrative format (Krzyzaniak & Lin, 2022; Love et al., 2020). Demographic data consists of items regarding the applicant and author(s) such as contact information, institution, and sources of information used to compile the eSLOE (Krzyzaniak & Lin, 2022). Background information includes information on when the grade breakdown, when the applicant rotated, and if an exam was part of the rotation. Evaluation of student part A, B, and C contain a combined 14 questions with Likert Scale responses related to the evaluation of the applicant for emergency medicine that may provide an aggregate rating of a Likert Score from six to 59 (Boone & Boone, 2012; Dowling & Midgley, 1991; Love et al., 2020). The five responses of Evaluation of Student Part A are scored on a three-point Likert Scale: Fully entrustable = 3, Mostly entrustable = 2, Pre-entrustable = 1, and unable to assess = 0 (Dowling & Midgley, 1991;). The seven responses of Evaluation of Student Part B are scored on a five-point Likert Scale with anchors at Exceptional EM candidate = 5, Minimally acceptable for an EM resident = 1, and Not acceptable for EM resident = 0. Part C of Evaluation of Student includes one question scored on a four-point Likert Scale: Minimal: Will excel with just a little guidance and support = 4, Standard: No problems expected, will succeed with standard guidance and support = 3, Moderate: May need slightly more than the standard support from time to time, no major issues anticipated = 2, Most: Has the potential to succeed, but will likely require extra support throughout residency = 1. Part C of Evaluation of Student includes one question scored on a five-point Likert Scale: Top 10% = 5, Top 1/3 = 4, Middle 1/3 = 3, Lower 1/3 = 2, Unlikely to be on our rank list = 1. The combined possible score on the

eSLOE ranges from six to 59 points. A score of 59 points is the highest possible score, while a score of six points is the lowest possible score, which Martin and McNamara (2014) note mean that a residency program director wants to meet, or avoid meeting, the individual respectively.

Linguistic Inquiry and Word Count 2015

The purpose of this instrument is to perform quantitative text analysis of aggregated narrative portions of an application to an emergency medicine residency program. This will provide a continuous score for communal language use based on text found within narrative portions of an application. Specifically, this will be done with the communal language dictionary using LIWC and result in a score with a range of 1 to 99 (Linguistic Inquiry and Word Count, n.d.; Pietraszkiewicz et al., 2019). LIWC was developed in the early 1990s with validity evidence on the software and dictionaries (Francis & Booth, 1993; Tausczik & Pennebaker, 2010). As it has iterated, additional validity evidence has been published (Dudău & Sava, 2020; Kahn et al., 2007). Recent data has demonstrated accuracy and Cohen's kappa as high as 0.95 and 0.88, respectively (Ferreira et al., 2020). The program is able to analyze text based on developed dictionaries of categories of words, and provide an output of usage (Tausczik & Pennebaker, 2010). In regard to specifically communal language, an internal consistency of 0.86 was found, with further validity evidence on convergent and discriminant validity (Pietraszkiewicz et al., 2019). LIWC been used as an instrument in numerous other studies, including in research similar to the proposed study (Babal et al., 2019; Figueiredo, 2020; Grimm et al., 2020; Li et al., 2017).

<div align="center">Procedures</div>

Institutional Review Board (IRB) approval was initially sought for this quantitative, causal-comparative study of applicants to a single institution's emergency medicine residency

program for the 2022-2023 application cycle (See Appendix B). The Association of American

Medical Colleges maintains a data policy and privacy statement to allow programs that receive

personal information and applications through ERAS for "non-commercial uses, including

research" (p. 2) (Association of American Medical Colleges, 2021a, 2021b). Data submitted is

maintained by the residency director or program, and then able to be accessed for administrative

or research purposes.

Following IRB approval, archival data was abstracted with each application given a

unique identifier code to abstain from using personally identifiable information in the datasets.

The unique identifier code key was maintained in an excel file on a password protected

computer. Data, including the variables of interest and the unique identifier code was entered

with statistics run using Statistical Package for the Social Sciences version 27 (SPSS), SigmaXL,

and LIWC which was also stored on a password protected computer. The abstracted data was

destroyed after the completion of this research study.

Data Analysis

Descriptive and inferential statistics, specifically a two-way ANOVA and an ANCOVA

was used to evaluate the research questions and hypotheses in the study. Pair-wise deletion will

be used to manage missing data. The rationale for using a two-way ANOVA is the ability to

evaluate for differences and interactions when there are two or more categorical independent

variables (MacFarland, 2011). In the proposed study, the involved categorical independent

variables are gender, specifically male or female, and ethnicity, which contains eight potential

responses. A two-way ANOVA will be used to answer RQ1 and $H_0 1 - H_0 3$. The rationale for

using an ANCOVA is that it is able to evaluate if the mean scores of groups differ after removing

or equating another variable's influence (Rutherford, 2011). An ANCOVA will be used to

answer RQ2 and $H_0$4. The independent variable for the ANCOVA is gender, specifically male or female. The dependent variable for the ANCOVA is communal language used in narrative portions of applications. The covariate for the ANCOVA is total word count.

Descriptive Statistics

Initially, visual data screening was performed to evaluate for missing and inaccurate datapoints. Then, descriptive statistics, including mean, median, and standard deviation, was run for general patterns in the data, to provide characteristics about the sample population, and begin evaluating normality of data (Marshall & Jonker, 2010). Gall et al. (2007) also highlight the use of measures of central tendency for descriptive data to provide easily and quickly interpreted findings. While the mean is one of the most common descriptive statistics, it is more sensitive to outliers than the mean (Vetter, 2017). The standard deviation provides information about the amount of variability present, in general representing distance from the mean on average for a data set. In addition to calculating values, box-and-whisker plots will be used to provide a visual representation of each variable and will allow for evaluation of extreme outliers. The box-and-whisker plot is a visual representation of data that allows for quick identification of the median, interquartile range, the fifth and 95th percentile of values or largest and smallest scores, and the extreme outliers.

Two-Way Analysis of Variance

A two-way ANOVA was used to answer RQ1 and $H_0$1 – $H_0$3. In order to perform a two-way ANOVA, certain assumptions must be met (Laerd Statistics, 2018-b). There are six assumptions that must be met, focused on the nature of variables used, observations, outliers, distribution, and variances. First, the dependent variable is to be continuous, meaning measured as interval or ratio. Second, the independent variable is categorical and independent. Third,

observations must be independent with no individual being part of more than one grouping (i.e., gender or self-identified ethnicity). Fourth, there should not be significant outliers. Fifth, the dependent variable should fall in a normal distribution. Finally, there should be homogeneity of variances.

Multiple methods and tests will be performed for evaluation of assumptions. A box-and-whisker plot can be used to evaluate of outliers (Vetter, 2017). A Shapiro-Wilk test can be performed to evaluate normality as Razali and Wah (2011) have found it to be the most powerful test of normality. Specifically, it has been found to be more powerful than a Kolmogorov-Smirnov or Anderson-Darling test for normality (Razali & Wah, 2011). Testing for homogeneity of variances may be done using Levene's test for equality of variance, which is also robust to nonnormality, and allows for evaluation of homogeneity of variances (Gastwirth et al., 2009). Significance can be identified using $\alpha = .05$ (Aberson, 2019; Maher et al., 2013). Eta-squared (η^2) is a commonly reported measure of effect size that is reported, with .01 representing a small effect size, .06 representing medium effect size, and .14 representing large effect size (Maher et al., 2013).

Analysis of Covariance

An ANCOVA was used to answer RQ2 and H_04, and requires multiple assumptions to be met. First, the dependent variable and covariate should be continuous in nature (Laerd Statistics, 2018-a). Second, the independent variable should be categorical and independent. Third, observations should be independent of each other. Fourth, there should not be significant outliers. Fifth, residuals for each category of the independent variable should fall in a normal distribution. Sixth, there should be homogeneity of variances. Seventh, there should be a linear relationship at each category of the independent variable between the covariate and dependent

variable. Eighth, there should be homoscedasticity. Finally, there should be homogeneity of the regression slopes. Again, evaluation of outliers may be done with a box-and-whisker plot (Vetter, 2017). A Shapiro-Wilk test may be used to evaluate residuals' normality (Laerd Statistics, 2018-a). Levene's test may be used for equality of variances. Assumptions seven and eight can be evaluated with grouped scatterplots. Regression lines and comparison of slopes may be evaluated through Assumption nine may evaluated through development of regression lines and comparing slopes found (Field, 2013). Deligonul (1998) notes the usefulness of visually evaluating a scatterplot for an elliptical shape for a bivariate normal distribution. Assessing the homogeneity of slopes is done to evaluate for the absence of interaction of the covariate and independent variable, which can be done through SPSS (Denis, 2018). Significance was again identified using $\alpha = .05$ (Aberson, 2019; Maher et al., 2013). Eta-squared (η^2) was again used to measure effect size (Maher et al., 2013).

Bonferroni Correction

As multiple statistical tests of significance were used, there is an increased risk of a type I error (Armstrong, 2014). One method of addressing this increased error is by performing a Bonferroni correction. This correction is specifically performed in order to address errors of repeated significance testing (Nahler et al., 2017; Ranstam, 2016). A Bonferroni correction may be used in conjunction with the two-way ANOVA and ANCOVA as part of the study. As Ranstam (2016) notes, there is an underlying assumption that tests involved are independent, which was met. Criticisms of the Bonferroni correction include that it is overly conservative and may make it more difficult when smaller sample sizes are being used (Ranstam, 2016; VanderWeele & Mathur, 2019).

CHAPTER FOUR: FINDINGS

Overview

Chapter Four presents descriptive statistics on the convenience sample of medical students applying to an emergency medicine residency during the 2022-2023 application cycle through ERAS. It reviews the data screening procedures and assumptions of a two-way ANOVA and ANCOVA analysis. Following review of assumptions, the chapter presents a statistical analysis of data. Results are presented for the two-way ANOVA using gender and ethnicity as categorical independent variables are gender. Then, results are provided for an ANCOVA using gender as an independent variable, communal language used in narrative portions of eSLOE as the dependent variable, and total word count as the covariate.

Research Questions

RQ1: Is there a difference among the electronic Standardized Letter of Evaluation aggregate ratings of male and female medical students applying to an emergency medicine residency based on their self-identified ethnicity?

RQ2: Is there a difference in communal language used in narrative portions of applications to an emergency medicine residency found by Linguistic Inquiry and Word Count 2015 among male and female medical students when controlling for total word count?

Null Hypotheses

$H_0 1$: There is no significant difference among electronic Standardized Letter of Evaluation aggregate ratings of medical students between male and female medical students.

$H_0 2$: There is no significant difference between electronic Standardized Letter of Evaluation aggregate ratings among applicants based on their self-identified ethnicity.

$H_0 3$: There is no significant interaction among the electronic Standardized Letter of

Evaluation aggregate ratings of medical students among male and female medical students based on their self-identified ethnicity.

H_04: There is no difference in communal language used in narrative portions of emergency medicine residency applications, as measured by LIWC, among male and female medical students when controlling for total word count.

Descriptive Statistics

Variables of interest and demographic data on applicants were evaluated including gender, ethnicity, age, training background, number of eSLOEs, and number of non-eSLOE letters of recommendation provided as part of an application. Additionally, variables on eSLOEs including gender(s) of author(s), rating, narrative length by word count, and communal language score by LIWC were evaluated.

There were more male applicants (54.9%) than female applicants (see Table 1), with an average age of 31.9 years old (median 30 years). Overall, the applicants primarily self-identified as White (39.9%) or Asian (32.4%), with most applicants not identifying a secondary ethnicity (85.5%) (see Table 2). There were similar amounts of applicants that were trained within allopathic medical schools in the United States (43.4%) and osteopathic medical schools in the United States (38.7%), with less undergoing non-U.S. training (see Table 3). Most applicants provided two eSLOEs (43.4%) and one non-eSLOE letter of recommendation (41.0%). The average rating on an eSLOE was 48.6 (median = 49), with an average of narrative length provided of 184 words (median = 173 words), and an average communal rating of 4.8 (median = 4.6). Of the 391 eSLOEs written, most were provided by a group of authors that include males and females (47.6%) or a single male author (27.4%) (see Table 4).

Table 3

Medical School Training of Sample Population

U.S. Allopathic	67 (38.7%)
U.S. Osteopathic	75 (43.4%)
Caribbean	26 (15%)
International Non-Caribbean	5 (2.9%)

Table 4

Authorship of Sample Population's Electronic Standardized Letter of Evaluations

|	
One Male Author	107 (27.4%)
One Female Author	47 (12.0%)
Multiple Authors (Mixed Gender)	186 (47.6%)
Multiple Male Authors	36 (9.2%)
Multiple Female Authors	15 (3.8%)

Results

Null Hypothesis One

Data screening was initially performed by visual inspection of sorted data fields to evaluate for extreme or impossible values of eSLOE aggregate rating, gender, and ethnicity. No extreme or impossible values were found, however there were multiple cases that had no eSLOE aggregate rating due to no eSLOE provided, or no primary ethnicity provided. Those cases were managed using a preplanned approach of pair-wise deletion.

As a two-way ANOVA was planned for analysis of null hypothesis one, assumptions of a two-way ANOVA were evaluated. There are six assumptions to meet in order to perform a two-way ANOVA (Laerd Statistics, 2018-b). First, the dependent variable is continuous, which was met by the design of the study. Second, the independent variable is categorical and independent, which was again met by the design of the study. Third, observations must be

independent, with no individual being part of more than one grouping. To meet this observation, any individual that provided a second self-identified ethnicity was then removed from the dataset and ensured the assumption was met. Fourth, a box-and-whisker plot was performed of eSLOE aggregate ratings to evaluate for outliers, demonstrating a single outlier (see Figure 1). This data point was further investigated due to being an outlier and appeared to represent an inaccurate data point that was not identified in the initial round of data screening, and so was removed. The fourth assumption was then met. Fifth, the dependent variable of eSLOE aggregate rating should fall in a normal distribution. To evaluate this, a Shapiro-Wilk test was performed as Razali and Wah (2011) found it to be more powerful than a Kolmogorov-Smirnov or Anderson-Darling test. A Shapiro-Wilk test was performed demonstrating $p = 0.006$ and that it did not meet the normality assumption (see Table 5). Osborne (2010) notes the box-cox transformation to be a best practice for normalizing data and equalizing variance. It is also noted that nonparametric tests may benefit from improved variable normality. Due to that, a box-cox transformation was performed, with an optimal lambda = 3.400. At that point the Shapiro-Wilk test was again performed to determine normality and if a nonparametric test would be required. The Shapiro-Wilk test then demonstrated $p = 0.210$, demonstrating a normal distribution and that the fifth assumption was met (see Table 6). Sixth, homogeneity of variances was evaluated using Levene's test [*Levene F*(1, 138) = 3.757, $p = 0.055$] which was not significant and the assumption was met. A total of 140 applicants remained after cleaning, which still exceeded the minimum required for a two-way ANOVA according to Gall et al. (2007) of 126 individuals.

Figure 1

Test Between-Subject Effects Box-and-Whisker Plot

eSLOE Aggregate Rating
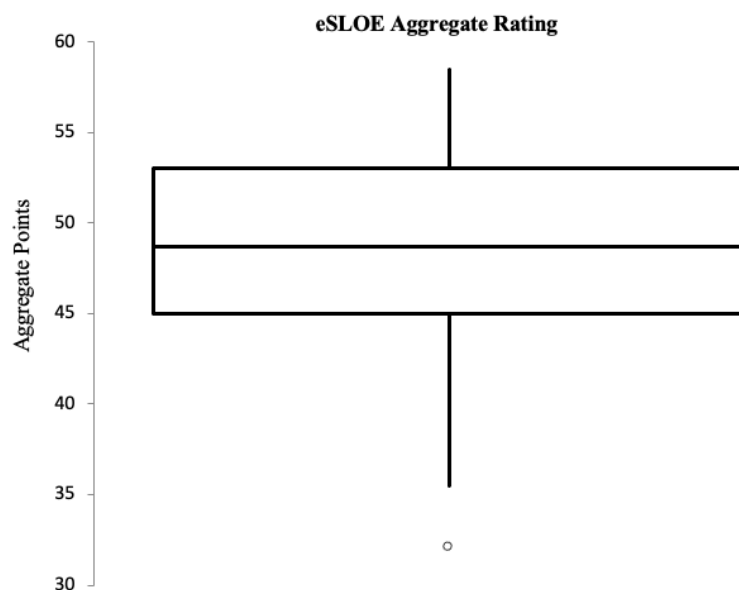
Note. One outlier was noted in the initial box-and-whisker plot of electronic Standardized Letter of Evaluation (eSLOE) aggregate ratings.

Table 5

Initial Two-Way ANOVA Shapiro-Wilk Test

Shapiro-Wilk	0.973
p value (Shapiro-Wilk)	0.006

Table 6

Two-Way ANOVA Shapiro-Wilk Test After Box-Cox Transformation

Shapiro-Wilk	0.986
p value (Shapiro-Wilk)	0.210

As all assumptions were met, a two-way ANOVA using an alpha = 0.025 after Bonferroni correction was applied (see Table 7). The significance level (alpha) can be calculated as: $0.05/2 = 0.025$ (Vickerstaff et al., 2019). This was used to evaluate if there was a significant

difference among eSLOE aggregate ratings of medical students between male and female

medical students (see Table 7). As $p = 0.633$, the two-way ANOVA fails to reject the null

hypothesis and demonstrates no significant difference in aggregate eSLOE ratings based on

gender of applicant. The effect size $\eta^2 = 0.002$ would qualify as less than a small effect.

Following statistical evaluation, this demonstrates failure to reject null hypothesis one.

Table 7

Test Between-Subject Effects

Source	df	SS	MS	F	p	η^2
Gender	1	9.01E+09	9.01E+09	0.229	0.633	0.002
Ethnicity	4	6.22E+10	1.55E+10	0.395	0.812	0.012
Interaction	4	5.74E+10	3.94E+10	0.364	0.834	0.011
Error	130	5.12E+12	3.94E+10			
Total	139	5.24E+12	3.77E+10			

Null Hypothesis Two

Data screening was performed as previously described in null hypothesis one with visual

inspection of data fields for extreme, impossible, or missing data. The assumptions for a two-

way ANOVA were evaluated as previously discussed in null hypothesis one. A two-way

ANOVA using an alpha $= 0.025$ after Bonferroni correction was applied (see Table 7). This was

used to evaluate if there was a significant interaction among the eSLOE aggregate ratings of

medical students based on their self-identified ethnicity. As $p = 0.812$, the two-way ANOVA

fails to reject the null hypothesis and demonstrates no significant difference in aggregate eSLOE

ratings based on ethnicity of the applicant. The effect size $\eta^2 = 0.012$ would qualify as a small

effect size (see Table 7). Following statistical evaluation, this demonstrates failure to reject null

hypothesis two.

Null Hypothesis Three

Data screening was previously described in null hypothesis one with visual inspection of data fields. Evaluation of assumptions for a two-way ANOVA were previously described in hypothesis one. A two-way ANOVA using an alpha = 0.025 after Bonferroni correction was applied was performed to evaluate if there was a significant interaction among aggregate eSLOE ratings of medical students among male and female medical students based on their self-identified ethnicity (see Table 7). The $p = 0.834$ fails to reject the null hypothesis and does not demonstrate a significant interaction among aggregate eSLOE ratings of medical students between gender and self-identified ethnicity. The effect size $\eta^2 = 0.011$ would qualify as a small effect size (see Table 7). Following statistical evaluation, this demonstrates failure to reject null hypothesis three.

Null Hypothesis Four

Data screening was initially performed by visual inspection of sorted data fields to evaluate for extreme or impossible values of gender, communal language, and total word count. No extreme or impossible values were found, yielding 391 eSLOEs, which was greater than the minimum required sample of 166 for an analysis of covariance (ANCOVA) with a power of .07 at the .05 alpha level according to Gall et al. (2007).

As an ANCOVA was planned for analysis of null hypothesis four, assumptions of an ANCOVA were evaluated. There are nine assumptions to meet in order to perform a two-way ANOVA (Laerd Statistics, 2018-a). The first three assumptions are met by the design of the study, and are that the dependent variable of communal language and covariate of word count are continuous, the independent variable uses two categorical independent groups, and observations are independent of each other. A box-and-whisker plot was performed to evaluate for the fourth

assumption that there are no outliers, which demonstrated outliers for communal language score and total word count (see Figures 2 and 3). Howitt and Cramer's (2011) approach was used for these outliers with review and then removal of 3 values. At that point assumption four was then met. Each category's normality was assessed using Shapiro-Wilk testing, as it has been found to be more powerful than a Kolmogorov-Smirnov or Anderson-Darling test (Razali & Wah, 2011). The Shapiro-Wilks test demonstrated $p = 0.732$ for Male, $p = 0.608$ for Female, meeting the fifth assumption (see Table 8). The Levene test for homogeneity of variances was not significant [*Levene F*$(1, 389) = 2.968, p = 0.086$], which met the sixth assumption. Scatterplots were assessed for linear relationship for each group (i.e., Male and Female) and appeared to demonstrate a linear relationship with lines of fit included, meeting assumption seven (see Figures 4 and 5). A scatterplot appeared to demonstrate homoscedasticity (see Figure 6), with Breusch-Pagan test performed for further confirmation demonstrating $p = 0.700$ and homoscedasticity (see Table 9). This demonstrated that assumption eight was met. Finally, homogeneity of the regression slopes was evaluated with an interaction hypothesis as discussed by Gignac (2019). An interaction hypothesis evaluates the association of the covariate and dependent variable interacting with the independent variable groups. More specifically it evaluates the slopes between independent groups of the covariate and dependent variables. The homogeneity of the regression slopes evaluation using the interaction hypothesis method demonstrated $p = 0.115$, meeting the assumption (see Table 10).

Figure 2

Communal Language Box-and-Whisker Plot

Gender
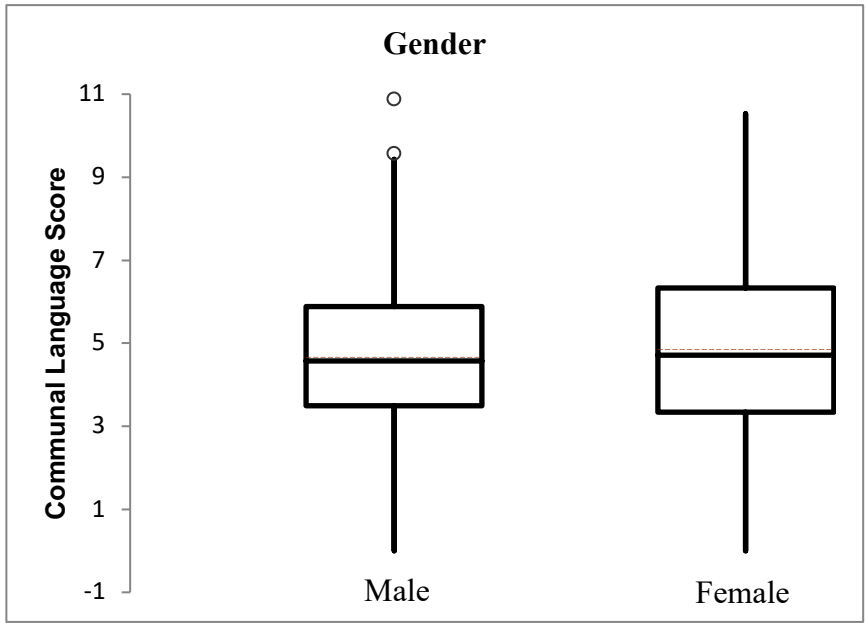
Note. Communal language score of the electronic standardized letter of evaluations' narrative portions was calculated by Linguistic Inquiry and Word Count 2015, with two outliers noted in the scores for males.

Figure 3

Word Count Box-and-Whisker Plot

Gender
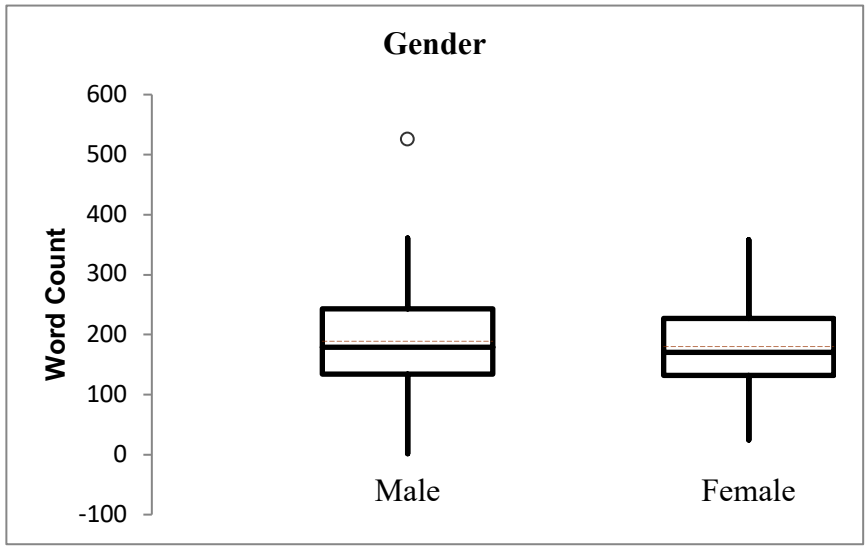
Note. Word count of the electronic standardized letter of evaluations' narrative portions was calculated by Linguistic Inquiry and Word Count 2015, with one outlier noted in the scores for

males.

Table 8

ANCOVA Shapiro-Wilk Test

	Male	Female
Shapiro-Wilk	0.995	0.995
p value	0.732	0.608

Figure 4

ANCOVA Linear Relationship Scatter Plot for Males
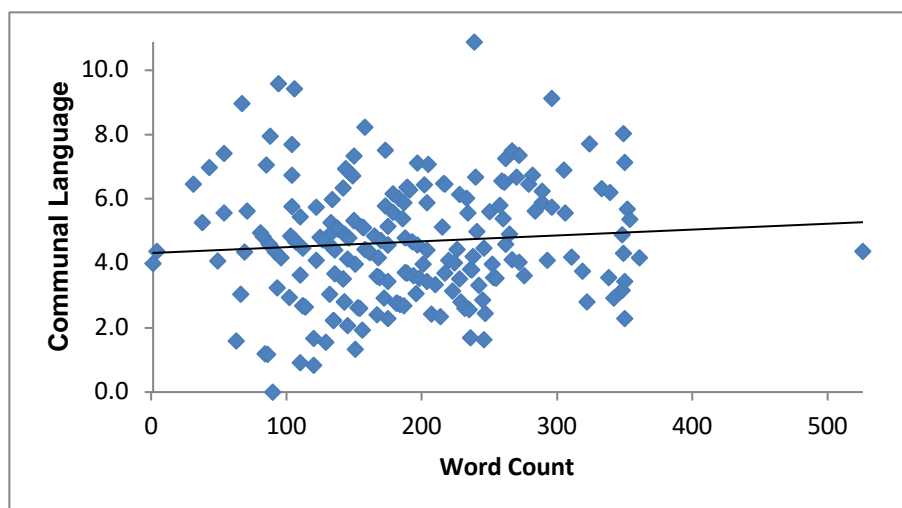

Figure 5
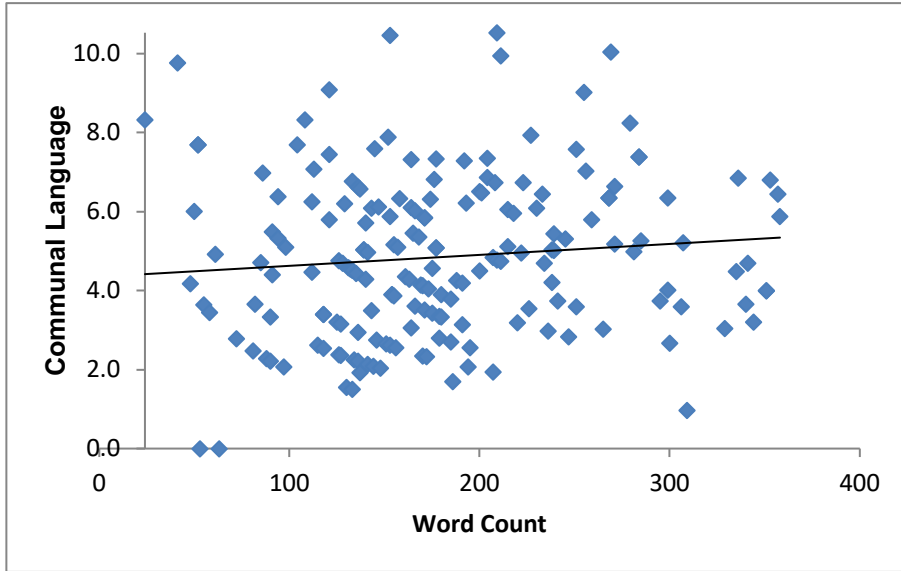
ANCOVA Linear Relationship Scatter Plot for Females

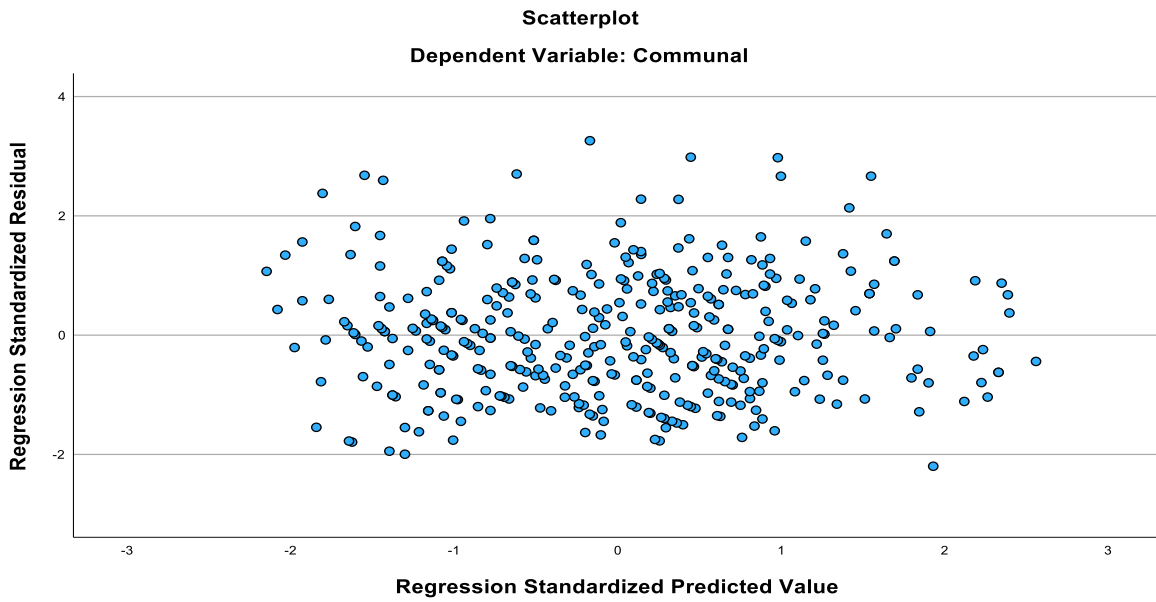**Figure 6**

ANCOVA Homoscedasticity Scatter Plot

Table 9

ANCOVA Breusch-Pagan Test

Chi-Square	0.149
df	1
p value	0.700

Table 10

ANCOVA Homogeneity of Regression Slopes

Source	*df*	*SS*	*MS*	*F*	*p*
Corrected Model	2	15.884	7.942	2.171	0.115
Intercept	1	1136.369	1136.369	310.62	<0.001
Gender*WordCount	2	15.884	7.942	2.171	0.115
Error	388	1419.455	3.658		
Total	391	10259.746			
Corrected Total	390	1435.339			

As all assumptions were met, an ANCOVA using an alpha = 0.025 after Bonferroni correction was applied was performed to evaluate if there was a difference in communal language used among male and female medical students when controlling for total word count. The ANCOVA demonstrated $p = 0.295$ and fails to reject the null hypothesis demonstrating a failure to reject the null hypothesis that there is no difference in communal language among male and female students when controlling for total word count (see Table 11). The effect size $\eta^2 = 0.003$ demonstrates an effect size less than small.

Table 11

Tests of Between-Subjects Effects

Source	SS	df	MS	F	p	η^2
Corrected Model	15.363	2	7.681	2.099	0.124	0.011
Intercept	1148.847	1	1148.847	313.916	<0.001	0.447
WordCount	11.988	1	11.988	3.276	0.071	0.008
Gender	4.028	1	4.028	1.101	0.295	0.003
Error	1419.976	388	3.660			
Total	10259.746	391				
Corrected Total	1435.339	390				

CHAPTER FIVE: CONCLUSIONS

Overview

Chapter Five discusses the results of the analyses performed and implications of those results. Specifically, it discusses the results associated with each research question within the context of prior literature available. Further, it examines the potential implications the findings of this study has. Following the potential implications, it considers the limitations of the study that must be considered. Finally, recommendations for future research are suggested based upon the existing literature and results of the study.

Discussion

The purpose of this quantitative, causal-comparative study was to examine for certain potential biases within application materials submitted through ERAS to emergency medicine residency programs. Specifically, it first sought to evaluate if gender or ethnicity impacted medical students' aggregate eSLOE scores. Second, it looked to evaluate if gender impacted the characteristics of language used, specifically communal language, within narrative evaluative portions of eSLOEs.

Gender and Ethnicity Impact on Aggregate eSLOE Ratings

The first research question asked if there was a difference among the eSLOE aggregate ratings of male and female medical students applying to an emergency medicine residency based on their self-identified ethnicity. This used gender as the first independent variable, defined as socially constructed differences between men and women (Risberg et al., 2009; Verdonk et al., 2009). Self-identified ethnicity was used as the second independent variable, defined as the scientific terminology of the socially constructed and multi-faceted identify of an individual (American Psychological Association, 2019; Santos et al., 2010). Aggregate eSLOE rating was

the dependent variable, which is the simple average of ratings on the eSLOEs for an individual (Hopson et al., 2019).

There have been iterative changes to the process of The Match for a position in an emergency medicine residency program for decades (Jackson et al., 2019; Love et al., 2014, 2020; Martin & McNamara, 2014). One aspect of this is the eSLOE, which has been considered "the gold standard" for use during the application process to an emergency medicine residency program (Jackson et al., 2019, p. 182). Significant research has been done on the eSLOE to evaluate and provide support in the form of evidence of validity (Girzadas et al., 1998; Jackson et al., 2019; Kukulski & Ahn, 2021; Love et al., 2020). Unfortunately, an instrument may have strong evidence of validity but still not be equitable as demonstrated by Davis et al. (2013) using the Medical College Admission Test. Multiple authors have demonstrated a potential inequity in the eSLOE. Hopson et al. (2019) found that females and white individuals were favored in eSLOE ratings when compared to males and black individuals. Andrusaitis et al. (2020) and Mannix et al. (2022) specifically found that female applicants performed overall better on eSLOE ratings but did not delve into ethnicity. Interestingly, Wilson et al. (2021) found that male applicants had a positive bias when specifically focusing on their home institution. Considering this from a theoretical framework, it is possible that feminism and homophily both may be demonstrated. From a feminist perspective, improving the outcomes for females to end oppression and exploitation could explain the overall bias in favor of females. Homophily, and the ongoing trend to replace the dominant majority with similar individuals, could provide a theoretical explanation why a male may perform better at their home institution, or why white individuals may have performed better than black individuals.

The results of the first research question of this study appear to contradict prior findings

demonstrating a bias based on an applicant's gender and ethnicity. Ideally, the lack of bias may reflect a recognition of underlying gender and ethnicity issues and push for equity that has appeared in the literature (Ansell & McDonald, 2015; Boatright et al., 2018; Madsen et al., 2017). Unfortunately, this study may reflect a regional specific trend to improvement in equity without significant improvement at the national level as demonstrated by Kearse et al. (2022) in general surgery training. The importance of a shift away from gender and ethnic-based biases is important nonetheless, with a need to now determine if this is due to geographic location, represents a trend over time, or may be a type II error.

Communal Language Influence by Gender and Word Count

The second research question asked if there was a difference in communal language used in narrative portions of applications to an emergency medicine residency found by LIWC among male and female medical students when controlling for total word count. Gender is the independent variable for this question, which refers to socially constructed differences found between men and women (Risberg et al., 2009; Verdonk et al., 2009). Communal language was the dependent variable, and defined as relationship-oriented language (Grimm et al., 2020; Pietraszkiewicz et al., 2019). Total word count of the narrative portion of the eSLOE represented the covariate, and can be defined as the total number of words within a certain text (Turrentine et al., 2019).

Prior studies have found linguistic differences in narrative recommendations of residency applicants with females having more communal language used (Aggarwal et al., 2018; Grimm et al., 2020; Pietraszkiewicz et al., 2019). Communal language has been suggested to be stereotypically reflective of females (Heilman, 2001; Pietraszkiewicz et al., 2019). Literature demonstrating these findings may be explained from a feministic perspective considering the

importance of language as it relates to gender as a social construct (Frug, 1992).

Again, this study seems to contradict the existing literature, and instead found no significant difference in communal language use based on gender in narrative recommendations of eSLOEs. Similar to the findings of research question one that contradict prior literature, it may be related to recent literature for equity (Boatright et al., 2018; Madsen et al., 2017). Alternatively, prior studies demonstrating the bias may have prompted work to address such findings (Grimm et al., 2020; Li et al., 2017; Lin et al., 2019). There is also the potential it may be a regional improvement in equity without an accompanying national trend (Kearse et al., 2022). Despite that, this finding also is important to highlight the potential shift, be it geographically or temporally. The need for further evaluation is clear to help determine ongoing trends, underlying reasons, and to rule out a type II error being responsible for the findings.

Implications

These results demonstrate that there does not appear to be differences in eSLOE ratings of medical students applying to emergency medicine residency programs based on gender or ethnicity during the most recent (2022-2023) Match. There are no significant differences in communal language used within the narrative portions of the eSLOE based on gender. This implies that a potential shift compared to prior studies demonstrating biases in application materials to emergency medicine residency programs (Andrusaitis et al., 2020; Li et al., 2017). This shift represents an improvement in The Match to emergency medicine with regards to diversity and equity for applicants. Such a shift in gender and ethnic-based biases have potential implications not only to individuals currently in medical training, but also to their future patients. Further, it may be representative of a larger change in respect to diversity and equity among other specialties within medicine, or medical training as a whole. Ultimately, this study helps

close the gap in the literature on certain current biases in the application process to an emergency

medicine residency program.

Limitations

There are multiple limitations to be aware of when considering using a causal-

comparative research design. First, these designs are not able to provide definitive cause-and-

effect relationships (Schenker & Rumrill, 2004). A primary reason for this is that the design does

not control external or extraneous variables, which also may influence differences between

groups. Another potential limitation in causal-comparative research is the sensitivity and

potential impact of language (Kettler, 2019). Kettler (2019) highlights this issue such as noting

the difference between sex and gender, with the former being biologic and the latter referring to

a socially constructed preferred identity. Kettler also points out a third limitation to be aware of

with these designs, that within group variance may be larger than between group differences. The

importance of that is that it is difficult or impossible to apply results from causal-comparative

studies to independent cases, and instead primarily applicable when considering groups.

Another set of limitations to be aware of come from using a convenience sampling

approach. Etikan et al. (2016) point out limitations with convenience sampling including

potential for bias and potential for outliers to have a larger impact. There is also a risk of threat to

external validity dependent on how the sample and total population align (Sedgwick, 2013).

These limitations are important in considering the results of the study and their applicability or

generalizability.

It is also important to recognize this study must be viewed in the context of the research

setting, including the time and place. External validity refers to the ability to apply a study's

findings to a broader sample or population (Findley et al., 2021). This concept includes

generalizability, or the ability to apply to a larger population. It also includes transportability, or the ability to apply to other target populations. There are some threats to the external validity as the sample was drawn from applicants to a single program in the Western United States. Khelifa and Mahdjoub (2022) discuss consideration of geographic influence on diversity, equity, and inclusion. Further, Kearse et al. (2022) suggest that certain geographic regions may be showing trends of improvement in diversity, equity, and inclusion, but that it may not mean there is a corresponding national trend.

Internal validity refers to the certainty of a researcher that findings of a dependent variable are due to changes of the independent variable. Confounding variables are a threat to the internal validity of a study. This study attempted to limit certain confounders such as by using only a single year's applicants to an emergency medicine residency program. Despite that, there are multiple potential confounding variables to be aware of. Potential confounding variables for this study include the geographic locations of applicants and authors of their eSLOEs, gender of the authors of the eSLOEs, number of rotations and timing of eSLOEs in respect to rotations, sequential eSLOEs written for those with multiple eSLOEs, and selection bias of applicants to a specific emergency medicine residency program.

Another threat to this study is due to the abstraction process and nature of archival studies. Luckey et al. (1982) discuss issues surrounding archival studies, specifically on the accuracy and reliability with such data. As this study used manual abstraction of data from an archival source, there is the potential that errors were introduced during the abstraction process. While data was screened for extreme and impossible values, there is potential that incorrect data that did not appear extreme or impossible may have been entered during abstraction.

Finally, a limitation with this study comes from the analytical and statistical approaches

used. This study uses multiple statistical tests for analysis, which can increase the risk of a Type I error (Armstrong, 2014). While the Bonferroni correction was used to address this and attempt to control for Type I error, it can lead to Type II errors. Another limitation associated with the approaches used come from use of Howitt and Cramer's (2011) suggested management of outliers. Bakker and Wicherts (2014) criticize the approach, with the potential increased risk of a Type I error noted.

Recommendations for Future Research

There are multiple avenues that could be considered for areas of future research.

1. Research directed at examining trends in residency application materials over time may be beneficial to not only identify trends in biases, but also to track changes.

2. Research seeking to determine reasons for changes over time in biases within residency applications.

3. Research on other aspects of language aside from communal language within narrative portions of residency applications.

4. Research seeking to identify other factors, such as age or disability, that may generate bias within application materials to a residency program.

5. Research seeking to evaluate fields outside the specialty of emergency medicine.

6. Research using a single statistical technique to avoid the Bonferroni correction allowing mitigation of Type II errors.

7. Research using an alternative approach to outlier management that may allow mitigation of Type I errors.

8. Qualitative approaches to this area of research may provide information on biases perceived from the standpoint of applicants, evaluators, and application reviewers.

References

Aberson, C. L. (2019). *Applied power analysis for the behavioral sciences.* Routledge.

Accreditation Council for Graduate Medical Education. (n.d.). *Common Program Requirements*. https://www.acgme.org/what-we-do/accreditation/common-program-requirements/.

Accreditation Council for Graduate Medical Education. (2020). *Common Program Requirements (Residency).* https://www.acgme.org/Portals/0/PFAssets/ProgramRequirements/CPRResidency2021.pdf.

Acker, J. (1990). Hierarchies, jobs, bodies: A theory of gendered organizations. *Gender & society, 4*(2), 139-158.

Acker, J. (2006). Inequality regimes: Gender, class, and race in organizations. *Gender & Society, 20*(4), 441-464.

Aggarwal, S., Grob, S., Banerjee, D., Putzel, P. J., & Tao, J. (2018). Key Word Use in Letters of Recommendation for Ophthalmology Residency Applicants According to Race, Gender, and Achievements. *Journal of Academic Ophthalmology, 10*(01), e163-e171.

Aisen, C. M., Sui, W., Pak, J. S., Pagano, M., Cooper, K. L., & Badalato, G. M. (2018). Gender differences in the urology residency match—does it make a difference?. *Urology, 111*, 39-43.

Almarzooq, Z. I., Lillemoe, H. A., White-Manigault, E., Wickham, T., & Curtin, L. S. (2021). The Single Match: Reflections on the National Resident Matching Program's Sustained Partnership With Learners. *Academic Medicine, 96*(8), 1116.

American Psychological Association. (2019). *APA Guidelines on Race and Ethnicity in Psychology*.

Andolsek, K. M. (2016). Improving the Medical Student Performance Evaluation to Facilitate Resident Selection. *Academic Medicine, 91*(11), 1475-1479.

Andrusaitis, J., Clark, C., Saadat, S., Billimek, J., Paradise, S., Wray, A., Wiechmann, W., & Boysen-Osborn, M. (2020). Does applicant gender have an effect on standardized letters of evaluation obtained during medical student emergency medicine rotations?. *AEM Education and Training, 4*(1), 18-23.

Ansell, D. A., & McDonald, E. K. (2015). Bias, black lives, and academic medicine. *New England Journal of Medicine, 372*(12), 1087-1089.

Armstrong, R. A. (2014). When to use the Bonferroni correction. *Ophthalmic and Physiological Optics, 34*(5), 502-508.

Association of American Medical Colleges. (n.d.-a). *About ERAS.* https://students-residents.aamc.org/applying-residencies-eras/about-eras

Association of American Medical Colleges. (n.d.-b). *Documents for ERAS® Residency Applicants.* https://students-residents.aamc.org/applying-residencies-eras/documents-eras-residency-applicants

Association of American Medical Colleges. (n.d.-c) *ERAS 2022 Residency Timeline.* https://students-residents.aamc.org/eras-tools-and-worksheets-residency-applicants/eras-2022-residency-timeline

Association of American Medical Colleges. (2017). *Recommendations for Revising the Medical Student Performance Evaluation (MSPE).* https://www.aamc.org/media/23311/download

Association of American Medical Colleges. (2021a). *Data Policy.* Retrieved May 30, 2021, from https://students-residents.aamc.org/media/7706/download

Association of American Medical Colleges. (2021b). *Privacy Statement.* Retrieved May 30,

2021, from https://www.aamc.org/privacy

Association of American Medical Colleges. (2023). *Emergency Medicine*.

https://www.aamc.org/media/39326/download

Axelson, R. D., Solow, C. M., Ferguson, K. J., & Cohen, M. B. (2010). Assessing implicit

gender bias in medical student performance evaluations. *Evaluation & the health*

professions, *33*(3), 365-385.

Babal, J. C., Gower, A. D., Frohna, J. G., & Moreno, M. A. (2019). Linguistic analysis of

pediatric residency personal statements: gender differences. *BMC medical education,*

19(1), 1-9.

Backhus, L. M., Lui, N. S., Cooke, D. T., Bush, E. L., Enumah, Z., & Higgins, R. (2019).

Unconscious bias: addressing the hidden impact on surgical education. *Thoracic surgery*

clinics, 29(3), 259-267.

Bakker, M., & Wicherts, J. M. (2014). Outlier removal, sum scores, and the inflation of the Type

I error rate in independent samples t tests: the power of alternatives and

recommendations. *Psychological methods, 19*(3), 409.

Belanger, N. (2018). "Inclusive Leadership" If we build it will they come? *Canadian Military*

Journal, 19(1), 32-39.

Benbassat, J., & Baumal, R. (2007). Uncertainties in the selection of applicants for medical

school. *Advances in Health Sciences Education, 12*(4), 509-521.

Bennett, J. M. (1989). Feminism and history. *Gender & History, 1*(3), 251-272.

Berriochoa, C., Reddy, C. A., Dorsey, S., Campbell, S., Poblete-Lopez, C., Schlenk, R., Spencer,

A., Lee, J., Eagleton, M., & Tendulkar, R. D. (2018). The residency match: interview

experiences, postinterview communication, and associated distress. *Journal of graduate*

medical education, *10*(4), 403-408.

Bird, J. B., Friedman, K. A., Arayssi, T., Olvet, D. M., Conigliaro, R. L., & Brenner, J. M. (2021). Review of the Medical Student Performance Evaluation: analysis of the end-users' perspective across the specialties. *Medical Education Online, 26*(1), 1876315.

Blackshaw, A. M., Watson, S. C., & Bush, J. S. (2017). The cost and burden of the residency match in emergency medicine. *Western Journal of Emergency Medicine, 18*(1), 169.

Boatright, D., Branzetti, J., Duong, D., Hicks, M., Moll, J., Perry, M., Pierce, A., Samuels, E., Smith, T., Angerhofer, C., & Heron, S. (2018). Racial and ethnic diversity in academic emergency medicine: how far have we come? Next steps for the future. *AEM education and training, 2*, S31-S39.

Boone, H. N., & Boone, D. A. (2012). Analyzing likert data. *Journal of extension, 50*(2), 1-5.

Boulis, A. K., & Jacobs, J. A. (2011). *The changing face of medicine: women doctors and the evolution of health care in America.* Cornell University Press.

Boysen-Osborn, M., Yanuck, J., Mattson, J., Toohey, S., Wray, A., Wiechmann, W., Lahham, S., & Langdorf, M. I. (2017). Who to interview? Low adherence by US medical schools to medical student performance evaluation format makes resident selection difficult. *Western Journal of Emergency Medicine, 18*(1), 50.

Byrd, W. M., & Clayton, L. A. (2001). Race, medicine, and health care in the United States: a historical survey. *Journal of the National Medical Association, 93*(3 Suppl), 11S.

Capers, IV, Q. (2020). How clinicians and educators can mitigate implicit bias in patient care and candidate selection in medical education. *ATS scholar, 1*(3), 211-217.

Capers, IV, Q., Clinchot, D., McDougle, L., & Greenwald, A. G. (2017). Implicit racial bias in medical school admissions. *Academic Medicine, 92*(3), 365-369.

Carvalho, I., Costa, C., Lykke, N., & Torres, A. (2019). Beyond the glass ceiling: Gendering tourism management. *Annals of Tourism Research, 75*, 79-91.

Catalanotti, J. S., Andolsek, K. M., & Berger, J. S. (2017). The AAMC medical student performance evaluation task force recommendations: do they go far enough?. *Journal of graduate medical education, 9*(1), 1.

Chapman, B. V., Rooney, M. K., Ludmir, E. B., De La Cruz, D., Salcedo, A., Pinnix, C. C., Das, P., Jagsi, R., Thomas, Jr., C. R., & Holliday, E. B. (2020). Linguistic Biases in Letters of Recommendation for Radiation Oncology Residency Applicants from 2015 to 2019. *Journal of Cancer Education,* 1-8.

Chapman, C. H., Hwang, W. T., Wang, X., & Deville, C. (2019). Factors that predict for representation of women in physician graduate medical education. *Medical education online, 24*(1), 1624132.

Chen, C., & Mullan, F. (2009). The separate osteopathic medical education pathway: uniquely addressing national needs. *Academic Medicine, 84*(6), 695.

Clifton, S. M., Hill, K., Karamchandani, A. J., Autry, E. A., McMahon, P., & Sun, G. (2019). Mathematical model of gender bias and homophily in professional hierarchies. *Chaos: An Interdisciplinary Journal of Nonlinear Science, 29*(2), 023135.

Cook, D. A., & Beckman, T. J. (2006). Current concepts in validity and reliability for psychometric instruments: theory and application. *The American journal of medicine, 119*(2), 166-e7.

Council of Residency Directors in Emergency Medicine. (2016). *Electronic Standardized Letter of Evaluation.*

Craig, E., Brotzman, E., Farthing, B., Giesey, R., & Lloyd, J. (2021). Poor match rates of

osteopathic applicants into ACGME dermatology and other competitive specialties. *Journal of Osteopathic Medicine, 121*(3), 281-286.

Cuddy, M. M., Dillon, G. F., Clauser, B. E., Holtzman, K. Z., Margolis, M. J., McEllhenney, S. M., & Swanson, D. B. (2004). Assessing the validity of the USMLE step 2 clinical knowledge examination through an evaluation of its clinical relevance. *Academic Medicine, 79*(10), S43-S45.

Custers, E. J., & ten Cate, O. (2018). The history of medical education in Europe and the United States, with respect to time and proficiency. *Academic Medicine, 93*(3), S49-S54.

Davis, D., Dorsey, J. K., Franks, R. D., Sackett, P. R., Searcy, C. A., & Zhao, X. (2013). Do racial and ethnic group differences in performance on the MCAT exam reflect test bias?. *Academic Medicine, 88*(5), 593-602.

Davis, K. (2008). Intersectionality as buzzword: A sociology of science perspective on what makes a feminist theory successful. *Feminist theory, 9*(1), 67-85.

Deligonul, S. (1998). Heuristics to Analyze Bivariate Data. *Quality Engineering, 10*(3), 437-443.

Demzik, A., Filippou, P., Chew, C., Reines, K., Brown, S., Wallen, E. M., Viprakasit, D., Smith, A. B., & Tan, H. J. (2021). Linguistic Differences in Personal Statements of Urology Residency Applicants by Self-Reported Race and Ethnicity. *Urology, 00*, 1-7.

Denis, D. J. (2018). *SPSS data analysis for univariate, bivariate, and multivariate statistics.* John Wiley & Sons.

De Oliveira, G. S., Akikwala, T., Kendall, M. C., Fitzgerald, P. C., Sullivan, J. T., Zell, C., & McCarthy, R. J. (2012). Factors affecting admission to anesthesiology residency in the United States: choosing the future of our specialty. *The Journal of the American Society of Anesthesiologists, 117*(2), 243-251.

Desbiens, N. A., & Vidaillet, H. J. (2010). Discrimination against international medical graduates in the United States residency program selection process. *BMC medical education, 10*(1), 1-5.

Dharssi, S., Woreta, F. A., & Boland, M. V. (2020). Ophthalmology Applicant Perceptions of Two Residency Application Services: The San Francisco Match Central Application Service and Electronic Residency Application Service. *Journal of Academic Ophthalmology*, *12*(02), e188-e194.

Dimant, O. E., Cook, T. E., Greene, R. E., & Radix, A. E. (2019). Experiences of transgender and gender nonbinary medical students and physicians. *Transgender health, 4*(1), 209-216.

Douglas, I. I., & Hendrix, J. (2021). Black Medical Student Considerations in the Era of Virtual Interviews. *Annals of Surgery, 274*(2), 232-233.

Dowling, G. R., & Midgley, D. F. (1991). Using rank values as an interval scale. *Psychology & Marketing, 8*(1), 37-41.

Dressel, P., Hartfield, B. W., & Gooley, R. L. (1994). The dynamics of homosocial reproduction in academic institutions. *American University Journal of Gender, Social Policy & the Law., 2,* 37.

Dudău, D. P., & Sava, F. A. (2020). The development and validation of the Romanian version of Linguistic Inquiry and Word Count 2015 (Ro-LIWC2015). *Current Psychology*, 1-18.

Dugdale, D. C., Zieve, D., Conaway, B., & A.D.A.M. Editorial Team. (2019). Doctor of medicine profession (MD). *MedlinePlus.* https://medlineplus.gov/ency/article/001936.htm

Durham, S. R., Donaldson, K., Grady, M. S., & Benzil, D. L. (2018). Analysis of the 1990–2007

neurosurgery residency match: does applicant gender affect neurosurgery match outcome?. *Journal of neurosurgery, 129*(2), 282-289.

Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological review, 109*(3), 573.

Ebert, T. L. (1991). The" difference" of postmodern feminism. *College English, 53*(8), 886-904.

Edmond, M. B., Deschenes, J. L., Eckler, M., & Wenzel, R. P. (2001). Racial bias in using USMLE Step 1 scores to grant internal medicine residency interviews. *Academic Medicine, 76*(12), 1253-1256.

Eggermont, D., Smit, M. A. M., Kwestroo, G. A., Verheij, R. A., Hek, K., & Kunst, A. E. (2018). The influence of gender concordance between general practitioner and patient on antibiotic prescribing for sore throat symptoms: a retrospective study. *BMC family practice, 19*(1), 1-8.

Espinosa, L. L., Turk, J. M., Taylor, M., & Chessman, H. M. (2019). Race and ethnicity in higher education: A status report.

Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. *American journal of theoretical and applied statistics*, *5*(1), 1-4.

Ezimora, C. (2020). Please make the tragic death of this physician mean something. *KevinMD.* https://www.kevinmd.com/blog/2020/04/please-make-the-tragic-death-of-this-physician-mean-something.html

Fairmont, I., Farrell, N., Johnson, A. P., & Cabrera-Muffly, C. (2020). Influence of gender and racial diversity on the otolaryngology residency match. *Otolaryngology–Head and Neck Surgery, 162*(3), 290-295.

Fay, M., & Williams, L. (1993). Gender bias and the availability of business loans. *Journal of*

Business Venturing, 8(4), 363-376.

Ferreira, M., Rolim, V., Mello, R. F., Lins, R. D., Chen, G., & Gašević, D. (2020, March).

Towards automatic content analysis of social presence in transcripts of online

discussions. In *Proceedings of the tenth international conference on learning analytics &*

knowledge (pp. 141-150).

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. SAGE.

Figueiredo, S. (2020, July). A psycholinguistic analysis of presidential? discourses concerning

the COVID-19 context: authenticity and emotional tone. In *Proceedings of International*

Academic Conferences (No. 10613009). International Institute of Social and Economic

Sciences.

Filippou, P., Mahajan, S., Deal, A., Wallen, E. M., Tan, H. J., Pruthi, R. S., & Smith, A. B.

(2019). The presence of gender bias in letters of recommendations written for urology

residency applicants. *Urology, 134*, 56-61.

Findley, M. G., Kikuta, K., & Denly, M. (2021). External validity. *Annual Review of Political*

Science, 24, 365-393.

Foley, M., & Williamson, S. (2018). Does anonymising job applications reduce gender bias?

Understanding managers' perspectives. *Gender in Management: An International*

Journal. 33(8), 623-635.

Francis, M. E., & Booth, R. J. (1993). Linguistic inquiry and word count. *Southern Methodist*

University: Dallas, TX, USA.

Freeman, B. D. (2016). Is it time to rethink postgraduate training requirements for licensure?.

Academic Medicine, 91(1), 20-22.

French, J. C., Zolin, S. J., Lampert, E., Aiello, A., Bencsath, K. P., Ritter, K. A., Strong, A. T.,

Lipman, J. M., Valente, M. A., & Prabhu, A. S. (2019). Gender and letters of recommendation: a linguistic comparison of the impact of gender on general surgery residency applicants. *Journal of surgical education, 76*(4), 899-905.

Friedman, R., Fang, C. H., Hasbun, J., Han, H., Mady, L. J., Eloy, J. A., & Kalyoussef, E. (2017). Use of standardized letters of recommendation for otolaryngology head and neck surgery residency and the impact of gender. *The Laryngoscope, 127*(12), 2738-2745.

Frug, M. J. (1992). A postmodern feminist legal manifesto (an unfinished draft). *Harvard Law Review*, 1045-1075.

Fu, F., Nowak, M. A., Christakis, N. A., & Fowler, J. H. (2012). The evolution of homophily. *Scientific reports, 2*(1), 1-6.

Gall, M. D., Gall, J. P., & Borg, W. R. (2007). *Educational Research: An Introduction 8th Edition*. Pearson/Allyn & Bacon.

Garrick, J. F., Perez, B., Anaebere, T. C., Craine, P., Lyons, C., & Lee, T. (2019). The diversity Snowball effect: the quest to increase diversity in emergency medicine: a case study of highland's emergency medicine residency program. *Annals of emergency medicine, 73*(6), 639-647.

Gastwirth, J. L., Gel, Y. R., & Miao, W. (2009). The impact of Levene's test of equality of variances on statistical theory and practice. *Statistical Science, 24*(3), 343-360.

Gennissen, L. M., Stegers-Jager, K. M., de Graaf, J., Fluit, C. R., & de Hoog, M. (2021). Unraveling the medical residency selection game. *Advances in Health Sciences Education, 26*(1), 237-252.

Gevitz, N. (1994). 'Parallel and distinctive': the philosophic pathway for reform in osteopathic medical education. *Journal of Osteopathic Medicine, 94*(4), 328-328.

Gignac, G. E. (2019). *How2statsbook (Online Edition 1).*

    https://drive.google.com/file/d/1Td3kclZo_ohF_Axovk7QQ0nhVgX7YNzI/view

Girod, S., Fassiotto, M., Grewal, D., Ku, M. C., Sriram, N., Nosek, B. A., & Valantine, H.

    (2016). Reducing implicit gender leadership bias in academic medicine with an

    educational intervention. *Academic Medicine, 91*(8), 1143-1150.

Girzadas, D. V., Jr., Harwood, R. C., Davis, N., & Schulze, L. (2004). Gender and the council of

    emergency medicine residency directors standardized letter of recommendation.

    Academic Emergency Medicine, 11(9), 988-991.

Girzadas, D. V., Jr., Harwood, R. C., Dearie, J., & Garrett, S. (1998). A comparison of

    standardized and narrative letters of recommendation. *Academic emergency medicine,*

    5(11), 1101-1104.

Gong, D., Winn, B. J., Beal, C. J., Blomquist, P. H., Chen, R. W., Culican, S. M., Dagi Glass, L.

    R., Domeracki, D. F., Goshe, J. M., Jones., J. K., Khouri, A. S., Legault, G. L., Martin, T.

    J., Mitchell, K. T., … Al-Aswad, L. A. (2019). Gender differences in case volume among

    ophthalmology residents. *JAMA ophthalmology*, *137*(9), 1015-1020.

Gorman, E. H. (2005). Gender stereotypes, same-gender preferences, and organizational

    variation in the hiring of women: Evidence from law firms. *American Sociological*

    Review, 70(4), 702-728.

Grace, M. K. (2018). Friend or frenemy? Experiential homophily and educational track attrition

    among premedical students. *Social Science & Medicine, 212*, 33-42.

Greenwood, B. N., Carnahan, S., & Huang, L. (2018). Patient–physician gender concordance and

    increased mortality among female heart attack patients. *Proceedings of the National*

    Academy of Sciences, 115(34), 8569-8574.

Greenwood, B. N., Hardeman, R. R., Huang, L., & Sojourner, A. (2020). Physician–patient racial concordance and disparities in birthing mortality for newborns. *Proceedings of the National Academy of Sciences, 117*(35), 21194-21200.

Grimm, L. J., Redmond, R. A., Campbell, J. C., & Rosette, A. S. (2020). Gender and racial bias in radiology residency letters of recommendation. *Journal of the American College of Radiology, 17*(1), 64-71.

Gruppuso, P. A., & Adashi, E. Y. (2017). Residency placement fever: is it time for a reevaluation?. *Academic medicine: journal of the Association of American Medical Colleges, 92*(7), 923.

Gustafsson, I. M. (2018). *Gender equality in HRM: a case study in a Finnish recruitment agency* [Master's Thesis, Hanken School of Economics]. Hanken Collections. https://helda.helsinki.fi/dhanken/bitstream/handle/123456789/189245/gustafsson.pdf?sequence=1&isAllowed=y

Hanson, M., Schoonover, A, Skarica, B, Harrod, T, Bahr, N, & Guise, J. (2019). Implicit gender bias among US resident physicians. *BMC Medical Education, 19*(1), 1-9.

Harfmann, K. L., & Zirwas, M. J. (2011). Can performance in medical school predict performance in residency? A compilation and review of correlative studies. *Journal of the American Academy of Dermatology, 65*(5), 1010-1022.

Heilman, M. E. (2001). Description and prescription: How gender stereotypes prevent women's ascent up the organizational ladder. *Journal of social issues*, *57*(4), 657-674.

Hern, H., Trivedi, T., Alter, H. J., & Wills, C. P. (2016). How prevalent are potentially illegal questions during residency interviews? A follow-up study of applicants to all specialties in the National Resident Matching Program. *Academic Medicine, 91*(11), 1546-1553.

Hern, H. G., Jr., Alter, H. J., Wills, C. P., Snoey, E. R., & Simon, B. C. (2013). How prevalent are potentially illegal questions during residency interviews?. *Academic Medicine, 88*(8), 1116-1121.

Hewett, L., Lewis, M., Collins, H., & Gordon, L. (2016). Gender bias in diagnostic radiology resident selection, does it exist?. *Academic radiology, 23*(1), 101-107.

Hoffman, A., Grant, W., McCormick, M., Jezewski, E., Matemavi, P., & Langnas, A. (2019). Gendered differences in letters of recommendation for transplant surgery fellowship applicants. *Journal of surgical education, 76*(2), 427-432.

Hopson, L. R., Regan, L., Bond, M. C., Branzetti, J., Samuels, E. A., Naemi, B., Dunleavy, D., & Gisondi, M. A. (2019). The AAMC standardized video interview and the electronic standardized letter of evaluation in emergency medicine: a comparison of performance characteristics. *Academic Medicine, 94*(10), 1513-1521.

Howitt, D., & Cramer, D. (2011). *Introduction to statistics in psychology.* Harlow: Pearson.

Hu, Y. Y., Ellis, R. J., Hewitt, D. B., Yang, A. D., Cheung, E. O., Moskowitz, J. T., Potts, J. R., Buyske, J., Hoyt, D. B., Nasca, T. J., & Bilimoria, K. Y. (2019). Discrimination, abuse, harassment, and burnout in surgical residency training. *New England Journal of Medicine, 381*(18), 1741-1752.

Hughes, R. H., Kleinschmidt, S., & Sheng, A. Y. (2020). Using structured interviews to reduce bias in emergency medicine residency recruitment: worth a second look. *AEM Education and Training, 5*(S1), S130-134

Jackson, J. S., Bond, M., Love, J. N., & Hegarty, C. (2019). Emergency medicine standardized letter of evaluation (SLOE): findings from the new electronic SLOE format. *Journal of graduate medical education, 11*(2), 182.

Jenkins, T. M., Franklyn, G., Klugman, J., & Reddy, S. T. (2019). Separate but Equal? The Sorting of USMDs and Non-USMDs in Internal Medicine Residency Programs. *Journal of General Internal Medicine, 35*(5), 1458-1464.

Jolly, P., Boulet, J., Garrison, G., & Signer, M. M. (2011). Participation in US graduate medical education by graduates of international medical schools. *Academic Medicine, 86*(5), 559-564.

Kaffenberger, J. A., Mosser, J., Lee, G., Pootrakul, L., Harfmann, K., Fabbro, S., Faith, E. F., Carr, D., Plotner, A., Zirwas, M., & Kaffenberger, B. H. (2016). A retrospective analysis comparing the new standardized letter of recommendation in dermatology with the classic narrative letter of recommendation. *The Journal of clinical and aesthetic dermatology, 9*(9), 36.

Kahn, J. H., Tobin, R. M., Massey, A. E., & Anderson, J. A. (2007). Measuring emotional expression with the Linguistic Inquiry and Word Count. *The American journal of psychology, 120*(2), 263-286.

Kassam, A. F., Taylor, M., Cortez, A. R., Winer, L. K., & Quillin, III, R. C. (2021). Gender and ethnic diversity in academic general surgery department leadership. *The American Journal of Surgery, 221*(2), 363-368.

Kearse, L. E., Jensen, R. M., Schmiederer, I. S., Zeineddin, A., Anderson, T. N., Dent, D. L., Payne, D. H., & Korndorffer Jr, J. R. (2022). Diversity, equity, and inclusion: a current analysis of general surgery residency programs. *The American Surgeon, 88*(3), 414-418.

Keim, S. M., Rein, J. A., Chisholm, C., Dyne, P. L., Hendey, G. W., Jouriles, N. J., King, R. W., Schrading, W., Salomone, J., Swart, G., & Wightman, J. M. (1999). A standardized letter of recommendation for residency application. *Academic Emergency Medicine, 6*(11),

1141-1146.

Kenny, S., McInnes, M., & Singh, V. (2013). Associations between residency selection strategies and doctor performance: a meta-analysis. *Medical education*, *47*(8), 790-800.

Kettler, R. J. (2019). Correlational and Causal Comparative Designs in School Psychology. In *Research Methodologies of School Psychology* (pp. 112-131). Routledge.

Khelifa, R., & Mahdjoub, H. (2022). Integrate geographic scales in equity, diversity and inclusion. *Nature ecology & evolution, 6*(1), 4-5.

Kiuhara, S. A., & Huefner, D. S. (2008). Students with psychiatric disabilities in higher education settings: The Americans with Disabilities Act and beyond. *Journal of Disability Policy Studies, 19*(2), 103-113.

Klammer, R. M., Haydel, M. J., Gallahue, F., Bruno, E. C., Langdorf, M. I., Cheaito, M. A., Lotfipour, S., & Kazzi, A. (2019). Program visits and residency interviews. *The Journal of emergency medicine, 57*(4), e133-e139.

Kleinbaum, A. M., Stuart, T. E., & Tushman, M. L. (2013). Discretion within constraint: Homophily and structure in a formal organization. *Organization Science*, *24*(5), 1316-1336.

Kobayashi, A. N., Sterling, R. S., Tackett, S. A., Chee, B. W., Laporte, D. M., & Humbyrd, C. J. (2020). Are there gender-based differences in language in letters of recommendation to an orthopaedic surgery residency program?. *Clinical Orthopaedics and Related Research, 478*(7), 1400-1408.

Koivunen, T., Ylöstalo, H., & Otonkorpi-Lehtoranta, K. (2015). Informal practices of inequality in recruitment in Finland. *Nordic journal of working life studies, 5*(3), 3-21.

Kominsky, A. H., Bryson, P. C., Benninger, M. S., & Tierney, W. S. (2016). Variability of

ratings in the otolaryngology standardized letter of recommendation. *Otolaryngology–Head and Neck Surgery, 154*(2), 287-293.

Krzyzaniak & Lin. (2022, July 19). *EM Match advice 38: Our 2 cents | The revamped Standardized Letter of Evaluation (SLOE) is here.* ALiEM. https://www.aliem.com/em-match-advice-38-our-2-cents-sloe-standardized-letter-of-evaluation/

Kukulski, P., & Ahn, J. (2021). Validity Evidence for the Emergency Medicine Standardized Letter of Evaluation. *Journal of Graduate Medical Education, 13*(4), 490-499.

Laerd Statistics. (2018-a). *One-way ANCOVA in SPSS statistics. https://statistics.laerd.com/spss-tutorials/ancova-using-spss-statistics.php*

Laerd Statistics. (2018-b). *Two-way ANOVA in SPSS statistics. https://statistics.laerd.com/spss-tutorials/two-way-anova-using-spss-statistics.php*

Lakoff, J., Howse, K., Cofie, N., Heeneman, S., & Dalgarno, N. (2020). Analysis of factors affecting Canadian medical students' success in the residency match. *Canadian Medical Education Journal, 11*(3), e43.

Lee, A. G., Golnik, K. C., Oetting, T. A., Beaver, H. A., Boldt, H. C., Olson, R., Greenlee, E., Abramoff, M. D., Johnson, A. T., & Carter, K. (2008). Re-engineering the resident applicant selection process in ophthalmology: a literature review and recommendations for improvement. *Survey of ophthalmology, 53*(2), 164-176.

Lee, J. S., Ji, Y. D., Kushner, H., Kaban, L. B., & Peacock, Z. S. (2019). Residency interview experiences in oral and maxillofacial surgery differ by gender and affect residency ranking. *Journal of Oral and Maxillofacial Surgery, 77*(11), 2179-2195.

Leonard, D. K., & Jiang, J. (1999). Gender bias and the college predictions of the SATs: A cry of despair. *Research in Higher education, 40*(4), 375-407.

Leopold, S. S. (2020). Editor's Spotlight/Take 5: Are There Gender-based Differences in Language in Letters of Recommendation to an Orthopaedic Surgery Residency Program?. *Clinical Orthopaedics and Related Research, 478*(7), 1396-1399.

Leopold, S. S. (2021). A Conversation with… Alvin E. Roth PhD, Economist, Game Theorist, and Nobel Laureate Who Improved the Modern Residency Match. *Clinical Orthopaedics and Related Research, 479*(5), 863-866.

Li, S., Fant, A. L., McCarthy, D. M., Miller, D., Craig, J., & Kontrick, A. (2017). Gender differences in language of standardized letter of evaluation narratives for emergency medicine residency applicants. *AEM education and training, 1*(4), 334-339.

Lin, F., Oh, S. K., Gordon, L. K., Pineles, S. L., Rosenberg, J. B., & Tsui, I. (2019). Gender-based differences in letters of recommendation written for ophthalmology residency applicants. *BMC medical education*, *19*(1), 1-5.

Linguistic Inquiry and Word Count. (n.d.). *LIWC analysis. https://www.liwc.app/help/liwc*

Love, J. N., Doty, C. I., Smith, J. L., Deiorio, N. M., Jordan, J., Van Meter, M. W., Edens, M. A., & Hegarty, C. B. (2020). The emergency medicine group standardized letter of evaluation as a workplace-based assessment: the validity is in the detail. *Western Journal of Emergency Medicine, 21*(3), 600.

Love, J. N., Howell, J. M., Hegarty, C. B., McLaughlin, S. A., Coates, W. C., Hopson, L. R., Hern, G. H., Rosen, C. L., Fisher, J., & Santen, S. A. (2012). Factors that influence medical student selection of an emergency medicine residency program: implications for training programs. *Academic Emergency Medicine, 19*(4), 455-460.

Love, J. N., Smith, J., Weizberg, M., Doty, C. I., Garra, G., Avegno, J., Howell, J. M. (2014). Council of Emergency Medicine Residency Directors' standardized letter of

recommendation: the program director's perspective. *Academic Emergency Medicine, 21*(6), 680-687.

Luckey, J. W., Broughton, A., & Sorensen, J. E. (1982). Archival data in program evaluation and policy analysis. *Evaluation and Program Planning, 5*(4), 319-326.

MacFarland, T. W. (2011). *Two-way analysis of variance: statistical tests and graphics using R.* Springer Science & Business Media.

Madera, J. M., Hebl, M. R., & Martin, R. C. (2009). Gender and letters of recommendation for academia: agentic and communal differences. *Journal of Applied Psychology, 94*(6), 1591.

Madsen, T. E., Linden, J. A., Rounds, K., Hsieh, Y. H., Lopez, B. L., Boatright, D., Garg, N., Heron, S. L., Jameson, A., Kass, D., Lall, M. D., Melendez, A. M., Scheulen, J. J., Sethuraman, K. N., Westafer, L. M., & Safdar, B. (2017). Current status of gender and racial/ethnic disparities among academic emergency medicine physicians. *Academic Emergency Medicine, 24*(10), 1182-1192.

Maher, J. M., Markey, J. C., & Ebert-May, D. (2013). The other half of the story: effect size analysis in quantitative research. *CBE—Life Sciences Education, 12*(3), 345-351.

Mannix, A., Monteiro, S., Miller, D., Parsons, M., Alvarez, A. A., Krzyzaniak, S. M., Gore, K., Eraso, D., Davenport, D., Chan, T. M., & Gottlieb, M. (2022). Gender differences in emergency medicine standardized letters of evaluation. *AEM Education and Training, 6*(2), e10740.

Marcelin, J. R., Siraj, D. S., Victor, R., Kotadia, S., & Maldonado, Y. A. (2019). The impact of unconscious bias in healthcare: how to recognize and mitigate it. *The Journal of infectious diseases*, *220*(2), S62-S73.

Margolis, M. J., Clauser, B. E., Winward, M., & Dillon, G. F. (2010). Validity evidence for USMLE examination cut scores: results of a large-scale survey. *Academic Medicine, 85*(10), S93-S97.

Marshall, G., & Jonker, L. (2010). An introduction to descriptive statistics: A review and practical guide. *Radiography, 16*(4), e1-e7.

Martin, D. R., & McNamara, R. (2014). The CORD Standardized Letter of Evaluation: Have We Achieved Perfection or Just a Better Understanding of Our Limitations? *Journal of graduate medical education, 6*(2), 353–354.

Masood, A. (2019). Doing gender, modestly: Conceptualizing workplace experiences of Pakistani women doctors. *Gender, Work & Organization, 26*(2), 214-228.

Maxfield, C. M., Thorpe, M. P., Desser, T. S., Heitkamp, D. E., Hull, N. C., Johnson, K. S., Koontz, N. A., Mlady, G. W., Welch, T. J., & Grimm, L. J. (2019). Bias in radiology resident selection: do we discriminate against the obese and unattractive?. *Academic Medicine, 94*(11), 1774-1780.

McDade, W. A. (2019). Increasing graduate medical education diversity and inclusion. *Journal of graduate medical education, 11*(6), 736.

Meeks, L. M., Herzer, K., & Jain, N. R. (2018). Removing barriers and facilitating access: increasing the number of physicians with disabilities. *Academic Medicine, 93*(4), 540-543.

Meeks, L. M., Jain, N. R., Moreland, C., Taylor, N., Brookman, J. C., & Fitzsimons, M. (2019). Realizing a diverse and inclusive workforce: Equal access for residents with disabilities. *Journal of graduate medical education, 11*(5), 498.

Miani, C., Hinrichs, S., Pitchforth, E., Bienkowska-Gibbs, T., Disbeschl, S., Roland, M., &

Nolte, E. (2015). Best practice: medical training from an international perspective. *Rand health quarterly, 5*(1).

Miller, D. T., McCarthy, D. M., Fant, A. L., Li-Sauerwine, S., Ali, A., & Kontrick, A. V. (2019). The standardized letter of evaluation narrative: differences in language use by gender. *Western Journal of Emergency Medicine, 20*(6), 948.

Mitsouras, K., Dong, F., Safaoui, M. N., & Helf, S. C. (2019). Student academic performance factors affecting matching into first-choice residency and competitive specialties. *BMC medical education, 19*(1), 1-13.

Mohsin, M., & Syed, J. (2020). The missing doctors—An analysis of educated women and female domesticity in Pakistan. *Gender, Work & Organization, 27*(6), 1077-1102.

Morris, D. B., Gruppuso, P. A., McGee, H. A., Murillo, A. L., Grover, A., & Adashi, E. Y. (2021). Diversity of the National Medical Student Body—Four Decades of Inequities. *New England Journal of Medicine, 384*(17), 1661-1668.

Nagarajan, K. K., Bali, A., Malayala, S. V., & Adhikari, R. (2020). Prevalence of US-trained International Medical Graduates (IMG) physicians awaiting permanent residency: a quantitative analysis. *Journal of Community Hospital Internal Medicine Perspectives, 10*(6), 537-541.

Nahler, G., Brunier, D. P., Mollet, A., Nahler, M., & Szucs, T. D. (2017). *Dictionary of pharmaceutical medicine.* Springer.

Naidich, J. B., Grimaldi, G. M., Lombardi, P., Davis, L. P., & Naidich, J. J. (2014). A program director's guide to the medical student performance evaluation (former dean's letter) with a database. *Journal of the American College of Radiology, 11*(6), 611-615.

The National Resident Matching Program. (n.d.-a). *Match Participation Agreement for*

Applicants. https://mk0nrmp3oyqui6wqfm.kinstacdn.com/wp-

content/uploads/2021/09/2022-MPA-Main-Match-Applicant-2.pdf.

The National Resident Matching Program. (n.d.-b). *Match Registration: Applicants*.

https://www.nrmp.org/applicant-registration/

The National Resident Matching Program. (n.d.-c). *Residency Applicant Checklist & Match Tips.*

https://www.nrmp.org/residency-applicant-checklist-match-tips/

The National Resident Matching Program. (2016). *The Match Terms and Topics.*

https://www.nrmp.org/wp-content/uploads/2016/07/The-Match-Terms-and-Topics.pdf

The National Resident Matching Program. (2018). *Charting Outcomes in The Match* (2nd Ed.).

The National Resident Matching Program. (2020). *Results and Data: 2020 Main Residency*

Match. https://mk0nrmp3oyqui6wqfm.kinstacdn.com/wp-

content/uploads/2020/06/MM_Results_and-Data_2020-1.pdf

The National Residency Matching Program. (2021-a). *NRMP Delivers Strong Residency Match*

During Uncertain Times. https://mk0nrmp3oyqui6wqfm.kinstacdn.com/wp-

content/uploads/2021/03/2021-MRM-Press-Release_FINAL.pdf

The National Residency Matching Program. (2021-b). *Results and Data: 2021Main Residency*

Match. https://mk0nrmp3oyqui6wqfm.kinstacdn.com/wp-

content/uploads/2021/05/MRM-Results_and-Data_2021.pdf

Nwakuya, M. T., & Nwabueze, J. C. (2018). Application of box-cox transformation as a

corrective measure to heteroscedasticity using an economic data. *American Journal of*

Mathematics and statistics, 8(1), 8-12.

O'Neill, T. R., Peabody, M. R., & Song, H. (2016). The predictive validity of the National Board

of Osteopathic Medical Examiners' COMLEX-USA examinations with regard to

outcomes on American Board of Family Medicine examinations. *Academic Medicine, 91*(11), 1568-1575.

Osborn, M. B., Mattson, J., Yanuck, J., Anderson, C., Tekian, A., Fox, C. J., & Harris, I. B. (2016). Ranking practice variability in the medical student performance evaluation: so bad, it's "good". *Academic medicine: journal of the Association of American Medical Colleges, 91*(11), 1540.

Osborne, J. (2010). Improving your data transformations: Applying the Box-Cox transformation. *Practical Assessment, Research, and Evaluation, 15*(1), 12.

Ostapenko, L., Schonhardt-Bailey, C., Sublette, J. W., Smink, D. S., & Osman, N. Y. (2018). Textual analysis of general surgery residency personal statements: topics and gender differences. *Journal of surgical education, 75*(3), 573-581.

Pelletier-Bui, A., Franzen, D., Smith, L., Hopson, L., Lutfy-Clayton, L., Parekh, K., Olaf, M., Morrissey, T., Gordon, D., McDonough, E., Schnapp, B. H., Edens, M. A., & Kiemeney, M. (2020). COVID-19: A Driver for Disruptive Innovation of the Emergency Medicine Residency Application Process. *The western journal of emergency medicine, 21*(5), 1105–1113.

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015.*

Pheister, M., Peters, R. M., & Wrzosek, M. I. (2020). The impact of mental illness disclosure in applying for residency. *Academic Psychiatry, 44*, 554-561.

Phillips, J. P., Peterson, L. E., Fang, B., Kovar-Gough, I., & Phillips, R. L., Jr. (2019). Debt and the emerging physician workforce: The relationship between educational debt and family medicine residents' practice and fellowship intentions. *Academic Medicine, 94*(2), 267-

273.

Phitayakorn, R., Macklin, E. A., Goldsmith, J., & Weinstein, D. F. (2015). Applicants' self-reported priorities in selecting a residency program. *Journal of graduate medical education, 7*(1), 21.

Pietraszkiewicz, A., Formanowicz, M., Gustafsson Sendén, M., Boyd, R. L., Sikström, S., & Sczesny, S. (2019). The big two dictionaries: Capturing agency and communion in natural language. *European journal of social psychology, 49*(5), 871-887.

Polanco-Santana, J. C., Storino, A., Souza-Mota, L., Gangadharan, S. P., & Kent, T. S. (2021). Ethnic/racial bias in medical school performance evaluation of general surgery residency applicants. *Journal of Surgical Education*, *78*(5), 1524-1534.

Powers, A., Gerull, K. M., Rothman, R., Klein, S. A., Wright, R. W., & Dy, C. J. (2020). Race-and gender-based differences in descriptions of applicants in the letters of recommendation for orthopaedic surgery residency. *JBJS Open Access, 5*(3).

Ranstam, J. (2016). Multiple P-values and Bonferroni correction. *Osteoarthritis and cartilage, 24*(5), 763-764.

Razali, N. M., & Wah, Y. B. (2011). Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *Journal of statistical modeling and analytics, 2*(1), 21-33.

Reimann, S., & Alfermann, D. (2018). Female doctors in conflict: how gendering processes in German hospitals influence female physicians' careers. *Gender Issues, 35*(1), 52-70.

Risberg, G., Johansson, E. E., & Hamberg, K. (2009). A theoretical model for analysing gender bias in medicine. *International Journal for Equity in Health, 8*(1), 1-8.

Ross, D. A., Boatright, D., Nunez-Smith, M., Jordan, A., Chekroud, A., & Moore, E. Z. (2017).

Differences in words used to describe racial and gender groups in Medical Student Performance Evaluations. *PloS one, 12*(8), e0181659.

Roth, A. E. (2003). The origins, history, and design of the resident match. *JAMA, 289*(7), 909-912.

Roth, A. E., & Peranson, E. (1997). The effects of the change in the NRMP matching algorithm. *JAMA, 278*(9), 729-732.

Rutherford, A. (2011). *ANOVA and ANCOVA: a GLM approach.* John Wiley & Sons.

Ruzycki, S. M., Earp, M., & Ma, I. W. (2020). Applicant gender and matching to first-choice discipline: a cross-sectional analysis of data from the Canadian Resident Matching Service (2013–2019). *CMAJ open, 8*(2), E346.

Sandella, J. M., Gimpel, J. R., Smith, L. L., & Boulet, J. R. (2016). The use of COMLEX-USA and USMLE for residency applicant selection. *Journal of graduate medical education, 8*(3), 358.

Santen, S. A., Davis, K. R., Brady, D. W., & Hemphill, R. R. (2010). Potentially discriminatory questions during residency interviews: frequency and effects on residents' ranking of programs in the national resident matching program. *Journal of graduate medical education*, *2*(3), 336-340.

Santos, D., Palomares, N. B., Normando, D., & Quintao, C. C. (2010). Race versus ethnicity: Differing for better application. *Dental Press Journal of Orthodontics, 15*(3), 121-124.

Sapp, R. W., Sebok-Syer, S. S., Gisondi, M. A., Rotoli, J. M., Backster, A., & McClure Poffenberger, C. (2021). The Prevalence of Disability Health Training and Residents With Disabilities in Emergency Medicine Residency Programs. *AEM Education and Training, 5*(2), e10511.

Schenker, J. D., & Rumrill, P. D., Jr. (2004). Causal-comparative research designs. *Journal of vocational rehabilitation, 21*(3), 117-121.

Schmitt, A., Dyne, P. L., Broder, J., Cheaito, M. A., Harter, P. M., Mattu, A., Epter, M., & Kazzi, A. (2019). An overview of the allopathic match. *The Journal of Emergency Medicine, 56*(4), e61-e64.

Schwarz, C. M., & Zetkulic, M. (2019). You belong in the room: addressing the underrepresentation of physicians with physical disabilities. *Academic Medicine, 94*(1), 17-19.

Sedgwick, P. (2013). Convenience sampling. *BMJ*, *347*, f6304.

Shah, R., & Ahluwalia, S. (2019). The challenges of understanding differential attainment in postgraduate medical education. *British Journal of General Practice*, 69(686), 426-427.

Sharma, M. (2019). Applying feminist theory to medical education. *The Lancet*, *393*(10171), 570-578.

Shen, M. J., Peterson, E. B., Costas-Muñiz, R., Hernandez, M. H., Jewell, S. T., Matsoukas, K., & Bylund, C. L. (2018). The effects of race and racial concordance on patient-physician communication: a systematic review of the literature. *Journal of Racial and Ethnic Health Disparities, 5*(1), 117-140.

Simmenroth-Nayda, A., & Görlich, Y. (2015). Medical school admission test: advantages for students whose parents are medical doctors?. *BMC medical education, 15*(1), 1-6.

Simons, E. C. G., Saigal, C., & Downs, T. (2021). Diversifying Graduate Medical Education & the Urology Workforce: Re-imagining our Structures, Policies, Practices, Norms & Values. *Urology, 00*, 1-9.

Spector, A. R., & Railey, K. M. (2019). Reducing reliance on test scores reduces racial bias in

neurology residency recruitment. *Journal of the National Medical Association, 111*(5), 471-474.

Spottswood, S. E., Spalluto, L. B., Washington, E. R., Donnelly, E. F., Birch, A. A., Bradshaw, M. L., & Omary, R. A. (2019). Design, implementation, and evaluation of a diversity program for radiology. *Journal of the American College of Radiology, 16*(7), 983-991.

Stimmel, B., Katz, L., Brownstein, E., Osteen, A., & Smith, H. (1981). Fifth pathway programs in American medical schools. *Bulletin of the New York Academy of Medicine, 57*(2), 149.

Swide, C., Lasater, K., & Dillman, D. (2009). Perceived predictive value of the medical student performance evaluation (MSPE) in anesthesiology resident selection. *Journal of clinical anesthesia, 21*(1), 38-43.

Takeshita, J., Wang, S., Loren, A. W., Mitra, N., Shults, J., Shin, D. B., & Sawinski, D. L. (2020). Association of Racial/Ethnic and Gender Concordance Between Patients and Physicians With Patient Experience Ratings. *JAMA network open, 3*(11), e2024583-e2024583.

Talamantes, E., Henderson, M. C., Fancher, T. L., & Mullan, F. (2019). Closing the gap—making medical school admissions more equitable. *New England Journal of Medicine, 380*(9), 803-805.

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology, 29*(1), 24-54.

Teherani, A., Hauer, K. E., Fernandez, A., King, T. E., Jr., & Lucey, C. (2018). How small differences in assessed clinical performance amplify to large differences in grades and awards: a cascade with serious consequences for students underrepresented in medicine.

Academic Medicine, 93(9), 1286-1292.

ten Cate, O., & Smal, K. (2002). Educational assessment center techniques for entrance selection in medical school. *Academic Medicine, 77*(7), 737.

Turrentine, F. E., Dreisbach, C. N., St Ivany, A. R., Hanks, J. B., & Schroen, A. T. (2019). Influence of gender on surgical residency applicants' recommendation letters. *Journal of the American College of Surgeons*, *228*(4), 356-365.

Urlings-Strop, L. C., Stijnen, T., Themmen, A. P., & Splinter, T. A. (2009). Selection of medical students: a controlled experiment. *Medical Education, 43*(2), 175-183.

VanderWeele, T. J., & Mathur, M. B. (2019). Some desirable properties of the Bonferroni correction: is the Bonferroni correction really so bad?. *American journal of epidemiology, 188*(3), 617-618.

Verdonk, P., Benschop, Y. W., De Haes, H. C., & Lagro-Janssen, T. L. (2009). From gender bias to gender awareness in medical education. *Advances in health sciences education, 14*(1), 135-152.

Vetter, T. R. (2017). Descriptive statistics: Reporting the answers to the 5 basic questions of who, what, why, when, where, and a sixth, so what?. *Anesthesia & Analgesia, 125*(5), 1797-1802.

Vickerstaff, V., Omar, R. Z., & Ambler, G. (2019). Methods to adjust for multiple comparisons in the analysis and sample size calculation of randomised controlled trials with multiple primary outcomes. *BMC medical research methodology, 19*(1), 1-13.

Wetz, R. V., Seelig, C. B., Khoueiry, G., & Weiserbs, K. F. (2010). Out-of-match residency offers: the possible extent and implications of prematching in graduate medical education. *Journal of Graduate Medical Education, 2*(3), 327.

Whitehead, A. L. (2013). Gendered organizations and inequality regimes: Gender, homosexuality, and inequality within religious congregations. *Journal for the Scientific Study of Religion, 52*(3), 476-493.

Wilkinson, D., Casey, M. G., & Eley, D. S. (2014). Removing the interview for medical school selection is associated with gender bias among enrolled students. *Medical Journal of Australia, 200*(2), 96-99.

Wilson, D., Laoteppitaks, C., & Chandra, S. (2021). A comparison of Standardized Letters of Evaluation for emergency medicine residency applicants. *Western Journal of Emergency Medicine, 22*(1), 20.

Woolf, K., Potts, H. W., & McManus, I. C. (2011). Ethnicity and academic performance in UK trained doctors and medical students: systematic review and meta-analysis. *BMJ*, 342.

Zavlin, D., Jubbal, K. T., Noé, J. G., & Gansbacher, B. (2017). A comparison of medical education in Germany and the United States: from applying to medical school to the beginnings of residency. *GMS German Medical Science*, 15. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5617919/

Zeidan, A. J., Khatri, U. G., Aysola, J., Shofer, F. S., Mamtani, M., Scott, K. R., Conlon, L. W., & Lopez, B. L. (2019). Implicit bias education and emergency medicine training: step one? awareness. *AEM education and training*, *3*(1), 81-85.

Zhang, X., Lin, D., Pforsich, H., & Lin, V. W. (2020). Physician workforce in the United States of America: forecasting nationwide shortages. *Human resources for health, 18*(1), 1-9.

APPENDIX A: Electronic Standardized Letter of Evaluation

APPENDIX B: Institutional Review Board Approval

LIBERTY UNIVERSITY.
INSTITUTIONAL REVIEW BOARD

May 1, 2023

Steven Warrington
Eric Lovik

Re: IRB Approval - IRB-FY22-23-776 GENDER AND ETHNICITY DIFFERENCES IN EMERGENCY MEDICINE RESIDENCY APPLICATIONS

Dear Steven Warrington, Eric Lovik,

We are pleased to inform you that your study has been approved by the Liberty University Institutional Review Board (IRB). This approval is extended to you for one year from the following date: May 1, 2023. If you need to make changes to the methodology as it pertains to human subjects, you must submit a modification to the IRB. Modifications can be completed through your Cayuse IRB account.

Your study falls under the expedited review category (45 CFR 46.110), which is applicable to specific, minimal risk studies and minor changes to approved studies for the following reason(s):

5. Research involving materials (data, documents, records, or specimens) that have been collected, or will be collected solely for nonresearch purposes (such as medical treatment or diagnosis).

Thank you for your cooperation with the IRB, and we wish you well with your research project.

Sincerely,

G. Michele Baker, PhD, CIP
Administrative Chair
Research Ethics Office