ABOUT PSI CHI
Psi Chi is the International Honor Society in Psychology, founded in 1929. Its mission: “recognizing and promoting excellence in the science and application of psychology.” (Note: Our new mission statement is available at http://www.psichi.org/?page=purpose) Membership is open to undergraduates, graduate students, faculty, and alumni making the study of psychology one of their major interests and who meet Psi Chi’s minimum qualifications. Psi Chi is a member of the Association of College Honor Societies (ACHS), and is an affiliate of the American Psychological Association (APA) and the Association for Psychological Science (APS). Psi Chi’s sister honor society is Psi Beta, the national honor society in psychology for community and junior colleges.

Psi Chi functions as a federation of chapters located at over 1,100 senior colleges and universities around the world. The Psi Chi Central Office is located in Chattanooga, Tennessee. A Board of Directors, composed of psychology faculty who are Psi Chi members and who are elected by the chapters, guides the affairs of the Organization and sets policy with the approval of the chapters.

Psi Chi membership provides two major opportunities. The first of these is academic recognition to all inductees by the mere fact of membership. The second is the opportunity of each of the Society’s local chapters to nourish and stimulate the professional growth of all members through fellowship and activities designed to augment and enhance the regular curriculum. In addition, the Organization provides programs to help achieve these goals including conventions, research awards and grants competitions, and publication opportunities.

JOURNAL PURPOSE STATEMENT
The twofold purpose of the Psi Chi Journal of Psychological Research is to foster and reward the scholarly efforts of psychology students as well as to provide them with a valuable learning experience. The articles published in the Journal represent the work of undergraduates, graduate students, and faculty. To further support authors and enhance journal visibility, articles are now available in the PsycINFO® and EBSCO Academic Search Complete® databases. In 2016, the Journal also became open access (i.e., free online to all readers and authors) to broaden the dissemination of research across the psychological science community.

JOURNAL INFORMATION
The Psi Chi Journal of Psychological Research (ISSN 2325-7342) is published quarterly in one volume per year by Psi Chi, Inc., The International Honor Society in Psychology.

For more information, contact Psi Chi Central Office, Publication and Subscriptions, P.O. Box 709, Chattanooga, TN 37401-0709, (423) 756-2044. www.psichi.org; psichijournal@psichi.org.

Statements of fact or opinion are the responsibility of the authors alone and do not imply an opinion on the part of the officers or members of Psi Chi.

Advertisements that appear in Psi Chi Journal do not represent endorsement by Psi Chi of the advertiser or the product. Psi Chi does not endorse or is responsible for the content of third-party promotions. Learn about advertising with Psi Chi at http://www.psichi.org/?page=Advertise

Permission must be obtained from Psi Chi to reprint or adapt a table or figure; to reprint quotations exceeding the limits of fair use from one source, and/or to reprint any portion of poetry, prose, or song lyrics. All persons wishing to utilize any of the above materials must write to the publisher to request nonexclusive world rights in all languages to use copyrighted material in the present article and in future print and nonprint editions. All persons wishing to utilize any of the above materials are responsible for obtaining proper permission from copyright owners and are liable for any and all licensing fees required. All persons wishing to utilize any of the above materials must include copies of all permissions and credit lines with the article submission.
76  Institutional Review Board: Ally Not Adversary
Melanie M. Domenech Rodríguez, Samantha M. Corralejo, Nicole Vouvalis
Utah State University
Alan K. Mirly
Idaho State University

85  Who’s Laughing Now? The Effects of Sexist and Rape Humor
Nalyn Sriwattanakomen
Washington & Jefferson College

98  Exploring Sexual Self-Concept in a Community-Based Sample
    of LGB and Heterosexual Adults
Dustin K. Shepler*, Kevin P. Johnson, and Melanie A. Ho
Michigan School of Professional Psychology

106 Examining Personality Factors in Deception Detection Ability
Samuel D. Spencer
Michigan State University, Mankato

114 The Relationship Between Religiousness and Friendship Quality
Namele Gutierrez, Elizabeth J. Krumrei-Mancuso*, and Cindy Miller-Perrin*
Pepperdine University

121 Romantic Satisfaction in Young Adults as a Function of Sexual Debut
Jaclyn M. Vancour and Marianne Fallon*
Central Connecticut State University

131 The Effects of Parental Support and Self-Esteem
    on Internalizing Symptoms in Emerging Adulthood
Lewie E. Moore II* and Madelynn D. Shell*
The University of Virginia’s College at Wise

141 Lyrical Memory:
    Mnemonic Effects of Music for Musicians and Nonmusicians
Sarah N. Lummis, Jennifer A. McCabe*, Abigail L. Sickles, Rebecca A. Byler,
Sarah A. Hochberg, Sarah E. Eckart, and Corinne E. Kahler
Goucher College

*Faculty mentor
Institutional Review Boards (IRBs) were established in the United States with the passage of the National Research Act of 1974. The National Research Act led to the creation of the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research (NCPHSBBR). The Commission was charged with identifying the basic ethical principles to be upheld when researchers conduct biomedical and behavioral research involving human subjects. In addition, the commission was tasked with the development of guidelines to carry out research in accordance with these basic ethical principles. The basic ethical principles are outlined in a document known as the Belmont Report (NCPHSBBR, 1979). The Belmont Report advances three principles—respect for persons, beneficence, and justice—that should guide researchers in the preparation and execution of their research projects. The observable manifestations of these principles are seen in current-day practices for obtaining informed consent, providing a thoughtful and thorough risk-benefit assessment, and providing a rationale for the selection of research participants.

In the 1970s, the National Research Act was ground-breaking and provided a much-needed response to chronic and abusive behavior on the part of government-sponsored researchers. However, some might argue that the United States was slow to enact these policies. The Nuremberg Code (Germany) was released in 1947 after the end of World War II. In 1964, the World Medical Association adopted the Declaration of Helsinki. The United States waited another 10 years before enacting comprehensive federal legislation. In the ensuing five decades, IRBs have been working to protect the rights of human participants in research activities across the nation. Unfortunately, IRBs are often perceived as barriers to research and cumbersome foes in the quest for research productivity.

The purpose of this editorial is to provide an alternative framework for understanding IRBs. The authors of this article are all scholars who also officially participate in the IRB. This places us in a unique position to bridge the gap between researchers and IRBs. In this editorial, we advance the notion that IRBs serve as close colleagues in the research enterprise, providing methodological and...
ethical support, and doing so in a timely manner and free of charge. IRBs were founded to preserve and promote human rights, and to protect individuals from harm while volunteering their time and talents to advance scientific knowledge. Yet their functioning has been criticized; we provide guidance on which of these criticisms are well-founded and which may have strong ties to the human rights foundation. We make recommendations on best practices in collaborating with local IRBs so as to establish and/or maintain a strong collaborative relationship with this “ally not adversary.”

The Evolution of the Protection of Fundamental Human Rights

The existence of IRBs is due directly to human rights violations under the guise of research and perpetrated by professionals using the tools and methods of research. Modern-day IRBs were established by the National Research Act of 1974. However, there is evidence of ethical oversight for federally funded research in the United States as early as 1966 (Ghooi, 2014).

The Tuskegee Syphilis Study (1932–1972) by the United States Public Health Service is credited with being the catalyst that led to the National Research Act, yet there had been ample violations of basic human rights in the name of research prior to Tuskegee. Accounts of atrocities in the United States are documented in various accounts (Skloot, 2010; Washington, 2008). Internationally, during World War II, experiments on altitude, freezing, sterilization, and immunizations, among others, were carried out using brutal and unnecessary methods. Findings were either later replicated with greater validity while using humane methods or the research had already been carried out with animals and provided valid results (Moe, 1984). The abuses during the war led to the Nuremberg Military Tribunals in 1946–47 (Nuremberg Military Tribunals, 1949). Shortly after the tribunals, the Universal Declaration of Human Rights (1948) was published.

The public record of the Nuremberg Military Tribunals (1949) included 10 basic principles for researchers to follow during experimentation, namely, (a) voluntary consent is “absolutely essential,” (b) results of research must be for the greater good of society and the knowledge not attainable through other means, (c) medical experiments should be based on prior research with animals, (d) the research should seek to avoid physical and mental suffering as well as injury, (e) experiments are not justified that may cause death or disability, (f) risks should not exceed benefits, (g) facilities should be prepared so as to protect human participants, (h) research should be carried out only by qualified persons, (i) participants may stop participation at any time, and (j) researchers use their judgment and are prepared to terminate the experiment when injury, disability, or death is likely to occur.

These basic principles are fundamental to other codes of ethics for researchers. For example, the Declaration of Helsinki, initially adopted in 1964 by the World Medical Association (WMA, 2013) General Assembly, covers all of these points either explicitly or implicitly and adds important details that are clearly tied to present-day IRB protocol structures. For example, the Declaration of Helsinki covers compensation of participants, vulnerable participants, formal protocols and review boards, adverse events, privacy and confidentiality, consequences for participants for choosing to withdraw from research, provisions for what occurs if participants are harmed, ongoing safety and monitoring, and appropriate dissemination of results (WMA, 2013).

The Declaration is unique from the Nuremberg codes in its international scope and clear articulation of expectations for researchers at a global scale. Specifically, the Declaration notes that protocols must take into account laws, regulations, norms, and standards, of the country in which the research is carried out. The Declaration of Helsinki also reaches beyond humans to specify that the environment must be protected from harm in the course of medical research.

The Belmont Report (NCPHSBBR, 1979) adds an important dimension to the consideration of human participants in research by explicitly adding behavioral research to the scope of attention for the federal regulations. The Belmont Report is a statement of policy of the Department of Health and Human Services. It specifies three principles that must be examined in making decisions regarding the ethicality of human subjects research: (a) beneficence, (b) respect for persons, and (c) justice.

An important contribution of the Belmont Report was to provide a clear, operational definition of research:

“Research” designates an activity designed to test a hypothesis, permit conclusions to be drawn, and thereby to develop or
IRB Editorial | Domenech Rodríguez, Corralejo, Vouvalis, and Mirly

contribute to generalizable knowledge (expressed, for example, in theories, principles, and statements of relationships). Research is usually described in a formal protocol that sets forth an objective and a set of procedures designed to reach that objective. (United States Department of Health and Human Services [USDHHS], 2009, para. 6)

The specific laws that were developed to address the guidance of the Belmont Report are codified in 45 C.F.R. 46 (USDHHS, 2009). The law follows, and extends, the Belmont Report definition of research:

Research means a systematic investigation, including research development, testing and evaluation, designed to develop or contribute to generalizable knowledge. Activities which meet this definition constitute research for purposes of this policy, whether or not they are conducted or supported under a program which is considered research for other purposes. For example, some demonstration and service programs may include research activities. (USDHHS, 2009, 45 C.F.R. 46.102(d))

These laws further specify the definition of human subject as “a living individual about whom an investigator (whether professional or student) conducting research obtains (1) Data through intervention or interaction with the individual, or (2) Identifiable private information” (45 C.F.R. 46.102(f)).

In sum, the current protections for human participants in research evolved from clear violations to fundamental human rights in the United States and abroad. Lawyers, judges, scholars, ethicists, and many others around the world convened to consider how to move forward. The varied yet converging guidelines from around the world suggest there are core human values that must be prioritized and put into practice in research.

Modern Day IRBs

Modern day IRBs function as independent bodies for research review and oversight. Many IRBs are affiliated with universities, medical centers, and research centers. However, they are not under the direct oversight of those institutions but rather are affiliated. This independence is critical to avoid conflicts of interest with institutional demands. A conflict of interest resulting from the interests of those within an institution can arise, for example, when a granting agency requires IRB approval prior to releasing funds for research. Institutions are often interested in receiving these funds, and there may be temptation to provide rote approvals in order to secure funds. This is why many institutions do not set their IRB leadership up to report to the individual overseeing sponsored funding for the institution. These conflicts of interest were foreseen in the Declaration of Helsinki (WMA, 2013). The declaration clearly states that research must be reviewed by an ethics committee that “must be transparent in its functioning, must be independent of the researcher, the sponsor, and any other undue influence” (para. 23).

A growing number of independent IRBs provide human research review and oversight to researchers who are not affiliated with any institution (USDHHS, 1998). Thousands of IRBs are registered with USDHHS nationally and internationally. A full list of IRBs can be found online (http://ohrp.cit.nih.gov/search/search.aspx).

Types of Reviews

IRBs are tasked with making the determination for whether or not a protocol review is required and, if so, which level of review a protocol should receive. An IRB may decline to review a protocol when the proposed project does not meet the definition of research articulated above, and/or its activities do not pertain to humans. If a protocol presents human subjects research, then it may be reviewed under one of three main categories: (a) exempt, (b) expedited, and (c) full board. An IRB determines that research is exempt from IRB review when, generally, it involves research on common instructional practices, anonymous surveys or interviews, passive observation of public behavior without collection of identifiers, and retrospective record reviews, or secondary data analyses without recorded identifiers. Readers are encouraged to consult 45 C.F.R. 46.101 for a detailed list of exemptions.

Initial expedited reviews must present no more than minimal risk to human participants in research and must also fall in one of six categories (Office of Human Research Protections [OHRP], 2016). These categories include limited clinical studies involving drugs and medical devices, collection of blood samples, noninvasive collection of blood specimens, noninvasive collection of data routinely collected in clinical practice, research involving materials that were collected for nonresearch purposes (e.g., records, specimens), and...
some audio or video recordings. The final category for expedited consideration is where much sociobehavioral research falls, specifically:

Research on individual or group characteristics or behavior (including, but not limited to, research on perception, cognition, motivation, identity, language, communication, cultural beliefs or practices, and social behavior) or research employing survey, interview, oral history, focus group, program evaluation, human factors evaluation, or quality assurance methodologies. (OHRP, 2016, para. 7)

Finally, full board reviews are conducted for research considered to carry some increase over minimal risk, or which are not described in the allowable expedite categories. The federal regulations define minimal risk as: “The probability and magnitude of harm or discomfort anticipated in the research are not greater in and of themselves than those ordinarily encountered in daily life or during the performance of routine physical or psychological examinations or tests” (45 C.F.R. 46.102i). The determination of what constitutes an increase over minimal risk is a judgment that is evaluated by the IRB. Once a study has been determined to surpass minimal risk, it is forwarded to the full IRB for review and discussion/determination at the next convened meeting. Some characteristics of research automatically lead to full board review; for example, studies involving vulnerable populations such as prisoners. Although not automatic, studies that involve the collection of sensitive data that could lead to reporting (e.g., child abuse), that use deception, or are carried out in international settings without a clear local authority to oversee the research activities, may receive full board review.

Composition of the Board
IRBs are required by federal regulation to have a minimum of five members (45 C.F.R. 46.107) who are qualified to review research, have varied expertise, and provide demographic diversity. Each IRB must have at least one scientific member, one nonscientific member, and at least one community member. IRBs vary in size, number, and relative composition of members. At our institution, there is only one IRB, which has 11 members, most of whom are scientific scholars. We have one community member and one prisoner advocate. Composition of IRBs varies outside of the United States as do the regulations for such composition. For example, in India, IRB chairs cannot belong to the same institution with which the IRB is affiliated (Ghooi, 2014)

Researcher Training
In the summer of 2000, the National Institutes of Health (NIH, 2000) released a notification of the requirement for researchers to be formally trained in human subjects research protections. It is now commonplace for institutions to have this requirement for researchers, thus nonfederally funded researchers are often held to the same standard. Since then, certification programs have been developed and are in common use such as the Collaborative IRB Training Initiative known as CITI and NIH’s Protecting Human Research Participants course. However, some institutions use their own home-grown training. The federal regulation does not specify a program for training or even a set frequency for the renewal of certifications.

Accreditation
IRBs function independently. The freedom from institutional oversight in decision-making protects against conflicts of interest. However, the independence also begs the question “who oversees the IRB?” Accreditation is an option for IRBs who want an independent assessment of their functioning and ongoing plan for improvement. The major accrediting body for IRBs in the United States is the Association for the Accreditation of Human Research Protection Programs (AAHRPP). To date, there are 229 IRBs accredited by AAHRPP.

Criticisms of IRBs
A wide variety of criticisms exist regarding the IRB. Some are more basic concerns of efficiency and timeliness. Other concerns focus on data-driven concerns such as inconsistencies in function and judgments. Yet others express frustrations with IRBs’ overreach or intrusive practices.

Efficiency and Timelines
Delays in feedback or approval can occur even on exempt protocols when IRBs become overwhelmed with reviews (Dziak et al., 2005). When protocols require full board review, this can further lengthen the research process, which may pose challenges for graduate students or junior tenure-track faculty who are working on tight timelines and for whom research productivity is tied to formal evaluations that can have significant outcomes on progress or employment. These issues are real and quite
Disparate Judgments

When researchers develop their studies and an IRB reviews them, both parties are conducting a risk-benefit assessment. One concern researchers have expressed is the disproportionate weight given to immediate risks and benefits compared to long-term benefits (Ceci & Bruck, 2009; Saxton et al., 2015). Others highlight the difference between individual autonomy and protection and community ethics (Malone, Yerger, McGruder, & Froelicher, 2006). The criticism is similar for both accounts: IRBs focus too much of their concern on what is right in front of them and fail to recognize the importance of the greater good (i.e., individual participants and immediate risks or benefits to the individual versus eventual benefit for the community). In any work requiring judgment, there will be variability in outcomes for decisions. Some of this should be expected by researchers and not necessarily catalogued as problematic or dysfunctional. Research could be carried out to examine the kinds of protocols, methods, or other research conditions that lead to the greatest variability in decisional outcomes. We suspect that some of the variability is also due to the differences in composition of IRB members. For example, on a board where there are more community and/or nonscientific members, there may be systematic differences in outcomes when compared to a board that has mostly scientific members. Future research may also gather information about the relationship between board composition and different outcomes.

Inconsistent Practices

IRBs are critiqued for the inevitable diversity of structures, practices, and outcomes of IRBs across the country. Critics point out that conclusions drawn are often inconsistent both across reviewers and across IRBs. For example, Dziak et al. (2005) submitted the same protocol to 15 different IRBs across the country and, while all protocols were eventually approved, the type of review, form of notification and recruitment, and time to approval varied. Some have complained that different protocols submitted to the same IRB may be inconsistently reviewed because of conflicts of interest (Keith-Spiegel & Koocher, 2005).

Recommendations for how the IRB can address criticisms have been offered, some of which are already being addressed with new regulations. Specific recommendations include better recruitment and training of staff and faculty members, educating researchers, improving turnaround times, nonbiased reviews, and proscience sensitivity (Fiske, 2009; Keith-Spiegel, Koocher, & Tabachnick, 2006). Keith-Spiegel and Koocher (2005) pointed out that, if researchers feel that they are not receiving fair reviews, they may begin to submit dishonest protocols that will be more easily approved. Researchers may begin to describe their studies inaccurately, omit important information, or forego the IRB process altogether (Keith-Spiegel & Koocher, 2005).

For example, a professor may ask students to participate in a classroom activity under the guise of an evaluation related to the course (e.g., a comparison of students’ exam performance when they do or do not use a study guide), but intend from the outset to contribute generalizable knowledge (e.g., evident by a stated intention and/or by the use of randomization to guide/no guide conditions). The professor may justify bypassing the IRB because “the semester will be over before IRB approves the project” or “it’s not really research.” Ironically, a professor using her students to advance her research agenda may be highly problematic because (a) her activities constitute research that requires oversight and (b) her position as professor of the course places students in a heightened position to be coerced into participation. In addition, (c) there is a conflict of interest that could threaten the integrity of the findings. There could be an additional problem (d) related to the professor having access to FERPA-protected information without student consent.

Scope of Practice

Some critics claim mission creep is occurring when IRBs take issue with and/or responsibility for aspects of research studies that are outside of their bounds (e.g., research methodology, university policies; Fiske, 2009; Malone et al., 2006). These disagreements may be simply differences in judgment. For example, a researcher may consider that the use of an idiosyncratic Spanish translation of a survey is appropriate to answer her research questions. An IRB member with expertise in working...
with Latina/o populations may note that the reliability and validity of the instrument has not been established, and that poorly conceived research methods and tools may threaten the enactment of the principles of respect for persons and justice. Other IRB members may feel comfortable trusting the researcher’s expertise to carry out the work, claiming that the knowledge gained may benefit the marginalized community. Both perspectives may be correct. The work of the committee is to arrive at a sufficient consensus to make a determination. Future research could examine archival data from IRBs on various determinations to help arrive at a consensus of mission creep and examine its prevalence.

In addition, the scope of practice of the IRBs is often questioned. For example, the federal regulations characterize research as both generalizable and systematic. Some researchers have questioned the need for review of qualitative or social science research claiming that it is not systematic and/or intended to be generalizable. Additional criticism notes that the historical context centers on human rights concerns in biomedical research (Ceci & Bruck, 2009) and thus should not apply to behavioral research. Yet other researchers have outlined challenges in providing an a prior systematic account for board review as is the case in qualitative methods that require the flexibility to adapt to findings as they are occurring. These are reasonable observations that suggest that an examination of procedures may be warranted. In our experience, some qualitative researchers have questioned the need for IRB oversight. When offered the option to discuss whether their activities are intended to be research, there has consistently been an adamant and decisive “yes” response. Thus, where researchers intend to do research and agree basic human rights are worthy of protection, it seems logical to turn attention to the suitability of existing procedures to meet the needs of the researcher population. Revisions based on recommendations from qualitative experts, for example, could provide a wonderful base from which to create a meaningful process and materials for qualitative research.

In essence, some of the main criticism are warranted. There are existing systems that may help address some of the valid criticisms (e.g., accreditation). There are also opportunities to create better and more relevant systems for researchers who find that the current IRB doesn’t “fit” for them. Finally, we believe much data are needed to better understand which challenges are tied to research (e.g., quality of the proposals submitted) and which are tied to the IRB process (e.g., differential prioritization of human subjects protections). New regulations on the horizon also will address some of these issues in the near future. We turn out attention to them.

New Regulations

The long-awaited updates to the federal Common Rule (effective beginning January 2018) have been designed to alleviate some of the aforementioned concerns. For example, many updates throughout the new Common Rule were “designed to more thoroughly address the broader types of research … such as behavioral and social science research.” (USDHHS, 2017, para. 5). This includes the expansion of exemption categories, which will undoubtedly result in more qualitative and social science work being reviewed as exempt, rather than expedited. Social, educational, and behavioral researchers in particular should see a decrease in the turnaround time on their protocols as a result of the shift in allowable exemptions. We expect that biomedical and clinical investigators will see decreased turnaround times as a result of the lighter load on IRB members who typically work as volunteers, but who are responsible in many cases for conducting the expedited reviews.

Another piece of the new regulations that should lighten the burden of IRB review is the elimination of annual (or more frequent) continuation review for most studies. The updated Common Rule provides for the elimination of annual continuation review for studies reviewed via the expedited procedure that pose minimal risk. The elimination of this responsibility should free up the time of IRB staff to allow initial reviews of protocols to move along at a more expedient rate.

Inconsistent practices between institutions and IRBs is another issue that the updated Common Rule seeks to address. Although the rest of the Common Rule will be effective in January 2018, beginning in 2019, all studies funded by the agencies adopting the Common Rule will require single-IRB review. The single-IRB review means that one IRB will be the responsible entity for oversight of research at multiple sites. The NIH plan to adopt this requirement in September of 2017, and it is expected that many IRBs will come into compliance along with NIH’s adoption of the requirement. The requirement that a study be reviewed by a single IRB, with some exceptions, should ensure more
consistent timelines and reviews between members of cooperative research projects. Although doing this effectively may be a burden on IRBs, we expect that the benefits of single-IRB review will accrue to the research teams and, hopefully, to their participants.

Collaborating With IRBs

We offer our recommendations based on our knowledge of the research integrity literature, our own experiences collaborating with colleagues broadly and in research activities specifically, and as members of the IRB in different roles (Chair, Graduate Assistant, Director, Prisoner Advocate). These are not intended to imply to any potential problems with a local IRB reside solely outside of the IRB, nor that solutions are only found outside of the IRB. Rather, we share some pointers for common points of miscommunication that we have experienced.

Do Some Homework

We recommend approaching the human protections activities with the same rigor with which researchers engage the preparation of a manuscript: consult original sources and seek data. There are many points of misinformation that often are the result of hearsay, conflicting experiences at other institutions, or even differential practices across departments/units within the same institution. Researchers would benefit from reviewing original documents that provide a historical perspective and current expectations: 45 C.F.R. 46, the Belmont Report, and your IRB’s Standard Operating Procedures. Along with these, we recommend having handy the code of ethics of your professional association, the regulations of the researcher’s institution, and regulations of any funding agencies. On the data end, researchers could ask for annual reports documenting key outcomes such as speed of determinations to verify whether their experience signals a chronic problem or a unique situation.

Check Your Mindset

When collaborating with colleagues, we recommend that researchers prepare their mindset. For example, researchers may want to assume expertise and expect collegiality from the IRB staff and members. Many research administrators are highly trained in the regulatory and ethical landscape surrounding human subjects research. When researchers receive requested changes or clarifications that are confusing or seem unnecessary, we recommend picking up the phone and asking for clarification from a staff person at the IRB. Also, remember that, while the electronic protocol systems many institutions use can seem terse or even punishing, the people behind that system are often trying to make a point as quickly as possible for your ease. Language or guidance that may appear “short” may have a well-developed rationale that, if carefully explicated, would create more burden or clutter in the protocol review system.

We also recommend assuming good will on the part of the IRB. In our experience, a small but vocal subset of researchers respond to queries for further information or clarification with defensiveness, as if being summarily charged with unethical conduct. We are often surprised by these responses. IRB staff and board members are tasked with protecting human participants in research. This is by definition, a prevention activity. Obtaining clarification or requesting further materials serve to verify that protections are in place. If researchers assume that IRBs are collaborators in the research enterprise, then these requests can be seen as prompts rather than as admonitions. We cannot assume that all IRB staff and board members have good intentions or helpful motivations. We do know our own motivations and intentions, and we have experienced the difference when working with colleagues who assume good will and those who do not.

Use Existing Resources

Using the structure of the IRB protocol template when developing a research project can serve as an excellent outline for the method section of a future manuscript. Having an empty IRB protocol handy as researchers build their proposal can ensure that all human protections issues are addressed at the outset of a study. In our experience, it is not unusual for a student to engage in research that has to be approved by a committee. When the IRB then detects a potential threat to the protection of human participants such as insufficient justification for sample size, we find that students and their primary advisors respond with frustration. They might have secured money or arrived at a committee consensus for sample size, but the decision was not based on a scientific process (e.g., calculation of power using statistical software). Checking IRB protocols ahead of time can save time and hassle, and also ensure a more scientifically rigorous process from the outset. In our experience, an additional benefit may be that this thoroughness
provides an excellent spring board for publication.

Another important resource that is often overlooked is the members of the IRB. Many departments with active human subjects research portfolios have faculty sitting on the IRB. If researchers do not feel comfortable approaching an IRB staff person with questions or concerns about the human research review process, students or faculty can check in directly with an IRB member! Additionally, researchers who have less experience with protocol submission can request recommendations for securing a peer-mentor who is not on the board.

Consider Timelines
Preferably well in advance of submitting a protocol, researchers become familiar and prepare for a future protocol submission. New researchers forget that they have to complete formal training in human protections, and seasoned researchers forget to check deadlines for their recertifications. This training can take a substantial amount of time. In addition, the frequency of IRB meetings and the volume and type of submissions varies tremendously across institutions. Check with your IRB office to see expected turn-around times and add a little buffer in case of snow days, natural disasters, or other unanticipated events. If your institutional IRB is accredited, you can likely check existing reports for timelines. Also remember that asking an IRB to speed up a process is often tantamount to “skipping” your colleagues in line. There are circumstances where it is appropriate (e.g., just in time reviews for NSF or NIH funding), but nothing beats advance planning.

Remember History
When researchers are frustrated about turn-around times or lengthy forms that may delay research start times, we recommend taking a moment to consider the historical context that gave rise to IRBs. We invite the reader to consider how you might feel about the current oversight if you, or a loved one, had been a participant in any of the many egregious experiments that led to regulatory bodies. Lengthy forms may be dismissed as tedious. Or they may be considered a great support to ensure a thorough review of the research activities. We recommend researchers shift perspectives to productive stances that facilitate communication and understanding.

Conclusion
Absent clear data to the contrary, there is no need to characterize IRBs as foes in the research enterprise. Researchers work daily to advance science, increase discipline-specific understanding, and use the data for the betterment of society. IRBs are tasked with supporting those activities while overseeing the protection of human participants in research. History has given compelling examples for the need for oversight in research. Researchers and IRB staff/members (who sometimes share both identities), are working different pieces of the same common goal to “do good.” We do not believe that researchers or IRB staff/members operate with a desire to harm or vex anyone. Rather than see each other in opposition, we encourage asking the question: how do we communicate effectively to reach our common goal to move science forward without doing harm? We hope that we have provided some useful ideas.

References

Author note. Melanie M. Domenech Rodríguez and Samantha M. Corralejo, Department of Psychology and Institutional Review Board, Utah State University; Nicole Vouvalis, Institutional Review Board, Utah State University; Alan Mirly, Physician Assistant Studies Department, Idaho State University and Institutional Review Board, Utah State University.

Correspondence concerning this article may be sent to Melanie M. Domenech Rodríguez, 2810 Old Main Hill, Logan, UT 84321.
E-mail: melanie.domenech@usu.edu.
Who’s Laughing Now? The Effects of Sexist and Rape Humor

Nalyn Sriwattanakomen¹
Washington & Jefferson College

ABSTRACT. The present study explored the effects of sexist humor and rape jokes on behavior toward women and rape myth acceptance in both women and men. Prior studies have suggested that exposure to sexist humor leads men who are high in hostile sexism to behave in a discriminatory manner toward women, and feel less ashamed for doing so (Ford, Boxer, Armstrong, & Edel, 2008). It was therefore predicted that sexist and rape jokes would increase rape myth acceptance and sexist behavior. Participants were 96 undergraduates. In partial support of predictions, sexist jokes increased rape myth acceptance among high-hostile sexist men more than nonsexist jokes, $F(1, 42) = 5.58, p = .02, \eta_p^2 = .12$. Additionally, among participants high in hostile sexism, sexist jokes increased rape myth acceptance more in men than in women, $F(1, 42) = 4.97, p = .03, \eta_p^2 = .11$. However, there was no evidence to suggest that sexist jokes increased sexist behavior by making participants less likely to donate to a women’s organization. The implications of these empirical findings and participants’ qualitative responses to the topics of rape and sexism are discussed, and amendments to Ford and Ferguson’s (2004) prejudiced norm theory are proposed.

D
epite efforts to create an egalitarian learning environment in colleges, sexism remains a virulent problem on campuses. Diary studies, in which college students were asked to daily record sexist events that they witnessed or by which they were targeted, have revealed that female students are exposed to as many as 3.45 sexist events per day, and that these sexist events harm their sense of comfort, mood, and self-esteem (Brinkman & Rickard, 2009; Swim, Hyers, Cohen, & Ferguson, 2001). Among these sexist events are sexist jokes, which belong to the larger category of disparagement humor: jokes that demean some social group. Aside from creating distress in its targets, disparagement humor has been proposed to perpetuate cultural and societal prejudice, thereby widening power imbalances between dominant and disadvantaged groups (Ford & Ferguson, 2004).

Recent studies have suggested that the ability of disparaging jokes to change attitudes depends on preexisting societal attitudes toward the disparaged groups and the groups’ societal status (Ford, 2000; Ford, Woodzicka, Triplett, Kochersberger, & Holden, 2014; Greenwood & Isbell, 2002). Social groups toward which society holds ambivalent attitudes may be more vulnerable to being the target of discriminatory attitudes as a result of disparaging humor.

Ford et al. (2014) tested the effect of targeting “ambivalent” social groups such as Muslims on people’s tolerance of discriminatory events in a series of experiments. Tolerance of discriminatory events was measured by the degree of self-directed negative affect (e.g., shame) that individuals reported they would experience upon witnessing
Effects of Sexist and Rape Humor  
Sriwattanakomen

such events. Their findings support the idea that disparagement humor increases tolerance of discriminatory events only for groups toward which society holds ambivalent attitudes. Understanding why and how this occurs is crucial to addressing the issue of sexist humor.

In initial research on the attitudinal effects of disparagement humor, it was proposed that exposure to such humor would increase discriminatory attitudes toward the disparaged group by making negative stereotypes more accessible (Olson, Maio, & Hobden, 1999). If attitudinal change toward disparaged groups were a function of increased stereotype accessibility or a priming effect, then disparaging statements should produce an effect similar to that of the disparaging jokes. However, studies comparing the effect of sexist jokes to sexist statements and neutral jokes have shown that they have approximately the same effect on sexist attitudes, and sexist jokes produce significantly more sexist attitudes (Ford, 2000; Ford, Wentzel, & Lorion, 2001; Gray & Ford, 2013). This finding strongly suggests that there is something unique about humorous communication. The prevailing explanation for the unique effect of disparagement humor is that humor endorses a noncritical mindset—that is, a mindset that interprets only the superficial features of a communication (e.g., tone, facial cues) rather than its deeper meaning. In so doing, disparagement humor trivializes “discrimination under the veil of benign amusement, thus precluding challenges or opposition that nonhumorous [discriminatory] communication would likely incur” (Ford, Boxer, Armstrong, & Edel, 2008, p. 159). However, the trivialization of discrimination is only the first step in the process of channeling exposure to disparagement humor into freer expression of discriminatory attitudes.

Ford and Ferguson (2004) have proposed prejudiced norm theory as the mechanism by which disparagement humor changes attitude expression. The theory states that disparagement humor does not change the evaluative content of stereotypes—that is, the degree to which an individual agrees with stereotypes about a group. Moreover, an individual’s evaluative content of stereotypes is generally regarded as stable across situations, however this theory also posits that it increases tolerance of discriminatory events among people already high in prejudice. This increases the likelihood that these individuals will more freely express their true discriminatory attitudes. The process is broken down into four major parts (Ford & Ferguson, 2004). The first step is the activation of the humorous, noncritical mindset. The second step involves the recipient of the humor perceiving this activation as a local norm of tolerance. People who are exposed to the humor interpret the joke as a signal that everyone in that local context will tolerate (i.e., refrain from critically examining) discrimination against the disparaged group. In the third step, humor recipients use the perceived local norm of tolerance to regulate their own tolerance of discrimination. In the fourth and final step, humor recipients’ preexisting prejudice levels interact with the local norm of tolerance to alter expressions of prejudice. Thus, if the humor recipient is high in prejudice, he or she will be more likely to express prejudice against the disparaged group in that social context. If the humor recipient is low in prejudice, the recipient is unlikely to switch to a noncritical mindset upon hearing the joke (i.e., unlikely to find the joke humorous), therefore the recipient will be no more likely to express prejudice than if the recipient had not received the joke at all. Ford and Ferguson (2004) have based these two predictions on the premise that people high in prejudice have more weakly internalized convictions to behave in a nonprejudiced manner than do nonprejudiced people.

As previously mentioned, sexist humor is a subset of disparagement humor. Given the mixed messages that society sends women—such as greater attention placed on sexual assault investigations under Title IX yet a profusion of product advertisements that portray women as sexual objects through the commodification of their body parts—women are a social group toward which society holds ambivalent attitudes. As such, it would follow from prejudiced norm theory that sexist humor, especially that pertaining to women as sexual objects, would have detrimental effects on attitudes and behaviors toward women.

Sexist Humor: Attitudinal and Behavioral Effects

Since the development of the Ambivalent Sexism Inventory by Glick and Fiske in 1996, researchers have been able to predict who will engage in sexist humor. The inventory has two subscales, one that taps benevolent sexism and another that taps hostile sexism. Hostile sexism consists of negative, antagonistic attitudes toward women whereas benevolent sexism consists of more positive but traditional attitudes toward women.

Greenwood and Isbell (2002) found that people who scored high in hostile sexism judged
misogynistic jokes to be more amusing and less offensive than did people who scored low in hostile sexism. No support was found for gender differences among hostile sexists in terms of amusement or offensiveness ratings. There was a significant negative correlation between amusement and offensiveness ratings; the more amusing a person found the jokes, the less offensive they rated them. Taken together, these findings and those by earlier researchers (Ford, 2000; Ford, Wentzel, & Lorion, 2001) suggest that preexisting sexist attitudes predict how amusing and offensive one finds a sexist joke. It follows, then, that individuals’ amusement and offensiveness ratings can predict sexism levels.

Many studies performed on sexist humor and disparagement humor have used written vignettes, which, although appropriate for the construct of the experiments, may produce issues with ecological validity. Typically, people encounter sexist jokes not by reading the transcript of a conversation between two strangers but by hearing jokes told by friends, acquaintances, or comedians on television. Greenwood and Isbell (2002) addressed this issue by having participants listen to an audio recording that allegedly belonged to a separate study about impression formation featuring two unknown male undergraduates telling blonde jokes in the context of a discussion about a broad range of topics. This presentation technique better represents the actual mode in which people hear sexist jokes.

As research has begun to address the effects of sexist humor on individuals, researchers have continued to factor in preexisting hostile sexism levels under the expectation that hostile sexism mediates the effects of the jokes, which is consistent with prejudiced norm theory. Generally, findings have supported this model.

For example, Ford et al. (2001) found that high-hostile sexist men who were exposed to sexist jokes least expected others to be offended by a hypothetical supervisor’s sexist comments in a written workplace scenario. Also as expected, high-hostile sexist men exposed to sexist jokes anticipated the least shame and disappointment after imagining that they had behaved as the supervisor had, calling a female staff member by an inappropriate pet name and communicating low performance expectations. However, all men exposed to the sexist jokes, regardless of hostile sexism level, reported less shame and disappointment upon imagining themselves as the supervisor and therefore expressed higher tolerance of discrimination. These two findings demonstrated that exposure to sexist humor increased high-hostile sexist men’s tolerance of discriminatory events; the high-hostile sexist men perceived a local norm of tolerance of discrimination. They also suggest that sexist humor may also induce men low in hostile sexism to perceive a sexist event as socially acceptable.

Attention has expanded to include the effects of sexist humor on discriminatory behavior. Ford et al. (2008) found that exposure to sexist humor led men high in hostile sexism to “release” (i.e., express) their prejudice as assessed by decisions made in role-playing tasks. In one study, men were asked to indicate how much they would be willing to give on a scale of $0 to $20 to a women’s rights organization that was soliciting donations. Among participants exposed to the sexist jokes, higher hostile sexism scores predicted smaller donations. The act of donating less to the women’s organization was interpreted as a sign of acting in a discriminatory manner toward women. Thus, exposure to sexist jokes encouraged men high in hostile sexism to release their prejudice.

In a second experiment, Ford et al. (2008) replaced the donation task with a budget-cutting task. Male undergraduates were instructed to cut a total of $24,000 out of the budgets of actual student organizations. Regardless of hostile sexism level, all men who were exposed to sexist jokes allocated a greater percentage of the funding cuts to the women’s organization. The higher men scored in hostile sexism, the more they cut from the women’s organization. The higher men scored in hostile sexism, the more they cut from the women’s organization. The higher men scored in hostile sexism, the more they cut from the women’s organization. The higher men scored in hostile sexism, the more they cut from the women’s organization. The higher men scored in hostile sexism, the more they cut from the women’s organization. The higher men scored in hostile sexism, the more they cut from the women’s organization. The higher men scored in hostile sexism, the more they cut from the women’s organization. The higher men scored in hostile sexism, the more they cut from the women’s organization. The higher men scored in hostile sexism, the more they cut from the women’s organization. The higher men scored in hostile sexism, the more they cut from the women’s organization. The higher men scored in hostile sexism, the more they cut from the women’s organization. The higher men scored in hostile sexism, the more they cut from the women’s organization. The higher men scored in hostile sexism, the more they cut from the women’s organization. The higher men scored in hostile sexism, the more they cut from the women’s organization. The higher men scored in hostile sexism, the more they cut from the women’s organization. The higher men scored in hostile sexism, the more they cut from the women’s organization. The higher men scored in hostile sexism, the more they cut from the women’s organization.
that the worst effect that sexist humor has on men is an increased likelihood to discriminate against women by opposing or expressing apathy about their political and occupational advancement. However, opposition to women’s advancement may not be the only consequence. There is a body of research that has suggested that exposure to sexist humor can actually increase rape proclivity, which is a person’s self-reported likelihood of committing rape (Romero-Sánchez, Durán, Carretero-Dios, Megías, & Moya, 2010; Thomae & Viki, 2013).

Romero-Sánchez and colleagues (2010) studied the relationship between hostile sexism, exposure to sexist humor, and rape proclivity among male undergraduates in Spain. Rape proclivity was assessed by reading date-rape vignettes and indicating how likely one would be to behave like the rapist. Men high in hostile sexism reported higher rape proclivity in general, and sexist jokes increased rape proclivity across all men, particularly those high in hostile sexism. These findings have been replicated in other studies (Abrams, Viki, Masser, & Bohner, 2003). However, in an experiment by Thomae and Viki (2013), controlling for the sexual content of these sexist jokes eliminated the increase in rape proclivity across all men, suggesting that it is the combination of sexist and sexual content that increases rape proclivity.

Overall, the tendency for high-hostile sexist men to report higher rape proclivity upon exposure to sexist jokes supports prejudiced norm theory. Within this theoretical framework, the sexist jokes created a local norm of tolerance for men who had weak internalized convictions about expressing adversarial attitudes toward women, thus allowing them to more freely express their true attitudes toward rape.

The Present Study
The key purposes of the present study were fourfold. The first was to examine how rape jokes influence rape myth acceptance, hostile attitudes toward women, and behavior toward women. Thus far, no published studies have explored the effects of rape humor. The jokes in the aforementioned studies have made light of gender stereotypes or made references to women performing sexual acts, but none of the jokes have addressed rape per se. Additionally, only men’s rape proclivity has been tested. This selective testing makes sense given that men commit the vast majority of rape, but women can still hold attitudes that in some way support or justify rape, if not indicate a proclivity to rape. Such attitudes can be measured by the degree to which a person accepts rape myths such as the idea that women who dress in skimpy clothing are “asking” to be raped. Multiple scales for measuring rape myth acceptance have been created (Burt, 1980; Deitz, Blackwell, Daley, & Bentley, 1982), but one of the most reliable and widely used scales is the Illinois Rape Myth Acceptance Scale (IRMA), which recently has been adapted by McMahon and Farmer (2011) into a form that better reflects the colloquial language of college students. Namely, men are referred to as “guys,” and women are referred to as “girls.”

The second key purpose was to compare the effects of rape jokes to those of nonsexual sexist jokes, which have been examined in the literature, and to do so by examining actual changes in sexist attitudes, which have not yet been addressed in the literature. Prior studies have examined only the expression of sexist attitudes after joke exposure. This comparison of rape jokes and nonsexual sexist jokes functioned to determine if one category of joke was more influential than the other on attitudes and behavior.

The third key purpose of the present study was to address issues of ecological validity and internal validity by having participants listen to jokes and by controlling for the effect of sexual content in jokes. Finally, the present study examined gender differences in responses to sexist and rape humor, another topic that has been largely ignored in the literature. Given the literature that suggests that humorous communications operate differently than straightforward statements or events (Ford, Wentzel, & Lorion, 2001), it could be that the sexist humor makes women more expressive of prejudice toward themselves, which in turn harms their self-esteem and induces negative affect. However, the opposite effect is also possible. Given that sexist events have been shown to harm women’s self-esteem and induce negative affect (Brinkman & Rickard, 2009), it is conceivable that sexist humor will also induce negative affect and thereby make women less likely to express prejudice toward their own social group. Men, on the other hand, are not threatened by the content of such sexist humor, therefore sexist humor is less likely to induce negative affect and decrease the expression of sexism.

The study employed a 2 (sexual content of jokes: sex, no sex) × 2 (sexist content of jokes: sexist, nonsexist) × 2 (participant gender: woman, man) × 2 (hostile sexism level: high, low) factorial design. Combinations of the sexual content and
sexist content factors yielded four joke categories: gender stereotype, rape, racial, and ribald. The rationale for having these four categories was to address the confounds of sexual content and offensiveness that arose in prior studies. As demonstrated by Thomae and Viki (2013), the inclusion of sexual content confounded the effect of joke type, making low-hostile sexist men more likely to report higher rape proclivity. Another potentially confounding factor was the differential offensiveness of the jokes. In the many studies conducted by Ford, the “control” condition consisted of neutral jokes that were not at all offensive. Hence, the joke categories of racial and ribald served as “controls” for offensive and sexual content. The racial category consisted of jokes that were offensive but neither sexist nor sexual in nature, serving as a comparison for the gender stereotype jokes, which were sexist but not sexual in nature. The ribald category consisted of jokes that were nonsexist but sexual in nature, serving as a comparison for the rape jokes, which were both sexist and sexual in nature. To measure the changes in attitudes toward women and rape myth acceptance, a pretest-posttest design was employed with items measuring sexism and rape myth acceptance in both tests.

Based on results from prior studies, it was hypothesized that there would be a main effect of joke type on attitudes toward women, rape myth acceptance, and behavior toward women. Specifically, it was predicted that sexist jokes (i.e., gender stereotype and rape) would increase hostile attitudes toward women and rape myth acceptance more than nonsexist jokes (i.e., racial and ribald), and that participants exposed to sexist jokes should be less likely to donate to either a women’s organization or a rape hotline on a behavioral measure adapted from Ford et al.’s (2008) donation task. Second, it was hypothesized that men would be more affected by the sexist jokes than would women. Third, it was predicted that rape jokes would increase rape myth acceptance more than any other joke group. Based on findings from prior studies (Abrams et al., 2003; Ford et al., 2001; Ford et al., 2008; Romero-Sánchez et al., 2010), it was also anticipated that all of the aforementioned effects would be more pronounced for people high in hostile sexism. Finally, based on Thomae and Viki’s (2013) findings that hostile sexism and rape myth acceptance are related, it was anticipated that hostile attitudes toward women scores and rape myth acceptance scores would correlate positively with each other.

**Method**

**Participants**

Participants consisted of 113 undergraduate students at a small liberal arts college. They were recruited from 32 introductory-level courses across various departments. Additional recruitment of male participants was undertaken through snowballing (i.e., participants bringing along friends) and by contacting fraternity leaders. Junior and senior psychology majors were eliminated from the pool of participants so as to reduce the number of participants who might expect and actively search for deception and therefore see through the deception that was critical to obtaining honest responses. Participants self-identified their gender. Those who identified themselves as “man,” “male,” or “masculine,” were coded as men. Those who identified as “woman,” “female,” or “feminine,” were coded as women. There were roughly equal number of men (n = 53) and women (n = 59). Most participants (n = 91) identified as White or European American. Of the remaining participants, eight identified as Black or African American, five identified as Hispanic, three identified as Asian, and six identified as mixed race. Four participants identified as international students. Of the 113 participants, one failed to indicate gender, one failed to choose an organization for the behavioral measure, and 15 declined to answer all attitudes toward women or rape myth acceptance questions on the pretest and/or posttest. These 17 participants were excluded from all analyses, which left a final sample of 96 participants.

**Materials**

The pretest, jokes, and posttest were presented on computers. Participants were provided with headphones, which were used to listen to the jokes and reduce the audibility of ambient laughter. The jokes were presented as audio clips of a joke-telling exchange in the college dining hall. Each audio clip consisted of an ostensibly male speaker telling a joke, followed by laughter from a group of two men and two women, who were described as friends of the participant. Ambient noise from the dining hall ran in the background. This format was used to increase ecological validity; college students are more likely to hear jokes within friend groups, not overhear them in conversations between strangers.

The jokes were selected on the basis of a pilot study involving 23 randomly selected junior and senior psychology majors. In the pilot study, each participant was randomly assigned 20 jokes from
Effects of Sexist and Rape Humor

Sriwattanakomen

two of the joke conditions and was asked to rate the humorousness of the jokes. Jokes were selected on the basis of two criteria: (a) having a median humor rating between 1 (not at all funny) and 2 (somewhat funny), inclusive; and (b) having a moderately variable distribution of humor ratings. The rationale for the criteria was that the most common humor rating for the majority of the gender stereotype and rape jokes was 1 (not funny at all), but it would be contrary to the purposes of the study to select jokes that no one found humorous, so jokes were selected that elicited at least a little amusement and appealed differently to a range of participants.

Two attitudinal assessments and one behavioral assessment were used to measure initial attitudes toward women and the attitudinal and behavioral effects of the jokes. The first attitudinal assessment gauged attitudes toward women. This assessment consisted of a modified version of the hostile sexism subscale of the Ambivalent Sexism Inventory and four items taken from Benson and Vincent’s (1980) Sexist Attitudes Towards Women Scale (SATWS). The SATWS has been found to have high internal consistency for college students, and items on the scale are highly correlated with appreciation of sexist humor (Benson & Vincent, 1980). Four items from the SATWS were chosen on the basis that they reflect relatively controversial opinions that are applicable to college students (e.g., “women rely more on intuition and less on reason than men do”) and that these opinions are not addressed by the hostile sexism portion of the Ambivalent Sexism Inventory. The same method used by McMahon and Farmer (2011) to make the IRMA more applicable to college students was used to modify the hostile sexism and SATWS items. In the pilot study, Cronbach’s alpha was .62 for the seven pretest items and .74 for the seven posttest items. Split-half reliability was within acceptable limits, \( r(21) = .81, p < .001 \). The items were divided between the pretest and the posttest by randomly selecting five items from the Ambivalent Sexism Inventory Hostile Sexism subscale and two items from the SATWS for each test. If any two items on a test were found to be very similar (e.g., “feminists are making entirely reasonable demands of guys” and “feminists are not seeking for women to have more power than guys”), one of them was swapped with an item on the other test so as to avoid redundancy on a single test.

This process resulted in a pretest whose questions consisted almost entirely of items from the Ambivalent Sexism Inventory and a posttest whose questions consisted of roughly equal numbers of items from the Ambivalent Sexism Inventory and the SATWS. Given that the SATWS items contained statements that were sexist but not necessarily hostile in nature (e.g., “girls rely more on intuition and less on reason than guys do”), the degree to which they tapped the different valences of sexism (i.e., hostile and not hostile) was not equal. This methodological limitation, which was discovered after the results had been analyzed, precludes any valid comparison of the pretest and posttest scores for attitudes toward women, therefore the results for the attitudes toward women assessment will not be reported.

The second attitudinal assessment consisted of a selection of the rape myth acceptance items from McMahon and Farmer’s (2011) modified version of the IRMA. All of the scales were divided to form a pretest and a posttest. Filler questions and a selection of modified questions from Marlowe & Crowne’s (1960) Social Desirability Scale were added to the pretest and posttest to obscure the purpose of the survey. In the pilot study, Cronbach’s alpha was .78 for the nine pretest items and .77 for the nine posttest items. Split-half reliability was within acceptable limits, \( r(22) = .76, p < .001 \). As with the attitudes toward women items, the rape myth acceptance items were randomly divided between the pretest and the posttest and then adjusted to reduce redundancy of items on a single test.

The behavioral assessment consisted of a modified version of the donation task from Ford et al. (2008): Participants were asked to select one of several social justice organizations to which to have $2 donated for their participation. The target organizations were a women’s rights organization and a sexual assault hotline.

Procedure

After institutional review board approval (PSY501-2015S-099) was given, participants were recruited and scheduled to take part in the experiment in the computer lab. Participants completed the pretest under the guise that it was an unrelated survey on social attitudes being conducted by the researcher’s supervisor. In keeping with this cover story, participants were informed that they could skip this section without penalty. After completing the pretest, participants were directed to a fake debriefing page and were then presented with a second informed consent form on the computer stating that the purpose of the present study was...
to examine the relationship between sociopolitical attitudes and humor appreciation. Participants were then prompted to put on their headphones and were randomly assigned to one of four joke conditions: gender stereotype (i.e., sexist, no sexual content), racial (i.e., nonsexist, no sexual content), rape (i.e., sexist, sexual content), and ribald (i.e., nonsexist, sexual content). The Appendix shows examples of jokes from each condition.

Participants were instructed to close their eyes and imagine themselves sitting in the dining hall with a mixed group of friends, “two girls and two guys.” They then clicked a button on the screen to hear a recording of an ostensibly male speaker telling a joke followed by canned laughter. Beneath the audio player for each joke, a text version of the joke was available onscreen. Participants were asked to rate the humorousness of the joke on a six-interval scale, from “not funny at all” to “extremely funny.” Each participant was exposed to a total of four jokes from their assigned category, plus three neutral jokes that were common to all four conditions. The purpose of the neutral jokes was to decrease suspicion about the purpose of the study.

Upon completion of the joke ratings, participants completed the posttest and completed a series of demographic questions intended to reduce suspicion and support the alleged purpose of the study. Next, participants were given two free-response questions intended to serve as a deception check and a manipulation check, respectively: (a) “What do you think the purpose of the study was?” and (b) “What did you think of the jokes (e.g., were they funny, offensive, worth repeating)? Were there any you didn’t understand?”

Upon completion of the survey, participants were presented with a debriefing page containing the behavioral measure of attitudes toward women. As a form of compensation, participants were given a choice of organizations to which the Psychology Department would donate $2. The organizations included the National Organization for Women and the National Sexual Assault Hotline (i.e., the organizations of interest), as well as five other social justice organizations (i.e., National Association for the Advancement of Colored People, Wounded Warrior Project, National Alliance on Mental Illness, National Network for Immigrant and Refugee Rights, Gay & Lesbian Alliance Against Defamation), each with a brief description. After participants indicated their preferred organization, they submitted their data and reported to the experimenter for debriefing.

Participants were verbally debriefed, either alone or in small groups if more than one participant finished at the same time. The verbal debriefing presented an opportunity to gather feedback and other qualitative information from participants including reactions to the questions and jokes.

Results

Manipulation Check
More than a third of participants ($n = 42$) wrote that they found the gender stereotype, rape, or racial jokes personally offensive or offensive to others. This finding suggested that these jokes expressed the politically incorrect material that they were meant to.

Deception Check
None of the participants expressed suspicion or correctly guessed in either the free-response section or the verbal debriefing that the study was about the effect of jokes on attitudes. However, in the free-response section, 43 participants indicated that they thought the study was about attitudes toward women, feminism, sexism, rape, or rape culture. These participants will henceforth be referred to as “suspicious participants.”

In a post-hoc analysis, participants were coded as either suspicious or unsuspecting. It was found that suspiciousness was linked to hostility toward women. Suspicious participants reported significantly more hostile attitudes toward women on both the pretest, $t(94) = -2.15$, $p = .03$, Cohen’s $d = .44$, and the posttest, $t(94) = -2.18$, $p = .03$, Cohen’s $d = .45$. However, no significant differences were found between the rape myth acceptance pretest scores for the two groups or between the rape myth acceptance posttest scores. A Pearson chi-square test for independence was performed on the relationship between suspiciousness and gender. The relationship did not reach significance, $\chi^2(1, N = 96) = 1.00$, $p = .32$, Cramer’s $V = .10$.

This pattern of results suggests that participants who held more hostile attitudes toward women were more sensitive to or aware of questions pertaining to sexism or rape. To better examine the effect of higher hostile sexism, analyses were performed both with and without the suspicious participants.

Hostile Sexism Designation
In keeping with procedures used in other studies that divided participants into low and high-hostile sexists (Ford, 2000), participants were designated...
as either low or high in hostile sexism based on a median split on the distribution of scores from the pretest attitudes toward women items ($Mdn = 3.04$), which consisted almost entirely of items from the hostile sexism subscale of Glick and Fiske’s (1996) Ambivalent Sexism Inventory. The mean score for those low in hostile sexism was 2.21 ($SD = 0.60$), and the mean score for those high in hostile sexism was 3.94 ($SD = 0.61$).

**Effects on Rape Myth Acceptance**

Rape myth acceptance was measured on a 6-point Likert-type scale ranging from 1 (strongly disagree) to 6 (strongly agree). Lower scores indicated less acceptance of rape myths, and higher scores indicated greater acceptance of rape myths. Cronbach’s alphas for the nine pretest and nine posttest rape myth acceptance items were .81 and .87, respectively, indicating good reliability.

Rape myth acceptance difference scores were calculated by averaging the pretest rape myth acceptance scores and the posttest scores and subtracting the pretest scores from the posttest scores. A 2 (gender: woman, man) × 2 (sexual content of jokes: sex, no sex) × 2 (sexist content of jokes: sexist, nonsexist) × 2 (hostile sexism level: high, low) Analysis of Variance (ANOVA) was performed on these difference scores to test the hypothesis that sexist jokes would increase rape myth acceptance, especially among men and those high in hostile sexism. Contrary to predictions, no main effect of sexist content was found, $p = .26, \eta^2 = .02$, nor did any other main or interaction effects reach significance.

With the suspicious participants excluded, another 2 (gender: woman, man) × 2 (sexual content of jokes: sex, no sex) × 2 (sexist content of jokes: sexist, nonsexist) × 2 (hostile sexism level: high, low) ANOVA was performed on the rape myth acceptance difference scores. There was a significant three-way interaction between sexist joke content, hostile sexism level, and gender, $F(1, 42) = 4.40, p = .04, \eta^2 = .10$. As shown in Figure 1, the direction of the difference between rape myth acceptance difference scores for high-hostile sexist men was opposite that for high-hostile sexist women. Furthermore, the magnitude of these differences between the genders was greater for high-hostile sexist participants than for low-hostile sexist participants. In support of these graphical inferences, simple main effects analyses revealed two significant effects. First, among men high in hostile sexism, sexist jokes increased rape myth acceptance ($M = 2.67, SD = 6.98$) more than nonsexist jokes ($M = -1.44, SD = 7.88$), $F(1, 42) = 5.58, p = .02, \eta^2 = .12$. Second, for participants high in hostile sexism, sexist jokes increased rape myth acceptance more for men ($M = 2.67, SD = 6.98$) than for women ($M = -2.20, SD = 5.31$), $F(1, 42) = 4.97, p = .03, \eta^2 = .11$. No other interaction or main effects reached significance, $ps > .05$.

**Effect on Behavior Toward Women**

Behavior toward women was measured indirectly through the donation task, in which the participant chose one social justice organization from a list of seven including the target organizations of the National Organization for Women and the National Sexual Assault Hotline. A Pearson chi-square test for independence was performed on the relationship between joke condition and whether or not participants donated to one of the women’s organizations to test the hypothesis that sexist jokes would worsen behavior toward women, particularly among those high in hostile sexism. The relationship failed to reach significance, $p = .09$, Cramer’s $V = .26$. With suspicious
participants excluded, another Pearson chi-square test for independence was performed on the relationship between joke condition and donation behavior. This relationship also failed to reach significance, $p = .06$, Cramer’s $V = .36$.

**Attitudes Toward Women and Rape Acceptance**

Given that attitudes toward women were measured by participants’ agreement with all statements from the hostile sexism portion of Glick and Fiske’s (1996) Ambivalent Sexism Inventory, pretest attitudes toward women can, in the scope of this study, be treated as a measure of hostile sexism. A Pearson’s $r$ correlation was performed on the pretest attitudes toward women scores and rape myth acceptance scores. A strong, significantly positive correlation was found between the two, $r(94) = .72$, $p < .001$, suggesting that hostile sexism and rape myth acceptance are closely related.

**Discussion**

The primary focus of this study was to determine the effect of sexual and sexist humor on attitudes toward women, rape myth acceptance, and behavior toward women. More specifically, it was hypothesized that exposure to sexist jokes (i.e., rape and gender stereotype) would increase rape myth acceptance and discrimination against women when choosing a charitable organization to which to donate. It was also hypothesized that rape jokes in particular would increase rape myth acceptance and that all of the above effects would be more pronounced for people high in hostile sexism. Overall, the findings lent no support to the behavioral hypotheses and mixed support to the rape myth acceptance hypotheses.

**Effects of Sexist and Rape Humor**

**Effect on rape myth acceptance.** With suspicious participants included, no support was found for the hypothesis that among high-hostile sexist men, gender stereotype or rape jokes increased rape myth acceptance more than nonsexist jokes. There was also a lack of evidence for the predictions that sexist jokes would increase rape myth acceptance across the board and that sexist jokes would increase rape myth acceptance more for men than for women. However, with the suspicious participants excluded, a pattern emerged that partially supported these hypotheses. Among high-hostile sexist men, sexist jokes increased rape myth acceptance more than nonsexist jokes. Moreover, sexist jokes did increase rape myth acceptance more for men than for women—but only among those high in hostile sexism.

These findings should be interpreted with caution because of the two different ways in which rape myth acceptance can be defined. On the one hand, increased rape myth acceptance could qualify as freer expression of prejudice toward a subset of women: rape victims. In this case, the sexist jokes presumably activated a local norm of tolerance of prejudice for men who were already prejudiced toward women, thereby increasing their willingness to agree with rape myth statements—that is, express prejudice toward women. This interpretation is consistent with prejudiced norm theory. On the other hand, however, changes in rape myth acceptance could be conceptualized as changes in the evaluative content of stereotypes. Rape myths equate to stereotypes about rape victims in that they depict situations or qualities of the rape victim that are often seen on popular television or in media coverage of real rape cases. One such example is the 2012 Steubenville High School rape case, in which members of the Steubenville community blamed the intoxicated female victim (Macur & Schweber, 2012) and CNN news coverage focused on the grief of the rapists, downplaying their culpability. Thus, the sexist jokes made high-hostile sexist men more adopting of stereotypes about rape victims. This interpretation challenges one component of prejudiced norm theory, which states that the evaluative content of stereotypes is a stable trait (Ford & Ferguson, 2004).

In interpreting the difference in changes in rape myth acceptance between high-hostile sexist men and women, another challenge to the literature on prejudiced norm theory arises. Although it is clear that sexist jokes increased rape myth acceptance more for high-hostile sexist men than for high-hostile sexist women, it could also be stated that sexist jokes decreased rape myth acceptance more for high-hostile sexist women than for high-hostile sexist men. Prior studies have disregarded the effects of sexist jokes on women’s expression of prejudice on the basis that their hostile sexism levels were too low (Ford, Wentzel, & Lorion, 2001). This finding illustrates that women can be high in hostile sexism and that sexist jokes do have an effect on their expression of prejudice or acceptance of stereotypes.

Together, these findings support the broader concept of prejudiced norm theory—that exposure to prejudicial humor increases prejudice in those who are already prejudiced. However, the finer
Effects of Sexist and Rape Humor

Effects of Sexist and Rape Humor

Sriwattanakomen

participants to self-report their personal investment to determine if this was the case would be to ask Lesbian Alliance Against Defamation. One way their delight that they could donate to the Gay & of the college’s Gay-Straight Alliance expressed interviews, in which some self-identified members out by qualitative data from the postexperimental the other organizations. This explanation is borne toward women but a personal investment in one of the targeted group, which could demand an amendment of prejudiced norm theory.

Effect on behavior. It was predicted that participants exposed to rape or gender stereotype jokes would be less likely to donate to a women’s organization—either the National Organization for Women or the National Sexual Assault Hotline—than would participants exposed to racist or ribald jokes. The results failed to support this prediction. Possible reasons for the lack of support include the confounds of the differential familiarity of the organizations, whereby more familiar organizations (e.g., the Wounded Warrior Project) attracted disproportionately more donations than more obscure organizations (e.g., the National Network for Immigrant and Refugee Rights), and the politically charged nature of certain organizations. For example, the National Organization for Women is known to support liberal candidates who support women’s reproductive rights including abortion. By contrast, the Wounded Warrior Project has no political affiliations, so participants might have deemed it a “safer” choice. Possible solutions to these limitations include asking participants why they chose a given organization, including only organizations that have no political affiliations, or selecting organizations of equal familiarity.

Other possible explanations for the lack of confirmatory findings include the construct validity of the behavioral measure. Unlike Ford et al.’s (2008) behavioral measures, which involved either (a) donating to one organization (a women’s organization) or (b) potentially cutting the budget of a women’s organization, the behavioral measure in the present study was diffuse. Choosing to donate to one organization over one of the women’s organization might have reflected not discrimination toward women but a personal investment in one of the other organizations. This explanation is borne out by qualitative data from the postexperimental interviews, in which some self-identified members of the college’s Gay-Straight Alliance expressed their delight that they could donate to the Gay & Lesbian Alliance Against Defamation. One way to determine if this was the case would be to ask participants to self-report their personal investment in each organization before exposure to the jokes.

Methodological limitations aside, yet another explanation for the lack of confirmatory findings is that the sexist jokes simply did not increase discriminatory behavior against women. Further inquiry is needed. It may be fruitful to use more than one behavioral measure so as to distinguish discrimination against women in general from women rape victims. Such multi-faceted behavioral measures could involve asking participants to choose an amount to donate to women in different scenarios.

Attitudes Toward Women and Rape Acceptance

The finding that initial hostile attitudes toward women correlated very strongly and positively with initial rape myth acceptance suggests that hostile sexism and rape myth acceptance are closely related. Given that the rape myths in this study hinged on the woman being perceived as malicious or less than a person, it is likely that hostile sexism underlies rape myth acceptance. Another possibility is that both traits belong to some larger complex such as a social dominance orientation.

Limitations

Of the limitations of the present study, the most substantial was the methodological issue with the assessment of attitudes toward women discussed earlier. No conclusions can be drawn for the effect of jokes on attitudes toward women. This limitation can be addressed by assessing hostile sexism in a temporally removed setting and then loading the pretest and the posttest with nonhostile sexist attitudes toward women items only. Alternately, efforts can be made to include equal numbers of hostile sexist and nonhostile attitudes toward women items on the pretest and the posttest. The temporally removed assessment of hostile sexism could also serve to reduce suspicion among participants. Aside from this methodological issue and the one concerning the behavioral measure, other methodological issues include participants’ concern with self-presentation, which inevitably occurs when using self-report or surveys as a measurement tool. By nature, people are more inclined to respond in a socially desirable manner than to respond honestly.

Other limitations of the present study include distractions, the sample gathered, and the gender of the researcher and lab assistants. First, at minimum, a dozen participants were observed texting or answering phone calls during the study. Such
distractions conceivably decreased the internal validity of the study by interrupting joke exposure or by introducing confounding variables in the form of interaction with and information from others. Second, the sample was one of convenience, although an attempt was made to make the sample representative. Moreover, the researcher might have influenced the way participants responded to the attitudes toward women and rape myth acceptance statements, as well as to the jokes. Bearing in mind that the researcher is a woman, participants might have been more guarded about their true attitudes toward women than they would have been had the experimenter been a man, as in other studies (Ford et al., 2008).

General Discussion

Overall, the present study lent some support to Ford and Ferguson’s (2004) prejudiced norm theory and the researcher’s predictions about gender differences. The present study also identified important areas for further study.

First, sexist humor increased rape myth acceptance among high-hostile sexist men and affected high-hostile sexist men more than high-hostile sexist women—but only when suspicous participants were excluded from analyses. The fact that these effects emerged only when suspicious participants were excluded suggests that the suspicious participants had some sort of dilative effect. That is, their responses on the pretest or the posttest offset the changes observed in unsuspecting participants. This observation raises the possibility that suspicous participants were higher self-monitors than unsuspecting participants, so they were more inclined to respond consistently on the pretest and the posttest.

It is also possible that the cause of their suspicousness was heterogeneous; perhaps some were high self-monitors whereas others were recently involved in a sexist incident or took a women’s studies class. Should high self-monitoring be found to be a mediator of responses to sexist humor or statements, prejudiced norm theory must be expanded to account for this effect. In any case, the identification of naturally existing groups (suspicious and unsuspecting) is essential to removing confounds and therefore promoting internal validity, much in the way that researchers in the field have identified high- and low-hostile sexists participants.

Second, although the quantitative effect of sexist humor on attitudes toward women could not be ascertained because of methodological issues, qualitative data from the free-response questions and the debriefing strongly suggest that attitudes toward women are mutable and that, in accordance with prejudiced norm theory, the expression of hostile attitudes could be dependent on preexisting hostile sexism levels. Some participants expressed that the statements about rape and women on the survey made them feel uncomfortable because they knew they ought to respond a certain way but felt differently—that is, in a more hostile manner. Two participants suggested adding a free-response block to each statement so they could explain why they somewhat agreed with the rape myths statements. In the free-response section, one participant wrote quite tellingly:

Although they were offensive, I found most of [the jokes] to be very funny. I think that people need to losen [sic] up a little bit (more like a lot). Most of our issues is [sic] because everyone takes themselves too seriously. A specific comment about rape...obviously it is terrible, obviously no one should every commit it. However, [...] I think women need to understand the peril that they could be in a given situation. If you knew that a hungry lion was [sic] waiting in a cave, would you walk into it? That is why I answered the way I did. I’m not a woman hater. I just think rational solutions for this horrendous act should be multi-dimensional. I explained this to provide you insight into my answer and so tht [sic] you will think I’m a less terrible person, hopefully.

The extreme concern with self-presentation paired with the obvious acceptance of rape myths (i.e., victim blaming) evident in this comment point to the difficulty of accurately measuring rape myth acceptance and attitudes toward women, as well as the role that suspicion of or vigilance to the topics of sexism and rape plays in the expression of attitudes toward women, especially rape victims. This comment also demonstrates the sort of ambiguous attitudes that college students hold toward the issues of rape and the behavior of women.

In conclusion, the findings of the present study show that sexist humor can indeed influence the way that people think and feel about rape victims—but only when suspicous participants were excluded from analyses. The fact that these effects emerged only when suspicous participants were excluded suggests that the suspicous participants had some sort of dilative effect. That is, their responses on the pretest or the posttest offset the changes observed in unsuspecting participants. This observation raises the possibility that suspicous participants were higher self-monitors than unsuspecting participants, so they were more inclined to respond consistently on the pretest and the posttest.

It is also possible that the cause of their suspicousness was heterogeneous; perhaps some were high self-monitors whereas others were recently involved in a sexist incident or took a women’s studies class. Should high self-monitoring be found to be a mediator of responses to sexist humor or statements, prejudiced norm theory must be expanded to account for this effect. In any case, the identification of naturally existing groups (suspicious and unsuspecting) is essential to removing confounds and therefore promoting internal validity, much in the way that researchers in the field have identified high- and low-hostile sexists participants.

Second, although the quantitative effect of sexist humor on attitudes toward women could not be ascertained because of methodological issues,
Effects of Sexist and Rape Humor | Sriwattanakomen

systematically probe participants’ thought processes as they respond to sexist or rape-myth accepting statements, and systematically examine the effect that suspicion has on to such statements, not to mention the relationship that suspicion may have with other traits. More importantly, however, future research must include women, as the present study has shown that women can be high in hostile sexism and that sexist jokes can alter how they respond to rape myths. Women’s responses to rape victims contribute enormously to rape culture, college climates, and how they appraise themselves should they be raped, which makes this topic especially exigent. The inclusion of women in future research in this area is also essential to inspecting the role that being exposed to disparagement humor that targets one’s own social group plays in how one responds in a sexist setting and whether any long-term attitudinal changes result.

As this study illustrates, sexist humor can lead people to trivialize rape. People continue to consume rape humor by watching inappropriate television programs, overlook or discount blatant sexist and sexual assault-endorsing qualities when electing people to public office, and rationalize rape myths. As long as this continues, colleges, not to mention society at large, will remain hazardous places for women to exist, places for sexual violence to seethe in silence.

References


The author thanks Elizabeth Bennett for her immense help designing the study, analyzing the results, proofreading this manuscript, and providing general support and consultation. The author also thanks Jeffrey Chrabaszcz for his statistical assistance, and Emma Crawford, Alexandra Norris, and Joshua Dewaine for their invaluable assistance tending the lab. Special thanks to all Psi Chi journal reviewers for their support. Lastly, the author thanks Timothy Klitz and Carla Myers for additional proofreading.

Correspondence concerning this article should be addressed to Nalyn Sriwattanakomen, Dynamic Decision Making Laboratory, Carnegie Mellon University, 4609 Winthrop St, 1st Floor Suite, Pittsburgh, PA, 15213. E-mail: nsriw1@gmail.com.
# APPENDIX

## Examples of Jokes From Each Condition

<table>
<thead>
<tr>
<th>Condition</th>
<th>Joke</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender Stereotype (Sexist × No Sex)</td>
<td>Adam was walking around the Garden of Eden feeling very lonely, so God asked, “What is wrong with you?” Adam said he didn’t have anyone to talk to. God said he was going to give him a companion. God said, “This person will cook for you and wash your clothes. She will always agree with every decision you make. She will bear your children and never ask you to get up in the middle of the night to take care of them. She will not nag you, and she will always be the first to admit she was wrong when you’ve had a disagreement.” Adam asked God, “What will this woman cost?” God replied, “An arm and a leg.” Adam said, “What can I get for just a rib?”</td>
</tr>
<tr>
<td>Racial (Nonsexist × No Sex)</td>
<td>Q. A black guy and a Mexican are in a car. Who is driving? A. A cop.</td>
</tr>
<tr>
<td>Rape (Sexist × Sex)</td>
<td>Statistically speaking, 9 out of 10 people enjoy gang rape.</td>
</tr>
<tr>
<td>Ribald (Nonsexist × Sex)</td>
<td>A little girl is in line to see Santa. When it’s her turn, she climbs up on Santa’s lap. Santa asks, “What would you like Santa to bring you for Christmas?” The little girl replies, “I want a Barbie and a G.I. Joe.” Santa looks at the girl for a moment and says, “But I thought Barbie comes with Ken.” “No,” says the little girl. “She comes with G.I. Joe; she fakes it with Ken.”</td>
</tr>
</tbody>
</table>
People often form ideas about their sexual performance and ability to sexually satisfy partners and evaluate their own ability to manage the sexual aspects of their lives. Furthermore, how individuals evaluate sexual aspects of their lives has importance for sexual satisfaction (Birnbaum, Reis, Mikulincer, Gillath, & Opraz, 2006) and relationship satisfaction (Brassard, Dupuy, Bergeron, & Shaver, 2015); these are key factors in relationship stability and commitment. Subsequently, it is useful to understand how those with different genders (e.g., women, men, transgender, genderqueer) and sexual identities (e.g., lesbian, gay, bisexual, pansexual, heterosexual) evaluate their own sexuality in forming conceptualizations of clients and theories of sexuality, alike. Limited (and sometimes contradictory) research currently makes it difficult to infer how gender and sexual identity are related to aspects of sexual self-concept. Recently, authors have commented on the distinction between sexual identity and sexual orientation. Hughes, Camden, and Yangchen (2016) point out that, while sexual orientation communicates an individual’s (mostly) enduring attraction, sexual identity “is the label that people adopt to signify to others who they are as a sexual being, particularly regarding sexual orientation” (Grollman, 2010, para.4 as cited in Hughes et al. 2016, p. 145). In the present study, we hoped to explore how those of different genders and sexual identities understand aspects of their sexual self-concept.

Compared to heterosexual individuals, those who identify as lesbian, gay, and bisexual (LGB) face unique stressors in life. Despite advances in gaining legal equality and social acceptance (Pew Research Center, 2013), recent research has continued to indicate that LGB people are targets of stigma by society in general (Nadal, 2013) and by police officers (Mallory, Hasenbush, & Sears, 2015), experience marginalization of romantic relationships (Lehmiller & Agnew, 2007), and confront unique challenges in identity formation (Rosario, Schrimshaw, & Hunter, 2008; Rosario, Schrimshaw, Hunter, & Braun, 2006). Previous findings have suggested that the impact of managing these experiences results in greater mental distress for people who identify as LGB than those who identify as heterosexual (Cochran, 2001; Meyer, 2003, Riggle & Rostosky, 2011; Wadsworth & Hayes-Skelton, 2015). How people who identify as LGB manage
these experiences has recently become a focus of researchers (Bruce, Harper, & Bauermeister, 2015; Hill & Gunderson, 2015; Meyer, 2015). Although these reports have furthered understanding of LGB psychology, they have not addressed how sexual orientation and sexual identity may be related to aspects of sexual self-concept. However, it seems possible that being a member of a group stigmatized because of sexual orientation may impact LGB individuals' sexual self-concept. This possibility has implications for understanding identity development and advocacy. For example, if holding a stigmatized identity or gender is associated with sexual self-concept, clinical interventions and psychoeducation programming may be important resources to make available to individuals. Similarly, if gender differences exist in sexual self-concept, it is important to consider whether differences are attributable to gender alone or if there might be an interaction effect between gender and sexual identity. Because gay and bisexual men often experience disproportionate stigma because of their sexual identities compared to lesbian and bisexual women, we might expect to find that an interaction between gender and sexual identity would result in gay and bisexual men reporting higher sexual anxiety and lower sexual esteem than any other combination of gender and sexual identity. Similarly, we might expect heterosexual men to report lower levels of sexual anxiety and higher levels of sexual esteem than heterosexual women and LGB men and women.

Two of the most discussed aspects of sexual self-concept are sexual esteem and sexual anxiety. Sexual esteem and sexual anxiety may be related to gender and sexual orientation and, more specifically, to sexual identity. For example, Snell, Fisher, and Schuh (1992) found “that there is a tendency for males to report greater sexual esteem than females” (p. 265). Considering that social norms are internalized for both heterosexual and LGB individuals, violating such norms may result in decreased sexual esteem and increased sexual anxiety for people who identify as LGB compared to heterosexual individuals. This is because, despite more recent social acceptance, historically, LGB sexual orientation has been considered to be a deviation from the social norm (Nadal, 2013). If these social norms are then internalized, LGB individuals may experience decreased sexual esteem and increased sexual anxiety. Consistent with this possibility, Dupras (1994) reported that gay men who were less accepting of their sexual orientation reported higher levels of sexual anxiety and lower levels of sexual esteem than gay men who were more accepting of their sexual orientation. Furthermore, prior investigations have indicated that gender is intricately linked to sexual orientation and sexual identity development (Martos, Nezhad, & Meyer, 2015), and sexual self-concept (Garcia, 1999). For example, Wiederman (2000) found that women reported lower levels of sexual esteem and higher levels of sexual anxiety than men. In this exploratory study, we sought to clarify whether differences in sexual esteem and sexual anxiety exist between men and women of different sexual identities.

**Sexual Esteem**

Snell (1998) described sexual esteem as “a generalized tendency to positively evaluate one’s own capacity to engage in healthy sexual behaviors and to experience one’s own sexuality in a satisfying and enjoyable way” (p. 521). Sexual esteem is different from global self-esteem and differentially predicts sexual communication in couples (Oattes & Offman, 2007). Sexual esteem has been found to positively correlate with sexual communication in that those with higher sexual esteem more readily communicate their preferences to their partners and more readily initiate satisfying sexual behaviors (Menard & Offman, 2009; Oattes & Offman, 2007). Sexual esteem has also been found to positively correlate with safer sex practices (Adler & Hendrick, 1991; Seal, Minichielo, & Omodei, 1997) and greater sexual satisfaction (Menard & Offman, 2009).

Many factors have been found to influence ratings of sexual experiences including body weight, body size, and facial attractiveness. Although Snell, Fisher, and Schuh (1992) found that men tended to report greater sexual esteem than women, Oattes and Offman (2007) found no gender differences in levels of sexual esteem; yet they reported that sexual esteem appears to fluctuate within different groups of women. Similarly, Wiederman and Hurst (1998) found that young women who reported high levels of sexual esteem were influenced only by their subjective views of attractiveness and women who reported lower levels of sexual esteem were influenced by the specific environment, social constructs, and degree of investment in their physical appearance. The authors pointed out that these factors related not only to a person’s sexual esteem, but also influence an individual’s sexual experiences. Notably, no previous research
exploring possible differences in sexual esteem among those of different sexual orientations or sexual identities could be located. Subsequently, we sought to further examine how sexual esteem may differ based on sexual identity while also re-examining the relationship between gender and sexual esteem.

**Sexual Anxiety**

Sexual anxiety is “the tendency to feel tension, discomfort, and anxiety about the sexual aspects of one’s life” (Snell, 1998, p. 521) and has been found to be inversely associated with sexual esteem (Shepler, 2012). Sexual anxiety is often associated with sexual dysfunction for both men and women, and as many as 50% of men and women have reported they are dissatisfied with sexual aspects of their lives (Laumann, Gagnon, Michael, & Michaels, 1994). Additionally, men and women attribute sexual dysfunction to attitudes toward sex, relationship satisfaction, and performance anxiety (McCabe, 2005). Sexual anxiety has also been found to explain some of the association between attachment styles and sexual functioning. For example, women with attachment anxiety were found to have increased sexual anxiety and decreased sexual satisfaction (Brassard et al., 2015).

Although some research has addressed impacts of sexual anxiety in heterosexual men and women, little empirical research has considered possible similar impacts regarding how sexual anxiety affects LGB men and women. McCabe (2005) indicated that sexual anxiety is important in both the development of sexual dysfunction and sexual identity. Therefore, it would be beneficial to determine if sexual anxiety levels are comparable among men and women of different sexual identities. Such information would be useful for researchers attempting to estimate population differences and for clinicians who are attempting to develop specific treatment plans to improve sexual esteem and reduce sexual anxiety related issues (e.g., sexual dissatisfaction and sexual dysfunction) or better understand how sexual identity and gender are related.

**Purpose**

Given the inconsistent findings regarding gender and sexual identity as it relates to sexual self-concept, we hesitated to offer specific hypotheses. Instead, we opted to explore four relationships in hopes of providing some clarification as to how sexual esteem and sexual anxiety were related to gender and sexual identity. First, due to the nature of sexual esteem and sexual anxiety, we expected to find an inverse relationship among these two variables given that higher sexual esteem is usually associated with positive sexual self-concept and that higher sexual anxiety is typically associated with negative sexual self-concept. Furthermore, we expect that this will be true regardless of gender or sexual identity. Second, we wondered whether LGB individuals would report different or similar levels of sexual esteem and sexual anxiety than their heterosexual counterparts. Given that we could not identify any research on sexual orientation or sexual identity and differences in sexual esteem and sexual anxiety, this question seemed particularly relevant to explore. Inconsistent reports of how gender is related to sexual esteem and sexual anxiety (Oattes & Offman, 2007; Snell et al., 1992) led us to explore how gender is related to sexual esteem and sexual anxiety. Finally, building on the first two points of inquiry, we considered whether an interaction effect for sexual identity and gender might be observed for sexual esteem and sexual anxiety. Specifically, we expect that the greatest level of sexual esteem will be reported by heterosexual men, followed by heterosexual women, lesbian and bisexual women, and gay and bisexual men with an inverse pattern emerging for sexual anxiety.

**Method**

**Participants**

Following approval from the Michigan School of Professional Psychology institutional review board, participants were recruited from a LGB Pride celebration in the Midwest. A total of 462 participants completed surveys. However, 18 people reported a gender identity other than man or woman, and an additional 11 failed to complete one or more of the three scales used to assess the dependent variables. Because gender was considered a grouping variable in the study design, and too few participants identified as outside of the gender binary (e.g., identified as transgender, genderqueer), data from these individuals were removed from further analysis. Furthermore, those who failed to complete the instruments used to measure the dependent variables were also excluded from analysis. In total, data from 29 individuals were excluded, resulting in analysis of data from 433 participants.

The 209 (48.30%) men who participated self-identified as gay (n = 163, 78.00%), bisexual (n = 16, 7.70%), and heterosexual (n = 30, 14.40%).
The underrepresentation of heterosexual men is likely due to the fact that data were collected at an urban LGB Pride celebration. One possible reason for the underrepresentation of bisexual men in the sample may be due to their current relationship status, For example, men who have bisexual attraction might have identified as gay if they were in a same-sex relationship at the time of participation. The 224 (51.70%) women who participated self-identified as gay/lesbian (n = 116, 51.80%), bisexual (n = 50, 22.30%), and heterosexual (n = 58, 27.80%). Although some researchers include other members of the broader sexual minority community (such as those who identify as “queer” or “intersex”), we opted not to do so in the present study due to concerns regarding statistical power. Participants self-identified as White/Euro-American (n = 262, 60.93%), Black/African American (n = 106, 24.65%), Hispanic/Latino/a (n = 16, 3.72%), Native American (n = 3, 0.70%), Bi/Multiracial (n = 34, 7.91%), Asian American/Pacific Islander (n = 5, 1.16%), and Middle Eastern (n = 3, 0.70%). Three participants (0.70%) did not report their race/ethnicity. However, because race was not required for analysis, data from these individuals was retained. Participants ranged in age from 18 to 75 years old (M = 30.00, SD = 11.82). On average, men were 32.51 years of age (SD = 12.74) and women were 27.65 years of age (SD = 10.39). Most participants (n = 257, 59.40%) reported being in a relationship.

Procedures
Masters and doctoral students in clinical psychology completed a brief training regarding how to invite participants to take part in a survey. The research team was stationed at a booth at an urban LGB Pride celebration. As celebration attendees passed by the booth, students approached the attendees, explained that they were collecting data for a study on sexuality, and invited them to complete the paper-pencil survey. No inducements were offered for participation. Attendees who opted to participate in the study were provided with a space to sit and a clipboard and pen to complete the informed consent and survey documents. The anonymous survey was composed of demographic items and standardized scales. Participation took approximately 10 minutes.

Instruments
Rostosky, Dekhtyar, Cupp, and Anderman (2008) used 8-item Likert scales to measure sexual anxiety and sexual esteem. Participants respond to items (e.g., “I worry about the sexual aspects of my life” for sexual anxiety or “I feel good about the way I express my own sexual needs and desires” for sexual esteem) using a 5-point Likert-type scale ranging from 1 = very uncharacteristic of me to 5 = very characteristic of me. Rostosky et al. reported acceptable internal reliability for both the sexual esteem (α = .90) and sexual anxiety (α = .88) scales. The mean score for our sample on the sexual esteem was 32.11 (SD = 6.37), and internal reliability was acceptable (α = .87); the mean score for sexual anxiety was 16.08 (SD = 7.04), and internal reliability was acceptable (α = .85).

Results
A Pearson bivariate correlation analysis was conducted to examine how the variables of interest (i.e., sexual esteem or sexual anxiety) were related to each other. As expected in the first hypothesis, sexual esteem was negatively related to sexual anxiety (r = -.35, p < .001). Notably, this suggests that participants likely responded meaningfully because this observation is consistent with other findings regarding the relationship between sexual esteem and sexual anxiety (Shepler, 2012). For this sample, the relationship between sexual esteem and sexual anxiety appeared stronger for men (r = -.41, p < .001) than for women (r = -.27, p < .001) and ranged from -.16 for bisexual to -.43 for heterosexual individuals. Subsequently, we examined whether the difference in the correlation strength differed based on gender. Review of Table 1 indicated that there was not a statistically significant difference for the strength of the correlation between sexual esteem and sexual anxiety for men (r = -.41, p < .001) and women (r = -.27, p < .001; or based on sexual identity (gay: r = -.37, p < .001; lesbian: r = -.40, p < .001; bisexual: r = -.16).

<table>
<thead>
<tr>
<th>TABLE 1</th>
<th>Results of Fisher's Z Test for Gender and Sexual Orientation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fisher's Z</td>
</tr>
<tr>
<td>Men, Women</td>
<td>-1.73</td>
</tr>
<tr>
<td>Gay, Lesbian</td>
<td>0.30</td>
</tr>
<tr>
<td>Gay, Bisexual</td>
<td>-1.52</td>
</tr>
<tr>
<td>Gay, Heterosexual</td>
<td>0.60</td>
</tr>
<tr>
<td>Lesbian, Bisexual</td>
<td>-1.65</td>
</tr>
<tr>
<td>Lesbian, Heterosexual</td>
<td>0.30</td>
</tr>
<tr>
<td>Bisexual, Heterosexual</td>
<td>1.83</td>
</tr>
</tbody>
</table>

SUMMER 2017
PSI CHI
JOURNAL OF PSYCHOLOGICAL RESEARCH
A 2 x 3 Multivariate Analysis of Variance was conducted to determine whether differences in sexual anxiety and sexual esteem exist between gender (man, woman) and sexual identity (gay/lesbian, bisexual, heterosexual) as predicted in the second and third hypotheses. The assumption of homogeneity of variance was not violated (Box’s $M = 19.56, p = .208$). There was not a statistically significant main effect for sexual identity on the combined dependent variables, Wilks’ Lambda = .98, $F(4, 852) = 2.01, p = .091$, partial eta squared = .01. However, results did reveal a significant multivariate main effect for gender, Wilks’ Lambda = .97, $F(2, 426) = 5.86, p = .003$, partial eta squared = .03 (see Table 2). These findings indicate that differences in sexual esteem and sexual anxiety were not found based on sexual identity independent of or in combination with gender, but that gender differences in sexual esteem and sexual anxiety were present. Specifically, review of Table 3 indicated that gender had a statistically significant effect for sexual esteem; $F(1, 427) = 3.99, p = .046$, partial eta squared = .009, and for sexual anxiety; $F(1, 427) = 10.91, p < .001$, partial eta squared = .025. Women ($M = 33.00, SD = 6.30$) reported higher levels of sexual esteem than men ($M = 31.30, SD = 6.40$), and men ($M = 17.14, SD = 7.20$) reported higher levels of sexual anxiety than women ($M = 15.10, SD = 6.80$; see Table 4). Contrary to our fourth hypothesis, results indicated that there was not a statistically significant interaction effect between gender and sexual identity on the combined dependent variables, Wilks’ Lambda = 1.00, $F(4, 852) = .56, p = .692$, partial eta squared = .00.

**TABLE 2**

<table>
<thead>
<tr>
<th>Effect or Interaction</th>
<th>Wilks’s Lambda</th>
<th>$F$</th>
<th>Hypothesis df</th>
<th>Error df</th>
<th>$p$</th>
<th>Partial Eta Squared</th>
<th>Noncentrality Parameter</th>
<th>Observed Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>.973</td>
<td>5.86</td>
<td>2</td>
<td>426.00</td>
<td>.003*</td>
<td>.027</td>
<td>11.724</td>
<td>.873</td>
</tr>
<tr>
<td>Sexual Orientation</td>
<td>.981</td>
<td>2.01</td>
<td>4.0</td>
<td>852.00</td>
<td>.091</td>
<td>.009</td>
<td>8.041</td>
<td>.605</td>
</tr>
</tbody>
</table>

Note. $N = 433, \alpha = .05$.

**TABLE 3**

<table>
<thead>
<tr>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>$F$</th>
<th>$p$</th>
<th>Partial Eta Squared</th>
<th>Noncentrality Parameter</th>
<th>Observed Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sexual Identity</td>
<td>SE</td>
<td>53.67</td>
<td>2</td>
<td>26.84</td>
<td>0.67</td>
<td>.514</td>
<td>1.33</td>
</tr>
<tr>
<td>SA</td>
<td>268.66</td>
<td>2</td>
<td>134.33</td>
<td>2.80</td>
<td>.062</td>
<td>.013</td>
<td>5.59</td>
</tr>
<tr>
<td>Gender</td>
<td>SE</td>
<td>160.93</td>
<td>1</td>
<td>160.93</td>
<td>3.99</td>
<td>.046*</td>
<td>.009</td>
</tr>
<tr>
<td>SA</td>
<td>524.43</td>
<td>1</td>
<td>524.43</td>
<td>10.91</td>
<td>.001*</td>
<td>.025</td>
<td>10.91</td>
</tr>
<tr>
<td>SI * Gender</td>
<td>SE</td>
<td>7.79</td>
<td>2</td>
<td>3.89</td>
<td>0.10</td>
<td>.908</td>
<td>.000</td>
</tr>
<tr>
<td>SA</td>
<td>85.63</td>
<td>2</td>
<td>42.81</td>
<td>0.89</td>
<td>.411</td>
<td>.004</td>
<td>1.78</td>
</tr>
</tbody>
</table>

Note. $N = 433, \alpha = .05$. SI = sexual identity.

**TABLE 4**

| Means and Standard Deviations for SE and SA Based on Sexual Identity and Gender |
|-------------------------------|-------------------------------|-------------------------------|
| Sexual Orientation | Gender | Total |
| Gay/Lesbian ($n = 279$) | Bisexual ($n = 66$) | Heterosexual ($n = 88$) | Men ($n = 209$) | Women ($n = 224$) | ($n = 433$) |
| Sexual Esteem | $M$ (SD) | $M$ (SD) | $M$ (SD) | $M$ (SD) | $M$ (SD) | $M$ (SD) |
| Sexual Anxiety | $M$ (SD) | $M$ (SD) | $M$ (SD) | $M$ (SD) | $M$ (SD) | $M$ (SD) |
| Gay/Lesbian | 32.16 (6.50) | 32.47 (5.70) | 31.70 (6.50) | 31.27 (6.37) | 32.90 (6.29) | 32.11 (6.37) |
| Bisexual | 16.01 (6.92) | 17.61 (7.90) | 15.16 (6.63) | 17.14 (7.18) | 15.08 (6.77) | 16.08 (7.04) |
Discussion
In this study, we explored whether sexual esteem and sexual anxiety levels were comparable among men and women with different sexual identities. To do this, we examined whether people who identify as LGB would report different levels of sexual esteem and sexual anxiety compared to heterosexual individuals as well as whether differences exist in sexual esteem and sexual anxiety based on gender. Sexual esteem and sexual anxiety are considered to be major components of sexual self-concept with importance for sexual satisfaction and relationship satisfaction (Birnbaum et al., 2006; Brassard et al., 2015). Similar to previous research (Shepler, 2012), sexual anxiety and sexual esteem were found to be inversely related. This finding, although not original or unique to the present study, importantly confirms the relationship between these two aspects of sexual self-concept.

We explored whether LGB individuals would report different or similar levels of sexual esteem and sexual anxiety than their heterosexual counterparts, and we found that LGB individuals reported comparable sexual esteem and sexual anxiety. One possibility for this finding is that no actual differences exist in sexual esteem and sexual anxiety based on sexual identity. Alternatively, results may be more limited in generalizability. Although no effort was made to assess level of sexual orientation acceptance in our study, it is worth noting that Dupras (1994) found that gay men who were less accepting of their sexual orientation reported higher levels of sexual anxiety and lower levels of sexual esteem. Future research should seek to determine (a) whether such differences can be replicated and (b) if replicable, are there differences in sexual esteem and sexual anxiety among heterosexual individuals, high sexual orientation accepting LGB individuals, and low sexual orientation accepting LGB individuals.

We also sought to explore how gender is related to sexual esteem and sexual anxiety. Findings showed that women reported significantly higher levels of sexual esteem compared to men, and men reported significantly higher levels of sexual anxiety than women. These results suggest that there are distinct differences in how men and women report and experience sexual esteem and sexual anxiety. Contrary to our findings, Wiederman (2000) reported that men reported higher levels of sexual esteem and lower levels of sexual anxiety than women while Oattes and Offman (2007) found no gender differences in sexual esteem. This contradiction requires further research to clarify how gender is related to sexual esteem and sexual anxiety. It seems likely that a third, untested variable may mediate the relationship between gender and aspects of sexual self-concept such as sexual esteem and sexual anxiety. One construct that may be especially important to consider as a possible mediator is sexual identity development. Women who have a more coherent and integrated sexual identity make healthier decisions in their sex lives and therefore experience a greater sense of sexual well-being (Hucker, Mussap, & McCabe, 2010). Furthermore, women who developed a sense of their sexual self-concept during adolescence are less reserved with their sexuality, less anxious about their sex life, and experience more comfort with their sexuality as they transition into adulthood (Hensel, Fortenberry, O’Sullivan, & Orr, 2011). Potentially, women who attend Pride may have more developed sexual self-concepts or have a tendency to have a different developmental trajectory than women who do not attend events like Pride. This may account for the gender-based differences in sexual esteem and sexual anxiety that we observed in the present study. Subsequently, future studies are needed to compare how representative these findings are of non-Pride attendees. Indeed, some (Bailey et al., 1999; Bailey et al., 2016) have questioned the generalizability of findings based on data collected from convenience sampling at LGB Pride events.

We also considered whether an interaction effect for sexual identity and gender might be observed for sexual esteem and sexual anxiety. However, no interaction effect was observed on sexual esteem and sexual anxiety for sexual identity and gender. This means that the main effect for gender is likely consistent across those of different sexual identity groups. Subsequent research may benefit from clarifying if this finding is robust for those who are at different levels of sexual identity development.

Strengths and Limitations
Collecting our data from Pride over the course of one weekend allowed for several benefits to the study. For example, our sample size was relatively large and more racially diverse compared to many other studies, allowing for greater generalizability. Also, by collecting data over one weekend, the impact of history and maturation effects were minimized. This seems especially important given the rapid changes in LGB rights and equality that were
Exploring Sexual Self-Concept

Shepler, Johnson, and Ho

One possibility is that those who attend such events may be, on average, more comfortable and accepting of their sexual orientation than those in the broader LGB population. Although similarity in sexual self-concept appears likely for individuals who are more achieved in their sexual identity development (regardless of sexual identity), this may not be true across differing levels of sexual identity development. For example, those of different sexual identity development statuses may not have similar levels of sexual esteem and sexual anxiety in earlier, comparable stages of sexual identity development. Nonetheless, our findings are important because much of what is known about sexual self-concept has been based on either samples of individuals with sexual disorders or adolescents/young adults. Future research may investigate sexual esteem and sexual anxiety across various levels of sexual identity development, measuring participants’ identity development status and sexual orientation acceptance.

Unlike previous researchers (Snell et al., 1992) who reported that men tended to report higher levels of sexual esteem than women, we found that women reported higher levels of sexual esteem (and lower levels of sexual anxiety) than men. Although results may accurately characterize gender-based differences, they may alternatively be due to a sexual identity-specific gender effect (i.e., an interaction) because the sample was composed predominately of those who identified as gay and lesbian. Unfortunately, due to the small number of heterosexual and bisexual individuals in the sample, the observed power for testing the interaction was considerably too low. Future researchers should seek to examine this possibility due to the implications for theory development. For example, if an interaction effect exists, this would indicate that it is important to not make broad characterizations about sexual anxiety and sexual esteem for men or women, but that care should be taken to specifically discuss these concepts for men and women who identify differently in terms of sexual identity. Another limitation of the present study is that a more completed analysis of the relationship between gender and the independent variables was not possible due to the small number of participants with transgender and other gender identities. Despite concerns regarding statistical power that limited our ability to explore the possibility of interaction effects, some significant differences were observed and results add to the discussion regarding sexuality and gender.

Conclusion

This study is one of the first to address a gap in the literature concerning possible interactions among sexual identity, gender, and sexual esteem and sexual anxiety in a community-based sample and subsequently expands our understanding of sexual self-concept. Results suggest that gender differences in sexual esteem and sexual anxiety are present although sexual identity group membership may not be related to sexual esteem and sexual anxiety. This is consistent with other research that has highlighted the role of gender as a major contributing factor to understanding individuals’ sexual satisfaction, relationship satisfaction, and overall sexual self-concept. Individuals’ sexual self-concepts are important factors in determining sexual satisfaction (Birnbaum et al., 2006) and relationship satisfaction (Brassard et al., 2015). Findings subsequently have implications for how those of different sexual identities and genders may experience their romantic relationships. Although further research is needed to confirm and extend the complexity of our understanding of how gender and sexual identity are related to sexual anxiety and sexual esteem, our findings provide direction for future researchers committed to examining such constructs. For example, findings leave open the possibility that within group differences in sexual esteem and sexual anxiety may be present for those who identify as LGB and future research is needed to clarify how sexual identity development status or self-acceptance of sexual orientation may be related to sexual esteem and sexual anxiety.

References


Because deception detection skills are highly valued and critical for many careers, especially those relating to law enforcement and corrections (Wang, Chen, & Atabakhsh, 2004), the field of psychology has contributed a great deal of research into attempting to identify individual differences in deception detection ability (Aamodt & Custer, 2006; Bond & DePaulo, 2008). Although meta-analysis examining individual differences in deception detection has not identified many definitive predictors of accurate deception detection ability (Aamodt & Custer, 2006), that is not to say that individual differences in deception detection do not exist.

For example, Albrechtsen, Meissner, and Susa (2009) discovered that people who only watched a small portion (thin slice) of a video clip, and presumably had to rely on their intuition to make veracity judgments, were more accurate at detecting deception than their counterparts who made veracity judgments after watching the entire video clip. Aamodt and Custer’s (2006) research indicated that the personality trait self-monitoring may be an individual difference influencing deception detection. The keen ability of self-monitors to observe the world around them and the behaviors of others may make them more adept at detecting deception.

ABSTRACT. Although meta-analysis has revealed that individual differences in deception detection ability do exist, the relationship between personality traits and deception detection ability has not been as heavily researched (Aamodt & Custer, 2006). The Big Five model of personality is often used to investigate personality differences in deception detection ability (Elaad & Reizer, 2015; John & Srivastava, 1999). Elaad, Reizer, and Hirschberg (2006) found significant relationships between deception detection and openness to experience, agreeableness, and extraversion, respectively. The current study predicted that high levels of openness to experience, agreeableness, and extraversion would correlate with accuracy in a video clip deception detection task. The study was administered to 228 undergraduate students from a midwestern university, and yielded no significant correlations between overall accuracy on the deception detection task and openness to experience ($r = .05$, $p = .47$), agreeableness ($r = .01$, $p = .87$), or extraversion ($r = .01$, $p = .85$). An independent-sample $t$ test revealed that participants exhibited a significant truth bias, $t(220) = 5.66$, $p < .001$, $d = .54$. The concept of truth bias is explored through Levine’s (2014) Truth Default Theory. Explanations for the lack of significant findings and other methodological issues are also addressed. Due to the complex nature of deception, further inquiries could investigate different types of deception detection tasks. Future research could also explore differences in attitudes and beliefs related to deception in general among college students because personality traits may not serve as a reliable predictor of deception detection ability.

Examine Personality Factors in Deception Detection Ability

Samuel D. Spencer
Minnesota State University, Mankato

Emily Stark is the faculty mentor.
These are but a few examples of the many individual differences that have been proposed to impact deception detection ability. However, the research examining personality traits as they relate to deception detection ability is limited (Aamodt & Custer, 2006), which indicates that more research on this topic is needed to identify potential attributes of an effective lie detector. The present study chose to focus on Big Five personality traits as a differentiator of deception detection ability because several of the Big Five traits, such as extraversion and agreeableness, focus on how an individual interacts with others, which may impact that individual’s deception detection ability. The Big Five taxonomy also appears to be the most commonly used taxonomy in the few studies that have explored the relationship between deception detection and personality (Elaad & Reizer, 2015; Hirschberg et al., 2006). Although limited in number and significant findings, these previous studies allow for guidance in hypothesis development and lend credence to the use of the Big Five taxonomy in deception detection research.

The Big Five Model of Personality
Throughout the latter half of the 20th century, there has been an increasing amount of research conceptualizing, describing, and measuring personality traits (John, Naumann, & Soto, 2008; McCrae & Costa, 2004). However, the plethora of different theories that have arisen has led to some confusion and disagreement in regard to which conceptualization best encapsulates personality (John et al., 2008). To clarify the many different approaches to personality, researchers have used factorial analysis and other lexical approaches to search for independent, overarching categories to simplify the conceptualization of personality (Cattell, 1945; Tupes & Christal, 1961). As more results were validated over the years, factor analysis continued to point in the direction of a conclusion that most of the words used to describe personality fall into five distinct categories.

These groups are collectively referred to as the Big Five taxonomy, with the traits of openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism all existing on a continuum and being measured by participant self-report surveys. For example, an individual scoring high in extraversion may exhibit an increased amount of social status in leadership positions, and an individual scoring lower in this trait may have poorer quality of relationships and speak up less in conversation. However, most individuals will fall somewhere in the middle of these continuums (John et al., 2008). Through the proliferation and validation of the Big Five model, research on personality traits and the predictions that can be inferred from their accurate measurement has continued to improve.

To further elucidate the Big Five model of personality, a brief description of the traits is warranted. Openness to experience involves discovering novel activities and approaches to problems, and is often described as an association with learning new things. John et al. (2008) conceptualized openness to experience as “describing the breath and complexity of an individual’s mental and experiential life” (p.120). The consensus regarding the description of openness to experience is that people with high levels of this trait tend to exhibit intellectual curiosity and the desire to pursue novelty and variety (McCrae & Costa, 1997).

Some of the traits that have been shown to comprise extraversion include talkativeness, enthusiasm, assertiveness, and energetic behavior (John et al., 2008). McCrae and Costa (1997) posited that people who exhibit high levels of agreeableness tend to be trusting of other people, and are usually very sympathetic and attuned to the thoughts, feelings, and actions of others. Neuroticism is generally associated with negative emotions such as anxiety, low self-confidence, irritability, and worrying. Individuals who score high in conscientiousness tend to be efficient, organized, thorough, and deliberate, and often pride themselves in adherence to societal norms and rules (John & Srivasta, 1999; McCrae & John, 1992).

One of the most important benefits of the Big Five taxonomy is that it allows access to an empirical method of quantifying personality differences among individuals with relative ease (John et al., 2008; McCrae & John, 1992). This ease of use, combined with the excellent construct validity the Big Five taxonomy has been shown to exhibit, lends credence to the use of this measure, which can usually be administered in less than 15 minutes (John et al., 2008). As more data continues to be collected from Big Five surveys, the collection of reliability and validity evidence supporting these measures will continue to grow.

Although accuracy rates of deception detection ability have typically been shown to be no greater than chance (Aamodt & Custer, 2006), research continues to look for traits or individual differences that may correspond to greater accuracy.
The streamlined administration and established psychometric properties of the Big Five has made this conceptualization of personality a good fit with deception detection research (Elaad & Reizer, 2015) because the few studies comparing personality to deception detection almost exclusively relied on the Big Five model.

**Big Five Taxonomy Used in Deception Detection Research**

The Big Five taxonomy’s conceptualization of personality, which has connections to interpersonal relationships and general outlook on life, may be a good fit for differentiating deception detection ability. This could be due, in part, to the importance of interpersonal relationships and outlook on life in deception detection, which involves making judgments of veracity in individuals. Although there has not been a plethora of research investigating the relationship between personality traits and deception detection, and much of what does exist is inconclusive (Aamodt & Custer, 2006), the extant literature has revealed a small number of studies that have compared Big Five personality traits to deception detection, which are highlighted below.

Hirschberg et al. (2006) found a positive correlation between accuracy in a deception detection task and openness to experience, and further postulated that high levels of openness to experience correlate with a greater level of intelligence and the ability to use critical thinking in discerning truth from lies. Openness to experience has been positively associated with emotional intelligence, which is an integral part of determining veracity through recognition of emotions and nonverbal communication (McIntyre, 2010). Self-described deception detection ability has also been shown to have a positive relationship with openness to experience (Elaad & Reizer, 2015), indicating that this may be a trait that is related to deception detection.

Extraversion has been extensively linked to social interaction and lying in general. Kashy and Depaulo (1996) found that extraverts tended to tell more lies than their introverted counterparts, based, in part, on the function that smaller white lies can serve in smoothing social interactions. This additional exposure to lying may cause extraverts to become more adept at telling and detecting lies. Furthermore, Elaad and Reizer (2015) found a correlation between extraversion and self-reported deception detection ability. Caution must be taken, however, when evaluating the results of self-report deception detection because confidence in accuracy does not always equate to actual accuracy regarding deception detection (Aamodt & Custer, 2006).

Hirschberg et al. (2006) also concluded that a positive relationship existed between agreeableness and deception detection ability. This conclusion was based on the premise that, because people exhibiting high levels of agreeableness tend to be more sensitive to the behaviors and attitudes of others, they would be able to discern truth from lies more effectively than those who are less agreeable. Although generalized trust, and to some degree, agreeableness, have been shown to predict a higher propensity to believe a story as true, regardless of its actual veracity (Carter & Weber, 2010; Elaad & Reizer, 2015), the relationship between agreeableness and actual deception detection ability remains unclear because this is a topic that has remained relatively untouched.

There are contrasting opinions regarding the relationship between neuroticism and deception detection. Some research has hypothesized that, because individuals scoring high in neuroticism may have a hard time making decisions, they tend to report low levels of deception detection ability (Elaad & Reizer, 2015). Others have speculated that high levels of neuroticism may predispose individuals to judge a statement as truthful because individuals scoring high in neuroticism may not be as well-equipped to handle the ramifications of being deceived (Hirschberg et al., 2006).

Many of the studies linking deception detection to Big Five personality traits do not mention conscientiousness (Hirschberg et al., 2006). These unsupported hypotheses and lack of definitive conclusions (Elaad & Reizer, 2015) suggest that the associations between deception detection and conscientiousness and neuroticism respectively are not as strong as those between the other three traits (extraversion, openness to experience, and agreeableness) and deception detection.

Beyond actual deception detection ability, a growing body of research has indicated that people may exhibit a truth bias. This truth bias suggests that people have a propensity to believe a statement is true, regardless of its actual veracity (Truth Default Theory, see Levine, 2014). This theory may relate to certain personality traits such as agreeableness and conscientiousness, although previous research has not yet examined these relationships.

**Deception Detection Tasks**

People are confronted with deception in their daily
lives in a wide variety of uncontrolled settings, and there are many extraneous variables affecting the outcome of any given situation involving deception. These factors make deception detection difficult to measure. Fortunately, video clip studies provide a scientifically sound experimental design that controls some of these extraneous variables and allows for comparison between subjects. In an experiment similar to the current study, Carter and Weber (2010) had participants watch eight video clips and make veracity judgments of each person speaking. By having all participants watch the same video clips, which were categorized as truth or lie, Carter and Weber’s design allowed for comparison between subjects on deception detection task scores, and provided a foundation on which this current study is based. Sweeney and Ceci’s (2014) deception detection study also found that, when compared to visual or written modalities, audiovisual modalities allowed for the most precise ratings of deception detection.

**Summary**

Although previous meta-analysis has not revealed a strong connection between personality traits and deception detection (Aamodt & Custer, 2006), a recently growing body of literature has been examining Big Five personality traits as they relate to deception detection ability (Elaad & Reizer, 2015; Hirschberg et al., 2006). The Big Five taxonomy has a potential link to deception detection ability due to the relationship that both concepts have with evaluation of an individual’s outlook on life as well as understanding interpersonal relationships. It certainly seems that further research is warranted into this subject, with the end goal of being able to identify traits or characteristics that may be able to serve as a differentiator between accurate and inaccurate lie detectors.

**Current Study**

The current study evaluated Big Five personality traits as a manner of distinguishing individual differences in deception detection ability. Although previous studies have attempted this comparison using a self-report measure of deception detection ability (Elaad & Reizer, 2015) and through a task using verbal language in simulated interviews to identify deception (Hirschberg et al., 2006), to the best of our knowledge, no research has ever investigated the relationship between Big Five personality traits and deception detection using a short video clip task.

By using a video clip deception detection task and an established Big Five Inventory (BFI; John et al., 2008), this design maximizes the potential for finding significant results, as well as minimizes the effects of confounding variables. It was predicted that participants’ levels of agreeableness, openness to experience, and extraversion would have a positive correlation with accuracy scores on the video clip deception detection task. In addition, it was hypothesized that participants would be more likely to choose truth over lie, regardless of the actual veracity of the story, and that agreeableness and conscientiousness would correlate positively with the frequency that truth is chosen.

**Method**

**Participants**

Participants were college-aged students recruited from a Midwestern university. Participants received extra credit points in their undergraduate psychology class in exchange for participation in this study. Of the 228 people who participated, 53 were men, 173 were women, and two did not respond to this question. The sample was mostly comprised of 116 European American individuals, 28 identified as African American, seven identified as Asian/Vietnamese/Hmong, seven identified as mixed/other, and two did not specify ethnicity. Participants’ ages ranged from 17 to 38, but 88.2% of the participants were in the 18 to 22 age group (M = 20.41, SD = 2.65). In regard to collegiate experience, 90 participants were first-year students, 37 were sophomores, 39 were juniors, 57 were seniors, and five classified themselves as “other.”

**Materials**

**Personality inventory.** Because the original BFI is very extensive and time consuming for participants to complete, John et al. (2008) created a short form of the BFI, which consists of 44 questions that are a valid and balanced representation of the five main personality dimensions. Participants rate each of the 44 items on a 5-point Likert scale from 1 (disagree strongly) to 5 (agree strongly).

Based on the results of the current study, internal consistency was calculated. The eight items that measured extraversion were combined to form a summed total extraversion score (α = .85). Similarly, the nine items comprising agreeableness were summed to form a total agreeableness score (α = .74). The eight items measuring conscientiousness were combined to form a total conscientiousness score (α = .77). The eight items
Personality Factors in Deception Detection

Spencer

measuring neuroticism were similarly combined to form a total neuroticism score ($\alpha = .79$). Finally, the 10 questions comprising openness to experience were also summed to create a total openness to experience score ($\alpha = .75$).

The internal consistency levels for the current study were similar to the results of John and Srivastava’s (1999) reliability testing, which yielded a mean alpha score of $\alpha = .83$. Further validity testing also showed that the BFI exhibits evidence of convergent validity with other measures of personality such as the Neuroticism, Openness, Extraversion Five Factor Inventory, and the Trait Descriptive Adjective test (John & Srivastava, 1999).

**Video clips.** As part of the current study, participants also watched eight video clips of undergraduate student volunteers telling a story. Four of the stories were false, and four of them were true. The true and false stores were presented in a random order. All participants watched the same eight video clips, in the same order. Participants rated each video as either true or false. Each participant’s total number of accurate judgments was combined to form an overall truth selection score. For a complete listing of descriptive statistics pertaining to the measures used, please see Table 1.

### Procedure
All aspects of this experiment were approved by Minnesota State University, Mankato’s institutional review board prior to the beginning of data collection (IRB Approval # 797429-2). Once selected for the study, participants watched eight separate video clips depicting a true or false story, and then made veracity judgments following each clip. After participants finished watching the video clips, they filled out a brief demographic section and a short form of the BFI.

### Results
Throughout the data analysis process, the hypotheses mentioned earlier were considered. Specifically, bivariate correlational analysis was used to evaluate relationships between personality traits and overall accuracy scores. Each participant’s average of truth selections was calculated and compared to what would be expected by chance using a one-sample $t$ test. Bivariate correlational analysis was also used to evaluate relationships between personality traits and overall truth scores.

#### Big Five Traits and Deception Detection Task Accuracy
Although this study hypothesized a correlation between openness to experience and overall accuracy scores, the results of the correlation indicated no significant relationship ($r = .05, p = .47$). Similarly, the hypothesis that agreeableness would correlate with accuracy was also not supported ($r = .01, p = .87$). The hypothesis that extraversion would correlate with accuracy scores was not supported either ($r = .01, p = .85$).

#### Accuracy and Frequency of Truth Selection
Initial analysis revealed that the average overall accuracy rate of all participants in this deception detection study was slightly above chance (58%). The frequency of participants’ rating of truth (versus lie) for each video were averaged and summed to form an average frequency truth rating for all participants ($M = 4.42, SD = 1.08$). When compared to what would be expected by chance ($M = 4$) using a one-sample $t$ test, participants showed a significant preference toward choosing truth, $t(220) = 5.66, p < .001, d = .54$.

#### Big Five Traits and Frequency of Truth Selection
Participants’ personality scores were correlated with overall frequency of truth ratings, and no significant results were found for any of the five traits (extraversion: $r = .06, p = .37$; agreeableness: $r = .01, p = .98$; conscientiousness: $r = .04, p = .52$; neuroticism: $r = .02, p = .82$; and openness to experience: $r = .01, p = .97$).

### Table 1

<table>
<thead>
<tr>
<th>Trait</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Openness to experience</td>
<td>3.25</td>
<td>0.60</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>3.73</td>
<td>0.60</td>
</tr>
<tr>
<td>Extraversion</td>
<td>3.35</td>
<td>0.78</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>3.96</td>
<td>0.54</td>
</tr>
<tr>
<td>Neuroticism</td>
<td>2.98</td>
<td>0.74</td>
</tr>
<tr>
<td>Accuracy in lie videos (/4)</td>
<td>2.09</td>
<td>0.81</td>
</tr>
<tr>
<td>Accuracy in truth videos (/4)</td>
<td>2.48</td>
<td>0.77</td>
</tr>
<tr>
<td>Overall accuracy (/8)</td>
<td>4.57</td>
<td>1.14</td>
</tr>
<tr>
<td>Overall frequency of truth</td>
<td>4.42</td>
<td>1.08</td>
</tr>
</tbody>
</table>
Discussion

Unfortunately, meta-analysis exploring the relationship between personality traits and deception detection ability has not produced conclusive findings to date. However, researchers have continued to investigate this topic, due to the shared emphasis that Big Five traits and deception detection place on interpersonal relationships and worldview. Some studies (Elaad & Reizer, 2015; Hirschberg et al., 2006) have found significant relationships between Big Five personality traits and deception detection ability. Additionally, the current study did not find similar significant results, indicating that Big Five traits may not be an effective differentiator of deception detection ability. Potential explanations for the lack of significant findings as well as limitations and directions for future research are presented in this section.

Methodological Limitations

The aforementioned lack of significant findings warrant an analysis of power to rule out the possibility that a small sample size led to an inability to detect significant findings. A program called G*Power 3.1.9.2 was used to conduct power analyses, with an alpha cutoff of .05 used for all analyses. Even if a hypothetical weak effect (\( r = .3 \)) existed, a sample size of 228 (the size of the sample in this study) would yield a beta value of .99, indicating that this sample size would be adequate to detect a small effect, assuming one did exist. Additionally, an effect size of .02 (the average correlation between Big Five traits and deception detection ability), would require a sample size of 19,620 to adequately detect an effect (\( \beta = .8 \)). These power estimates reveal that sample size was not a contributor to the lack of significant findings.

It is also noteworthy that a significantly greater number of women (173) participated in the study, compared to men (55). Previous literature has shown that women are more likely to respond to online research surveys than men (Smith, 2008). The reasons behind this gender disparity are not specifically clear, but it is nonetheless important to be aware of this potentially confounding variable. However, this was not a major concern for the current study because gender was not a primary variable of interest.

The type of deception detection task used might have played a role in the lack of significant findings. In Elaad and Reizer’s (2015) and Hirschberg et al.’s (2006) research, which were similar to the current study in regard to comparing Big Five traits to deception detection ability, different measures of deception detection were used (self-report and reviewing interview transcripts) compared to the short video clips used in the current study. These methodological differences might have contributed to the lack of significant findings, suggesting that accurate deception detection may involve more than just making a judgment based on a short video clip.

Another issue arising with the use of video clip deception detection tasks is that they may not induce the requisite motivation that real-life liars possess. Martin and Leach (2012) argued that traditional video clip deception detection tasks lack generality due to storytellers not being properly motivated to tell a realistic lie. Some research has included a monetary incentive to produce more realistic lie stories (Hirschberg et al., 2006). Monetary incentives for successful evasion of deception detection could allow for greater authenticity of the stories used in the video clip task.

Although previous research has indicated that individual differences in deception detection ability are relevant (Bond & DePaulo, 2008; Carter & Weber, 2010), Big Five personality traits do not appear to be an effective and reliable differentiator. This may be due, in part, to the difficulty of simulating the nuances and intricacies of real-life deception detection in a laboratory setting. The lack of significant findings in this study may also be attributed to the variable nature and expression of personality in general. Because manipulation is difficult to achieve in this type of research design (one cannot “assign” someone to a personality trait), it is impossible to rule out the effects of other causal variables. For example, Bond and DePaulo noted that sender credibility, among other factors, is very influential of the outcome of a deception detection scenario. Clearly, there are many different variables pertaining to deception detection ability, and personality may not play as large of a role as originally expected.

Truth Default Theory

Regardless of the veracity of the story, the data analyses performed did indicate that participants’ propensity to choose truth over lie in the video clip task was significantly greater than what would be expected due to chance. This data falls into line with Levine’s (2014) Truth Default Theory, and offers support for the theory that people may be more likely to believe a story, regardless of the veracity. Some evidence has suggested that
generalized trust and agreeableness may relate to accuracy in deception detection (Carter & Weber, 2010; Hirschberg et al., 2006). Although data analysis did not find a significant correlation between any of the Big Five traits and overall truth score, future research could expand on this idea by exploring the relationship between attitudes and beliefs about deception in general to generalized trust and agreeableness.

Although the results of the one-sample t test comparing truth selection frequency to what would be expected due to chance were significant, the relatively moderate effect size of .54 indicates that the relationship was not as robust as previous research has shown. Levine (2014) indicated that the truth bias depicted in Truth Default Theory may not be as strong when people are engaged in situations where they are primed to expect deception. Because participants were told that they were going to be taking part in a deception detection study, this might have put them “on guard,” so to speak. Caution should also be taken when interpreting these findings of a truth bias due to the confounding nature of the personality traits of participants in this study. People who tend to be more agreeable also tend to have a truth bias (Kashy & Depaulo, 1996). Future research may want to look at the relationship between Big Five personality traits and truth bias to further generalize these conclusions.

Implications for Future Research

One intriguing direction that future research could pursue is whether experience in college (e.g., first-year, sophomore) influences deception detection ability. Investigating the development of personality, social values, and cognitions as manifested through the progression into adulthood may also be of interest for the field of deception detection research. Although personality generally remains stable in adulthood, adolescence and early adulthood can often be a time of changing values, traits, and cognitions. Studying these developmental shifts in the form of comparing younger college students (first-year students) to older ones (seniors) regarding deception detection ability and truth bias would be a worthwhile continuation of this line of research.

Given the generally high level of interest in this line of research, and the readily available population of college students, future research may want to move away from an examination of the relationship between personality and deception detection ability, and instead focus on beliefs, attitudes, and perceptions about deception in general. Understanding how beliefs and attitudes about deception change over time and investigating the differences in beliefs between younger and older adolescents would be of particular interest.

However, just as deception detection research is conducted in the contrasting experiences of lab studies and real-life scenarios, cutting-edge research techniques and innovative ideas continue to blend the differences between these two approaches, and will hopefully allow researchers to gain a better understanding of the complex subject of deception detection. In summary, the current study’s lack of significant findings suggest that Big Five personality traits may not be a reliable predictor of deception detection ability. However, the significant findings related to a truth bias indicate that more research in needed concerning beliefs and attitudes about deception and how those relate to truth bias.

References


https://doi.org/10.1177/0261927X14535916  
https://doi.org/10.1002/pmh.1215  
http://dx.doi.org/10.1037/0003-066X.52.5.509  
https://doi.org/10.1016/S0191-8869(03)00118-1  
https://doi.org/10.1111/j.1467-6494.1992.tb00970.x  
http://dx.doi.org/10.1016/j.paid.2009.12.019  
http://scholarworks.sjsu.edu/cgi/viewcontent.cgi?article=1003&context=elementary_ed_pub  
http://doi.org/10.3389/fpsyg.2014.00590  
http://doi.org/10.1023/B:GRUP.0000021838.66662.0c

Author Note. Samuel D. Spencer, Department of Psychology, Minnesota State University, Mankato.  
This research was funded in part through a grant from the Undergraduate Research Center at Minnesota State University, Mankato. Special thanks to Dr. Emily Stark for all her guidance and wisdom, and to all *Psi Chi Journal* reviewers for their support.  
Correspondence concerning this article should be addressed to Samuel D. Spencer, Department of Psychology, Minnesota State University, Mankato, 103 Armstrong Hall, Mankato, MN 56001. E-mail: Samuel.spencer@mnsu.edu.
The Relationship Between Religiousness and Friendship Quality

Namele Gutierrez, Elizabeth J. Krumrei-Mancuso*, and Cindy Miller-Perrin*
Pepperdine University

ABSTRACT. The aim of the present study was to examine a possible relationship between religiousness and friendship quality through relational theory. We hypothesized that higher levels of religiousness would be positively correlated with friendship quality. Young adults (n = 118) from a small, Christian university participated in an online survey. We observed a main effect for religiousness of the participant and religiousness of the best friend being associated with greater support and depth in friendship, as well as an interaction between participants’ religiousness and best friends’ religiousness in regard to friendship support. The results indicated that participants’ best friends’ religiousness was a significant, positive predictor of support in friendships (r^2 = .08). In addition, an interaction was observed between participants’ religiousness and their best friends’ religiousness in predicting support (r^2 = .05). For participants low in religiousness, having a friend with higher levels of religiousness was associated with more support and depth in friendship, but for participants high in religiousness, having a friend with higher levels of religiousness was not associated with friendship qualities. Because friendship is instrumental in young adults’ development, it is important to evaluate religiousness as a possible factor that can positively impact friendships at this life stage.

As a function of frequent socialization away from home, young adulthood in the United States becomes a time in which young men and women are affected by friendships with peers (Cheadle & Schwadel, 2012; Desrosiers et al., 2011). Horizontal relationships with friends become increasingly instrumental in young adults’ individual development, which can provide benefits to an individual’s overall well-being (Cheadle & Schwadel, 2012; Desrosiers et al., 2011). For example, friendships that provide a high degree of support and intimacy, and meet other psychological needs are related to happiness, health, well-being, and self-esteem (Demir & Davidson, 2013; Olson & Perl, 2011). Given the benefits of friendship, it is important to examine factors that can potentially enrich the quality of friendships including the level of support, conflict, and depth within a friendship (Pierce, Sarason & Sarason, 1991).

Although research on friendship in adulthood is limited, scientific support exists for the idea that friendships play important roles in influencing people’s sense of self, skills, and priorities in life (Anye, Gallien, Bian, & Moulton, 2013; Demir & Davidson, 2013; Flora, 2013; Sawatzky, Gadermann, & Pesut, 2009). To examine the relationship between religiousness and friendship quality in young adults, this study utilized relational theory as its foundational basis. Relational theory, proposed by Hill and Pargament (2008), was based on Martin Buber’s writings on the connection between relationship and religion (Simpson, Newman & Fuqua, 2008). According to relational theory, an individual’s religious pursuit of the sacred fundamentally includes the pursuit of relationship, not only in hopes of creating a connection with God (vertical relationship), but also a connection with others (horizontal relationship; Hill & Pargament, 2008; Simpson et al., 2008). Within this framework, how individuals personally relate to and maintain a relationship with God impacts their approach toward interpersonal relationships.
and vice versa (Desrosiers, Kelley, & Miller, 2011; Simpson et al., 2008). It suggests that optimal human functioning involves an interconnectedness between the psychological and spiritual, and that compartmentalizing these aspects of life negates the benefits of taking a more holistic approach. Relational theory focuses on how religion can act as a guide to form, sustain, and shape people’s closest relationships. Essentially, the theory describes a pursuit of the sacred as it relates to pursuing intimate interpersonal relationships (Simpson et al., 2008). Therefore, this study aimed to understand how young adults’ relationship with God (vertical relationship) connects with their relationship with their friends (horizontal relationship).

Relational Functioning on the Horizontal and Vertical Level

One potential factor that can contribute to friendship quality is religion. Although there has been a general lack of agreement among researchers in regard to defining religiousness, the present study defined the construct as an individual’s pursuit of the sacred within well-established communities and traditions that incorporate common beliefs and practices (Davis, Hook, Van Tongeren, Gartner & Worthington, 2012). Within this definition, the sacred is comprised of a connection with God or a dedication to a religious purpose. Therefore, this study aimed to understand how young adults’ relationship with God (vertical relationship) connects with their relationship with their friends (horizontal relationship).

Past research has proposed a parallel process between how individuals think, feel, and behave in relationship with others and how they do so with God (Hall, 2007). When individuals express their relationship dynamic with God, it not only provides insight into their past relationship dynamics, but also into their “current...expectations of emotionally significant relationships” (Hall, 2007, p. 17). This has generally been understood and demonstrated through attachment theory (Ainsworth & Bowlby, 1991). For example, those with secure attachment, consciously or unconsciously expect that those who are emotionally significant to them will be available, responsive, open to sharing positive and negative emotion, and sincerely take interest in them. Similarly, those with secure attachment will expect the same from their relationship with God and approach their relationship with God in the same way (Hall, 2007). On the other hand, individuals with a preoccupied attachment expect that both others and God will be unreliable, unstable, and unavailable (Hall, 2007). Those with preoccupied attachment tend to respond with desperate, clingy or help-seeking forms of prayer, just as they also use these responses within their relationships with others in order to achieve closeness (Hall, 2007). Moreover, those with fearful attachment style tend to view others as uncaring, hostile, or rejecting, and view themselves as unlovable or unworthy (Hall, 2007). Even though those with this attachment style long for close relationships, they tend to avoid them because they may feel that they cannot seek support from emotionally significant others. In a similar way, they may want a close relationship with God, but they feel unable to achieve such a relationship. Therefore, they may be less willing to share their pain through prayer to God just as they would be less willing to share their pain through verbal communication with others (Hall, 2007). However, individuals who do not identify as securely attached are not only more likely to experience a religious conversion and find new meaning in their relationship with God during adolescence or adulthood, but also self-report a greater increase in religiosity over time (Hall, 2007; Kirkpatrick, 1998).

Connection Between Friendship and Religiousness

Past research has extensively examined the role of religiousness within romantic relationships (McCurry, Schrodt & Ledbetter, 2012; Reiter & Gee, 2008) as well as family relationships (Breelsford & Mahoney, 2008) and found that religiousness has been positively associated with open communication, connection, and support, and can possibly mitigate feelings of relational uncertainty and distress (Breelsford, Marinelli, Ciarrochi, & Dy-Liacco, 2009; McCurry et al., 2012; Reiter & Gee, 2008). Although findings have suggested religion’s positive impact on romantic and familial relationships, much less research has examined the positive role of religiousness in friendships (Mahoney, 2010).

The few studies that have assessed the relationship between religiousness and friendship have provided insightful trends. Both conflict resolution and forgiveness have been emphasized in the literature that assesses religion’s impact within interpersonal relationships. In a study that assessed 615 adolescents and young adults with a broad range of ethnicities and religious affiliations, forgiveness and empathy were suggested to be essential to openly speaking about religious and spiritual questions in
Religiousness and Friendship Quality | Gutierrez, Krumrei-Mancuso, and Miller-Perrin

friendships, which seemed to deepen friendships (Davis et al., 2012; Desrosiers et al., 2011). In addition, positive links have also been found between personal measures of religion such as individual religious practices and experiences, and spiritual disclosure and relationship satisfaction in friendships (Brelsford et al., 2009; Sawatzky et al., 2009). This result suggests a possible relationship between individuals who identify as religious and the level of friendship quality they experience because they may be more able to disclose and discuss religious values and ideas with their close friends. A study conducted among well-educated African American men also confirmed this finding (Mattis et al., 2001). After controlling for education, income, age, and advice offering, the results of Mattis et al. (2001) indicated that men’s subjective spirituality positively affected how supportive they viewed their same-sex friends.

The Current Study

Based on previous research on marriages, families, and friendships, the relational theory framework helps address the influence of religiousness on relationships. Religiousness has been linked with both conflict resolution and supportive, deep relationships. Thus far, the research that has assessed the positive association between religiousness and friendship has done so only indirectly by assessing social or community networks, overall quality of life, or the specific ability to forgive or have spiritual disclosure. Therefore, the current study focused on expanding understanding about how religiousness relates to friendship quality directly. In addition, it was a goal of the current study to consider levels of religiousness of participants’ best friends in addition to their own levels of religiousness. We hypothesized that individuals’ religiousness, their ratings of their best friends’ religiousness, and the interaction between both would be positively correlated to friendship quality defined as high social support and depth of values and ideas in friendship, and low amounts of conflict experienced in the friendship.

Method

Participants

Young adults (18–32 years old; M = 19.17, SD = 1.74) from a small, Christian university in southern California participated. The sample consisted of 118 students (67.80% women, 33.20% men). The racial and ethnic self-identifications were as follows: 51.7% identified as non-Hispanic White or Euro-American, 22.9% as Asian/Pacific Islander, 9.3% as Latino or Hispanic, 5.9% as African or African American, 5.1% as Other, 3.4% as Middle Eastern or Arab American, 0.8% as Native American or American Indian, and .8% preferred not to respond. The religious self-identifications were as follows: 86.4% identified as Christian, 6.8% as Atheist, 2.5% as Agnostic, 1.7% as Muslim, 1.7% as Jewish, and .8% as Mormon.

Materials

Demographic Questionnaire. Demographic information was gathered, including age, sex, racial group, and religious affiliation.

Religiosity Measure. This study involved participants rating their own religiousness and their best friends’ religiousness with the Religiosity Measure (Aalsma et al., 2013), a self-report measure consisting of four items, each rated on a 3-point Likert-type scale (not important, important, or very important). Example items include “How important is it to you to rely on religious teachings when you have a problem?” and “How important is it to you to believe in God?” Higher scores indicate greater religiousness (range = 4–12). The internal consistency has been reported as adequate (α = .83; Aalsma et al., 2013). This 4-item measure was completed twice by participants, once about themselves and once about their best friend. This approach was based on previous research, which has shown the quality of self versus other ratings of both internal and external characteristics such as intellect, positivity, and sociability to be fairly to highly accurate, meaning that, when other people rate someone on a personal characteristic, they rate that person closely to how the person rates themselves (Connelly & Ones, 2010; Kenny, 1994; Rauthmann et al., 2014). For the current study, internal consistency was good for both participants’ religiosity (4 items; α = .92) and their best friends’ religiosity (4 items; α = .93).

The Quality of Relationships Inventory (QRI). The QRI (Pierce et al., 1991) is a self-report measure used to assess relational quality through 25 items, each rated on a 5-point scale from 1 (not much) to 5 (a great deal). The questionnaire was divided into three subscales: Support (8 items, α = .85; range = 16–37), Conflict (9 items, α = .91; range = 9–37), and Depth (8 items, α = .84; range = 17–39). Support involves the ability to offer approval, comfort, or encouragement to another person. An example item is “To what extent could you count on this person for help with a problem?”
Conflict relates to disagreements, arguments, or a clash of opposing needs or wishes. An example item is “How critical of you is this person?” Depth is associated with the complexity and profundity of thoughts or ideas able to be discussed. An example item is “To what extent can you count on this person to give you honest feedback, even if you might not want to hear it?” Items from the conflict subscale are reverse scored so that higher scores indicate greater friendship quality (range = 25–125). For the current study, internal consistency for all subscales was adequate: Support (7 items; \( \alpha = .82 \)), Conflict (12 items; \( \alpha = .76 \)), and Depth (6 items; \( \alpha = .80 \)).

**Procedure**

This study was approved by Pepperdine University’s institutional review board. A convenience sample of participants from a small, Christian university was recruited. Participants accessed the study through the university’s online survey system that requires students’ assigned login information in order to gain access. All participants provided informed consent online prior to completing any measures. First, participants completed demographic information. Participants then completed the Religiousness Measure via self-report. They were then asked to respond to the Religiousness Measure according to how they would expect their best friend to respond to the questionnaire. Participants then completed the QRI. There were no missing data. As compensation, participants were given research credit for a psychology class.

**Results**

**Preliminary Analyses**

Descriptive analyses were conducted on all measures (see Table 1). Religiousness scores ranged from 4 to 12 among both participants’ ratings of themselves and their best friends. Participants’ self-reported levels of religiousness were relatively high (\( M = 9.48, SD = 2.56 \)) as were the best friends’ levels of religiousness (\( M = 8.63, SD = 2.92 \)). Relationship quality scores ranged from 52 to 110. Relationship quality scores were also relatively high (\( M = 80.06, SD = 9.54 \)). In general, friends experienced high degrees of support (\( M = 31.71, SD = 3.76 \)) and depth (\( M = 31.65, SD = 4.47 \)), while maintaining low levels of conflict (\( M = 16.69, SD = 5.60 \)).

Several analyses were conducted to examine whether demographic factors (i.e., age, sex, race) were significantly related to individuals’ religiousness, their ratings of their best friends’ religiousness, friendship quality, and the three subscales of Support, Depth, and Conflict. First, correlation analyses indicated that age was not significantly correlated with participant’s self-rated religiousness (\( p = .12 \)), participant’s best friend’s religiousness (\( p = .19 \)), friendship quality (\( p = .77 \)), the Support subscale (\( p = .25 \)), the Depth subscale (\( p = .51 \)), or the Conflict subscale (\( p = .42 \)). Second, an independent-samples t-test indicated that women scored higher on religiousness (\( M = 9.86, SD = 2.32 \)) than men (\( M = 8.65, SD = 3.00 \), \( t(116) = 2.44, p = .02 \)). Finally, a multivariate analysis of variance indicated that none of the racial or ethnic groups differed significantly from one another on participant’s self-rated religiousness (\( p = .47 \)), participant’s best friend’s religiousness (\( p = .64 \)), friendship quality (\( p = .23 \)), the Support subscale (\( p = .63 \)), or the Conflict subscale (\( p = .15 \)). On the basis of these analyses, sex was the only demographic variable controlled in the primary analyses.

Correlation analyses were conducted to examine direct links between predictor and outcome variables. Participants' self-rated religiousness was correlated with their ratings of their best friends’ religiousness (\( r = .43, p < .001 \)). Support in friendships was correlated with participants’ religiousness (\( r = .22, p = .01 \)) and their best friends’ religiousness (\( r = .28, p = .002 \)). Depth in friendships was correlated with the participants’ religiousness (\( r = .29, p < .01 \)) and their best friends’ religiousness (\( r = .33, p < .01 \)). Conflict in friendships was not correlated with either the participants’ religiousness (\( p = .08 \)) or the best friends’ religiousness (\( p = .26 \)).

**TABLE 1**

<table>
<thead>
<tr>
<th>Variable</th>
<th>SR</th>
<th>BFR</th>
<th>QRI</th>
<th>Support</th>
<th>Conflict</th>
<th>Depth</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>SR</td>
<td>.428**</td>
<td>.196</td>
<td>.129</td>
<td>.206*</td>
<td>.330**</td>
<td>.764**</td>
<td>.068</td>
</tr>
<tr>
<td>BFR</td>
<td>.129</td>
<td>.206*</td>
<td>.694**</td>
<td>.330**</td>
<td>.103</td>
<td>.764**</td>
<td>.068</td>
</tr>
<tr>
<td>QRI</td>
<td>.216*</td>
<td>.283*</td>
<td>.158</td>
<td>.300**</td>
<td>.584**</td>
<td>.103</td>
<td>.075</td>
</tr>
<tr>
<td>Support</td>
<td>-158</td>
<td>-103</td>
<td>-1.00</td>
<td>-.293**</td>
<td>-.103</td>
<td>-.106</td>
<td>-.061</td>
</tr>
<tr>
<td>Conflict</td>
<td>.393**</td>
<td>.300**</td>
<td>.764**</td>
<td>.129</td>
<td>.120</td>
<td>.110</td>
<td>.37</td>
</tr>
<tr>
<td>Depth</td>
<td>.141</td>
<td>.120</td>
<td>-.026</td>
<td>-.478**</td>
<td>-.026</td>
<td>.37</td>
<td>.39</td>
</tr>
<tr>
<td>Age</td>
<td>4</td>
<td>12</td>
<td>110</td>
<td>.42</td>
<td>.37</td>
<td>.37</td>
<td>.32</td>
</tr>
</tbody>
</table>

Note. SR = self-reported religiousness. BFR = best friend’s religiousness. QRI = Friendship Questionnaire. Support, conflict, and depth are the subscales of the QRI. **Correlation is significant at the .01 level (2-tailed). *Correlation is significant at the .05 level (2-tailed).
Hierarchical Regressions

Hierarchical regression analyses were used to examine the effect of participants' religiousness, their best friends' religiousness, and the interaction between these two variables on support and depth in friendships, while controlling for sex (see Table 2 and Table 3). The results indicated that participants' best friends' religiousness was a significant, positive predictor of support in friendships ($\beta = .27, p = .006; r^2 = .08$). In addition, an interaction was observed between participants' religiousness and their best friends' religiousness in predicting support ($\beta = -1.28, p = .009; r^2 = .05$). With regard to depth of friendship, only participants' best friends' religiousness was a significant predictor ($\beta = .28, p = .003; r^2 = .12$).

Post hoc probing of the significant interactions was conducted by testing the significance of separate regression lines for individuals high (1 SD above mean) versus low (1 SD below the mean) on religiousness after centering all variables to reduce multicollinearity. First, an interaction was observed for best friends' religiousness moderating the relationship between the participants' religiousness and the support in a friendship (see Figure 1). For participants low in religiousness, the slope for the regression line was: support = 31.97 + 5.97 * (best friends' religiousness), $t(114) = 3.74, p < .001$, indicating that participants experienced significantly greater support in their friendships when they rated their best friends as being higher in religiousness. However, for those participants high in religiousness, the slope of the regression line was: support = 32.27, $t(114) = -0.02, p = .99$, indicating that there was no significant relationship between best friends' religiousness and friendship support for participants high in religiousness. Thus, it seems that having a best friend who is higher in religiousness is associated with greater friendship support for individuals who themselves are low in religiousness, but not for individuals who

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SE(B)</th>
<th>$\beta$</th>
<th>$\Delta F$</th>
<th>$\Delta R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td>2.53**</td>
<td>.71</td>
<td>.31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 2</td>
<td></td>
<td></td>
<td></td>
<td>5.54**</td>
<td>.08</td>
</tr>
<tr>
<td>Participants' Religiousness</td>
<td>.05</td>
<td>.14</td>
<td>.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Best Friends' Religiousness</td>
<td>.34**</td>
<td>.12</td>
<td>.27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interaction Between Participants' and Best Friends' Religiousness</td>
<td>-.12**</td>
<td>.04</td>
<td>-1.28</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 1: Interaction Between Participants' Self-Religiousness and Their Best Friends' Religiousness on Support in the Friendship.

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SE(B)</th>
<th>$\beta$</th>
<th>$\Delta F$</th>
<th>$\Delta R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td>2.92**</td>
<td>.85</td>
<td>.30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 2</td>
<td></td>
<td></td>
<td></td>
<td>8.46**</td>
<td>.12</td>
</tr>
<tr>
<td>Participants' Religiousness</td>
<td>.20</td>
<td>.17</td>
<td>.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Best Friends' Religiousness</td>
<td>.43**</td>
<td>.14</td>
<td>.28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interaction Between Participants' and Best Friends' Religiousness</td>
<td>-.07</td>
<td>.05</td>
<td>-0.61</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 2: Interaction Between Participants' Self-Religiousness and Their Best Friends' Religiousness on Depth in the Friendship.
Similarly, the relationship between best friends’ religiousness and depth in a friendship differed significantly between participants with high versus low levels of religiousness (see Figure 2). The slope for those low in religiousness was: depth = .39 (best friends’ religiousness) + 2.90, t(114) = 2.90, \( p = .004 \), indicating that participants experienced significantly greater depth in their friendships when their best friends were higher in religiousness. On the other hand, the slope for those high in religiousness was: depth = .13 (best friends’ religiousness) + 32.55, t(114) = 1.01, \( p = .31 \), indicating that there was no significant relationship between best friends’ religiousness and friendship depth for participants high in religiousness. Thus, it seems that greater religiousness among best friends was associated with greater depth in relationships for individuals low in religiousness, but not for individuals high in religiousness.

**Discussion**

The current study’s objective was to examine a possible association between religiousness and friendship quality on the basis of relational theory. According to relational theory, an individual’s religious pursuit of the sacred fundamentally includes the pursuit of relationship, not only in hopes of creating a connection with God (vertical relationship), but also a connection with others (horizontal relationship; Hill & Pargament, 2008; Simpson et al., 2008). Within this framework, how individuals personally relate to and maintain a relationship with God impacts their approach toward interpersonal relationships (Desrosiers et al., 2011; Simpson et al., 2008). We hypothesized that individuals’ religiousness, their ratings of their best friends’ religiousness, and the interaction between both would be positively correlated to friendship quality defined as high social support and depth of values and ideas in friendship, and low amounts of conflict experienced in the friendship. The results partially supported the hypotheses.

Participants’ ratings of their best friends’ religiousness significantly related to level of support and depth in a friendship, which is consistent with past research indicating that a friend’s religiousness can be positively linked to support in friendships (Brelsford et al., 2009; Mattis et al., 2001; Pierce et al., 1991). In addition, we observed a significant interaction between participants’ religiousness and their friends’ religiousness in relation to support within a friendship. Interestingly, it seems that higher levels of religiousness among a best friend is associated with better friendship quality for participants who themselves are low in religiousness, but not for those who themselves are high in religiousness. It may be that those who are higher in religiousness tend to gain or seek more support from God or the divine (vertical relationship), although those who are low in religiousness tend do so from others such as their friends (horizontal relationships). No links were observed between religiousness and levels of conflict in relationships. This may be because mean levels of conflict were low within the sample overall, not leaving much room for improvement on the basis of religiosity.

On the basis of this study, we can conclude that religiousness or participants’ vertical relationship was associated with friendship quality or one of their horizontal relationships. It is possible that religiousness can strengthen friendships. For example, it may be that individuals who rate their best friends with higher levels of religiousness experience more support and depth in their friendships because these friends are more open to talking about deeper topics or caring for them during times of need, or they may simply perceive their friendships as such. Direct causality cannot be established on the basis of this correlational study. An alternative explanation for the findings is that religious values may overlap with values within friendships. For example, a religious value may be supporting loved ones, which a best friend usually does. Participants, then, may have been reporting on overlapping religious values and general friendship values such as love, connection, and treating others as they would like to be treated.

There should be caution in generalizing this study’s findings due to the fact that the study’s sample was a convenience sample of college students attending a small, private Christian university. With the overwhelming majority of students being Christian and religiously inclined, the study did not provide insight into the links between religiousness and friendship among individuals who do not identify as religious or who are of another religion. Participants’ self-reported levels of religiousness were relatively high (\( M = 9.4, SD = 2.5 \)) as were ratings of the best friends’ levels of religiousness (\( M = 8.6, SD = 2.9 \)). Another limitation of this study was that the measure used to assess religiousness was relatively short. A lengthier measure might have provided a more nuanced understanding of participants’ and their best friends’ religiousness. Religiousness has often been measured in a variety
Religiousness and Friendship Quality

Gutierrez, Krumrei-Mancuso, and Miller-Perrin

of ways, and having a measure that addresses both internal beliefs and outward action might have been better able to account for how individuals understand or express their religiousness. Additionally, the surveys used were self-report measures, which can sometimes be problematic when participants provide answers that are socially desirable.

Future research should consider replicating this study with a population that has larger religious variability, whether in the United States or internationally. In addition, it may be insightful to compare individuals at different ages, and to examine if changes occur in the relationship between religiosity and friendship quality throughout various developmental stages. Future research may also assess the level of relational satisfaction directly because it may be that those higher in relational satisfaction have different experiences of relational quality when religiosity is a mediating variable. Despite future research needed and the limitations of the current study, the conclusions add to the limited, but growing, body of knowledge of the impact of religiousness on interpersonal functioning, specifically within friendships.

References


Author Note. Namele Gutierrez, Pepperdine University; Elizabeth Krumrei Mancuso, Social Science Division, Pepperdine University; Cindy Miller-Perrin, Social Science Division, Pepperdine University.

Special thanks to all Psi Chi journal reviewers for their support.

Correspondence concerning this article should be addressed to Namele Gutierrez. E-mail: namele.gutierrez@gmail.com
A romantic relationship develops, partners contemplate when they might have sexual intercourse for the first time (i.e., sexual debut). Approximately 40% of heterosexual partners in committed romantic relationships have sex within the first month of dating (Busby, Carroll, & Willoughby, 2010; Peplau, Rubin, & Hill, 1977; Sassler, Addo, & Lichter, 2012; Willoughby, Carroll, & Busby, 2014). Expectations about the timing of first sexual experience with a new romantic partner are consistent with this research. Cohen and Shotland (1996) reported that college students expect sexual debut to occur within 2 months, on average, of the start of the relationship, although men expected to have sex approximately 4 weeks earlier than women. Given the relative rapidity with which sexual debut occurs within adult relationships and the documented sex differences in expectations of sexual debut, it is important to examine how sexual debut is associated with relationship outcomes for men and women. Thus, the overarching purpose of the present study was to examine whether (a) sexual debut correlated with relationship and sexual satisfaction, (b) this association was comparable for men and women, and (c) sexual debut predicted the duration of an already failed romantic relationship.

Busby et al. (2010) summarized two competing models that explained relationship outcomes associated with sexual debut. According to the sexual restraint model (Metts, 2004), sexual intercourse is potentially harmful to the development of a young relationship. If a couple has intercourse too early, the relationship may be built on sexual
companionship instead of emotional compatibility. Partners who delay their sexual debut can be more assured that their relationship is based on emotional connection rather than sexual chemistry. By contrast, the sexual compatibility model (Cassell, 2008) states that sexual interaction is necessary for the relationship’s early development because it helps romantic partners explore their sexual and emotional compatibility. During sexual intercourse, partners have the opportunity to express their needs and determine whether their partner can fulfill those needs. According to this model, partners who have an early sexual debut will be more satisfied because they have determined that their partner can meet their sexual and emotional needs.

The extant research examining the relationship between sexual debut and relationship satisfaction supports the sexual restraint model. Peplau et al. (1977) found that partners who waited to have sexual intercourse for at least a month after their first date reported higher relationship satisfaction than those with sexual debuts within a month. Other researchers reported that cohabitating or married couples with sexual debut within the first 2 months of their dating relationship experienced lower relationship satisfaction over time than couples who delayed intercourse (Busby et al., 2010; Sassler et al., 2012). Further, Willoughby et al. (2014) reported that delaying sexual intercourse by “a few weeks” was associated with higher relationship satisfaction in young adults in dating relationships.

However, such associations may not be comparable for heterosexual men and women. Indeed, women report that their first sexual experience in a relationship has a more positive impact on their relationship than for men, supporting the notion that women link emotional intimacy with sexual intimacy (Cate, 1993; Leigh, 1989). Although some researchers have found that delaying sexual intercourse is associated with increases in relationship and sexual satisfaction in married men and women (Bubsy et al., 2010), other researchers have reported that these associations are specific to women (Sassler et al., 2012). Further, Peplau and colleagues (1977) reported that sexual debut was not associated with men’s sexual satisfaction, but women who delayed sexual debut reported less sexual satisfaction and more guilt than women with earlier debut.

Previous investigations comparing relationships between sexual debut, relationship satisfaction, and sexual satisfaction across men and women are limited in three respects. First, although it appears that relationship and sexual satisfaction are more closely linked to sexual debut in women (Peplau et al., 1977; Sassler et al., 2012), much of the previous research has been conducted with adults in long-standing relationships who are cohabitating or married (Busby et al., 2010; Sassler et al., 2012). There is reason to believe that these patterns may not generalize to traditional age college students in dating relationships. For example, Varga (1997) found comparable sexual satisfaction scores between men and women in college dating situations, whereas Sprecher (2002) found that married and cohabitating women reported higher sexual satisfaction than their male partners. Although Willoughby et al. (2014) examined dating couples, the majority of respondents were in their later twenties, 15% of the sample reported having one or more children from previous relationships, and most couples were likely moving toward marriage. As such, the sexual restraint theory has not been adequately tested in traditionally aged college students in shorter-term dating relationships.

Second, sexual debut has traditionally been defined as the point in the relationship that vaginal or anal intercourse occurs for the first time. To our knowledge, no research has been conducted on sexual debut for oral sex. Indeed, college students’ attitudes about oral and vaginal/anal intercourse differ: Whereas 90% of college students identified vaginal and anal sex as being sexual intercourse, only 60% of college students considered oral sex to be sexual intercourse (Byers, Henderson, & Hobson, 2009). According to the Centers of Disease Control and Prevention (2012), about a fourth of 15- to 24-year-old men and women living in the United States reported that oral sexual debut occurred before vaginal sexual debut. Comparable percentages of young men and women reported vaginal debut occurring before oral debut. Given that oral debut occurs just as frequently before or after vaginal debut, the type of sexual debut may not differentially predict relationship outcomes.

Third, the majority of work on sexual debut has been conducted with young adults in an existing romantic relationship. However, if sexual debut predicts relationship and sexual satisfaction in current relationships, it may also predict the length of a failed relationship. Some indirect evidence supports this prediction. In their meta-analysis, Le and colleagues (2010) reported a moderate association between higher relationship satisfaction and longer relationship duration in married
couples. Further, Willoughby et al. (2014) found that relationship length moderated the association between sexual debut and relationship satisfaction in young, unmarried adults’ romantic relationships. Adults whose sexual debut occurred within the first few weeks of dating reported disproportionately lower relationship satisfaction after 1 year of dating than did adults who delayed intercourse. In as much as relationship satisfaction predicts relationship dissolution, these findings suggest that early sexual debut would predict shorter romantic relationships.

In the present study, we examined whether different forms of sexual debut (oral, vaginal/anal) predicted relationship and sexual satisfaction in undergraduate women and men who were and were not currently in a relationship. In addition to addressing gaps in previous research, findings from this study could have important implications for professionals who counsel young people about romantic relationships and sexual intercourse such as clinicians or high school guidance counselors. Given that most college students do not consider oral sex intercourse (Byers et al., 2009), these findings could support the belief that oral sex is qualitatively distinct from vaginal/anal sex or could demonstrate that the potential impact of oral sexual debut is comparable to that of vaginal/anal sexual debut. Such information could help young people make informed choices about becoming sexually intimate with their romantic partner.

Consistent with the sexual restraint model (Busby et al., 2010; Metts, 2004), we expected that young adults in romantic relationships with later sexual debut (i.e., 2 or more months) would report higher relationship satisfaction and higher sexual satisfaction than those in romantic relationships with earlier sexual debuts (i.e., less than 2 months). This pattern of findings was expected for both vaginal/anal and oral sex. Further, given the conflicting findings regarding sex differences in sexual debut, relationship satisfaction, and sexual satisfaction (Leigh, 1989; Sassler et al., 2012; Sprecher, 2002; Varga, 1997), it was unclear whether sex differences would emerge in the current study. Following from Willoughby et al. (2014), we expected that later sexual debut would be associated with longer relationship duration.

Method

Participants
Participants consisted of 229 undergraduates from a public regional university in the northeast who were enrolled in a psychology course that offered research participation as a curricular component. Most participants were taking introductory level psychology courses and represented a broad cross-section of the student population. Students earned course credit for their participation. Participants who had never been in a romantic relationship and those who had never had sexual intercourse were excluded from the study. An additional 37 participants were excluded for exceeding 25 years of age (n = 8), having an open relationship (n = 7), cohabitating with their partner (n = 19), or being married to their partner (n = 1). Two participants were excluded for response bias or not answering four or more items on the survey. Thus, subsequent analyses were based on 194 participants.

Average participant age was 19.80 years (SD = 1.46). Most participants were women (67.2%) and were in a monogamous relationship at the time of the study (58.3%). Participants identified as non-Hispanic White (71.9%), African American (10.4%), Latino/a American (9.9%), Asian American (2.6%), and other (5.2%). Participants identified their sexual orientation as being heterosexual (93.2%), gay (1.6%), and other (5.2%). Participants reported being in their current or most current relationship for an average of 21.77 months (SD = 18.01, Mdn = 18). All participants who were not in a relationship at the time of testing had broken up with their partner within the previous year. On average, participants were 16.20 years old (SD = 1.75) when they first experienced oral sex and 16.39 (SD = 1.67) years of age when they first experienced vaginal or anal sex. Further, participants reported oral sexual debuts of 4.49 months within their relationship (SD = 5.77, Mdn = 2), and vaginal/anal sexual debuts of 4.82 months (SD = 6.07, Mdn = 3).

Procedure
This study was reviewed by the institutional review board and deemed exempt. Participants were tested in person using paper-and-pencil methods and were, when possible, asked to sit with at least one seat in between each other to ensure privacy. Participants were informed that the purpose of the study was to examine college students’ romantic relationships. After signing consent forms, participants placed them in a specially designated folder separate from their responses to subsequent measures. Thus, participant identity could not be linked to responses on the questionnaire.

Next, participants received a stapled
questionnaire containing the measures in the following order: demographics questionnaire, Relationship Assessment Scale (RAS; Hendrick, 1988), Relationship Events Scale (RES; King & Christensen, 1983), and the Index of Sexual Satisfaction (ISS; Fischer & Corcoran, 1994). Because the demographics questionnaire required participants to report sexual activity, we presented the relationship satisfaction measure next to reduce the likelihood of participants linking sexual activity to sexual satisfaction. Further, to minimize the chances of participants directly associating relationship satisfaction with sexual satisfaction, we included the RES as an intervening filler measure.

Upon completing the questionnaires, participants received a written debriefing informing them that the true purpose of the study was to examine sexual debut and satisfaction in relationships. We provided contact information for on-campus services so participants could discuss their romantic relationship with a trained professional. We also requested that participants refrain from discussing the study with other potential participants. The procedure lasted about 15 minutes.

Measures
Demographics questionnaire. Participants reported their age, biological sex, sexual orientation, and race/ethnicity. They also reported their current relationship status, the length of their current or former relationship, the age they first had sexual intercourse (oral, anal, or vaginal), and at what time in their current or former relationship they engaged in sexual intercourse (oral, anal or vaginal). If participants were not in a relationship at the time of testing, they reported how long ago the dissolution occurred.

Relationship Assessment Scale (RAS; Hendrick, 1988). The RAS measures romantic relationship satisfaction (e.g., “How well does your partner meet your needs?”) using seven questions that are rated on a 5-point Likert-type scale from 1 (poorly) to 5 (extremely well). The internal consistency of the RAS is high with a Cronbach’s $\alpha$ of .86. In addition, the RAS features excellent concurrent and predictive validity. In the current study, interim reliability analyses revealed high interitem consistency (Cronbach’s $\alpha$ = .91). Responses for all items were averaged into a composite score spanning from 1 to 5 with higher numbers reflecting greater satisfaction.

Relationship Events Scale (RES; King & Christensen, 1983). The RES was used as a filler questionnaire to reduce demand characteristics. It is a 19-item true-or-false questionnaire used to assess progress in a romantic relationship through “milestones,” such as if participants have told each other “I love you” or if they are engaged to be married. This questionnaire was not scored or used in any analyses.

Index of Sexual Satisfaction (ISS). The ISS, developed by Walter Hudson (as cited and reproduced in Fischer & Corcoran, 1994) is a 25-item scale measuring feelings of sexual satisfaction in a romantic relationship (e.g., “I feel that sex is a normal function of our relationship”). Participants rate items on a 7-point Likert-type scale from 1 (none of the time) to 7 (all of the time). The ISS features a high internal consistency of .92, high test-retest reliability, and high concurrent validity. In the current study, the scale exhibited excellent internal consistency (Cronbach’s $\alpha$ = .94). All items were summed, subtracted by the number of items that were completed, and multiplied by 100. This number was then divided by the number of completed items multiplied by 6. Thus, scores could range from 0 to 100 with higher numbers indicating greater sexual satisfaction.

Results
Descriptive Statistics
The wide variability in sexual debut precluded examining sexual debut as a continuous variable. Thus, we grouped participants using a 2-month cutoff: those who engaged in first vaginal/anal intercourse at or before 2 months ($n = 91$) and those who waited after 2 months for first intercourse ($n = 102$). We also divided participants similarly based on oral sexual debut: 91 participants engaged in first oral sex at or before 2 months ($n = 91$) and those who waited after 2 months for first intercourse ($n = 102$). We also divided participants similarly based on oral sexual debut: 91 participants engaged in first oral sex at or before 2 months and 95 waited 2 or more months (7 participants did not report oral sexual debut). The 2-month cutoff created roughly equivalent sexual debut groups and is similar to the cutoff for early and late sexual debut used in previous literature (Busby et al., 2010; Peplau et al., 1977; Willoughby et al., 2014; but see Sassler et al., 2012, who grouped participants by first month, between 1 and 6 months, and more than 6 months).

We examined whether our early and late sexual debut groups were comparable across demographic characteristics that could be related to sexual or relationship satisfaction (see Tables 1 and 2). Of note, there was a slight, marginally higher proportion of women reporting later oral sexual debut than men ($\chi^2(1) = 3.49, p = .062, \text{Cramer’s } \nu =$
### TABLE 1

<table>
<thead>
<tr>
<th>Demographics of Participants With Early and Late Oral Sexual Debut</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total</strong></td>
</tr>
<tr>
<td>Number of Participants</td>
</tr>
<tr>
<td>Biological Sex (% Female)</td>
</tr>
<tr>
<td>Race/Ethnicity (% Racial/ Ethnic Minority)</td>
</tr>
<tr>
<td>Sexuality (% Sexual Minority)</td>
</tr>
<tr>
<td>Age in Years</td>
</tr>
<tr>
<td>Length of Relationship (in months)</td>
</tr>
<tr>
<td>Age of First Oral Intercourse</td>
</tr>
<tr>
<td>Age of First Vaginal/Anal Intercourse</td>
</tr>
<tr>
<td>Oral Sexual Debut in Relationship (in months)</td>
</tr>
<tr>
<td>Vaginal/Anal Sexual Debut in Relationship (in months)</td>
</tr>
</tbody>
</table>

*Note:* Standard deviations in parentheses. Missing data included length of relationship (n = 2) and oral sexual debut in relationship (n = 7).

### TABLE 2

<table>
<thead>
<tr>
<th>Demographics of Participants With Early and Late Vaginal/Anal Sexual Debut</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total</strong></td>
</tr>
<tr>
<td>Number of Participants</td>
</tr>
<tr>
<td>Biological Sex (% Female)</td>
</tr>
<tr>
<td>Race/Ethnicity (% Racial/ Ethnic Minority)</td>
</tr>
<tr>
<td>Sexuality (% Sexual Minority)</td>
</tr>
<tr>
<td>Age in Years</td>
</tr>
<tr>
<td>Length of Relationship (in months)</td>
</tr>
<tr>
<td>Age of First Oral Intercourse</td>
</tr>
<tr>
<td>Age of First Vaginal/Anal Intercourse</td>
</tr>
<tr>
<td>Oral Sexual Debut in Relationship (in months)</td>
</tr>
<tr>
<td>Vaginal/Anal Sexual Debut in Relationship (in months)</td>
</tr>
</tbody>
</table>

*Note:* Standard deviations in parentheses. Missing data included length of relationship (n = 2).

### TABLE 3

<table>
<thead>
<tr>
<th>Demographics of Participants Who Were Single or in a Romantic Relationship</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total</strong></td>
</tr>
<tr>
<td>Number of Participants</td>
</tr>
<tr>
<td>Biological Sex (% Female)</td>
</tr>
<tr>
<td>Race/Ethnicity (% Racial/ Ethnic Minority)</td>
</tr>
<tr>
<td>Sexuality (% Sexual Minority)</td>
</tr>
<tr>
<td>Age in Years</td>
</tr>
<tr>
<td>Length of Relationship (in months)</td>
</tr>
<tr>
<td>Age of First Oral Intercourse</td>
</tr>
<tr>
<td>Age of First Vaginal/Anal Intercourse</td>
</tr>
<tr>
<td>Oral Sexual Debut in Relationship (in months)</td>
</tr>
<tr>
<td>Vaginal/Anal Sexual Debut in Relationship (in months)</td>
</tr>
</tbody>
</table>

*Note:* Standard deviations in parentheses. Missing data included length of relationship (n = 2) and oral sexual debut in relationship (n = 7).
.137], but this marginal difference was not mirrored vaginal/anal sexual debut \( \chi^2(1) = 2.19, p = .139, \) Cramer’s \( \nu = .107 \]. Participants reporting early oral sexual debut within their relationships also reported experiencing oral sex for the first time at a younger age than those reporting later sexual debut in their relationships, \( t(184) = 2.52, p = .013, d = .38 \). This pattern also held true across early and late vaginal/anal sexual debut groups; participants with earlier vaginal/anal debuts within a relation reported earlier first oral sexual debuts than participants who waited longer to have vaginal/anal sex within their relationship, \( t(187) = 2.36, p = .019, d = .34 \). Participants with earlier vaginal/anal debuts within their relationship were significantly, but less than 1 year older than participants with later vaginal/anal debuts, \( t(161.18) = 2.78, p = .006, d = 1.33 \). Not surprisingly, participants who initiated both oral or vaginal/anal sex earlier in their relationship reported being in that relationship for less time than participants who had waited to have intercourse. Sexual debut within a relationship and the duration of said relationship are partially intercorrelated; if one waits 8 months to have intercourse, that relationship must survive at least 8 months.

Demographic characteristics for participants who were and were not in a romantic relationship at the time of testing are reported in Table 3. Participants in a relationship reported relationships being, on average, 5 months longer than participants who reported on a previous relationship. Participants’ ages at the time of their first vaginal or anal sexual experience did not differ across relationship status, \( t(190) = 0.42, p = .673, d = 0.06 \). Participants who were single at the time of testing reported significantly earlier oral sex experience than participants who were in a relationship, \( t(187) = 2.57, p = .011, d = 0.38 \). In addition, participants who were single reported significantly earlier sexual debut in their previous relationship than did participants who were currently dating (vaginal/anal sex: \( U = 3413.50, z = 2.83, p = .005 \); oral sex: \( U = 3309.00, z = 3.11, p = .002 \)).

Sex Differences in Sexual Debut and Satisfaction. To examine our hypotheses regarding sexual debut and biological sex differences, we conducted a series of 2 x 2 x 2 Analyses of Variance (ANOVAs) using sexual debut (early, late), biological sex (men, women), and relationship status (in a relationship, single) as between-subjects independent variables. We conducted separate analyses for different types of sexual debut (vaginal/anal and oral) to evaluate whether similar patterns emerged.

Relationship satisfaction. As illustrated in Figure 1, participants who waited longer to have intercourse with their partner (\( M = 3.87, SD = 0.86, 95\% CI [3.71, 4.04] \)) reported higher relationship satisfaction than those who had intercourse relatively early in the relationship (\( M = 3.56, SD = 0.78, 95\% CI [3.40, 3.72] \)), \( F(1, 185) = 6.58, p = .011, \eta^2 = .03 \). Not surprisingly, participants in a relationship (\( M = 4.23, SD = 0.84, 95\% CI [4.01, 4.35] \)) were more satisfied with their relationship than those who were single (\( M = 3.20, SD = 0.72, 95\% CI [3.02, 3.37] \)), \( F(1, 185) = 75.60, p < .001, \eta^2 = .29 \). Women and men did not differ in relationship satisfaction, \( F(1, 185) = 2.11, p = .15, \eta^2 = .01 \). No two-way or three-way interactions were significant, all \( F \)‘s < 1.43, all \( p \)‘s > .235. Consistent with our hypothesis, early
Sexual debut was associated with lower relationship satisfaction. We did not find evidence of biological sex differences in relationship satisfaction.

Consistent with our hypothesis, the pattern of results for oral sexual debut mirrored the results for vaginal/anal sexual debut. Participants in late oral sex relationships (M = 3.85, SD = 0.91) reported higher relationship satisfaction than those in early oral sex relationships (M = 3.58, SD = 0.79), F(1, 178) = 4.89, p = .034, ηp² = .03. Also, participants in a relationship (M = 4.21, SD = 0.85) were more satisfied with their relationship than those who were single (M = 3.22, SD = 0.83), F(1, 178) = 63.61, p < .001, ηp² = .26. Women and men reported comparable levels of relationship satisfaction, F(1, 178) = 2.43, p = .121, ηp² = .01, and no two-way or three-way interactions were significant, all Fs < .96, all p's > .329.

Sexual satisfaction. We used the same analytic approach to examine associations between biological sex, relationship status, and sexual debut with sexual satisfaction. The timing of vaginal/anal sexual debut was not related to sexual satisfaction, F(1, 185) = 2.41, p = .123, ηp² = .01. However, participants in a relationship (M = 82.26, SD = 17.25) reported significantly higher sexual satisfaction than participants reporting on a previous relationship (M = 75.16, SD = 16.11), F(1, 185) = 14.17, p < .001, ηp² = .07. Women and men did not differ in sexual satisfaction, F(1, 185) = 0.70, p = .403, ηp² = .004, and no two-way or three-way interactions were significant, all Fs < 1.01, all p's > .318. Of note, Levene’s test for equality of variances was marginal (2.04, p = .052). Contrary to our hypothesis, sexual debut was not associated with sexual satisfaction. As with relationship satisfaction, we did not find evidence of biological sex differences in sexual satisfaction.

Comparable to vaginal/anal sexual debut, timing of first oral sex in a relationship was not significantly associated with sexual satisfaction, F(1, 178) = 0.92, p = .340, ηp² = .005. Participants in a relationship (M = 81.93, SD = 17.46) reported significantly higher sexual satisfaction than participants reporting on a previous relationship (M = 72.66, SD = 16.92), F(1, 178) = 13.69, p < .001, ηp² = .07. Men and women reported comparable sexual satisfaction, F(1, 178) = 0.41, p = .524, ηp² = .002, and no two-way or three-way interactions were significant, all Fs < 1.80, all p's > .181.

Relationship Dissolution. We expected that later sexual debut, independent of relationship or sexual satisfaction would be associated with longer relationship duration. The following analyses were conducted with participants who were single at the time of testing and reported on a previous (failed) relationship (n = 78). We corrected a positively skewed distribution for relationship duration with a square-root transformation. Two participants did not report relationship duration. A linear, forced-entry multiple regression confirmed that relationship satisfaction and sexual satisfaction did not predict relationship length, adjusted R² < .01, F(2, 75) = 0.24, p = .701. Thus, we did not include sexual and relationship satisfaction in subsequent analyses to evaluate the relationship between sexual debut and relationship duration.

Vaginal/anal sexual debut. A 2 x 2 ANOVA using sexual debut (early, late) and biological sex (men, women) as between-subjects variables and relationship duration as the between-participants variable revealed that women (M = 4.36 months, SD = 1.93, 95% CI [3.81, 4.92]) reported longer relationships than men (M = 3.21 months, SD = 1.55, 95% CI [2.62, 3.80]), F(1, 74) = 8.54, p = .005, ηp² = .104. Consistent with our hypothesis, participants whose sexual debut within the relationship occurred after 2 months (M = 4.62 months, SD = 1.71, 95% CI [4.02, 5.23]) reported longer relationships than participants whose sexual debut had occurred earlier (M = 3.42 months, SD = 1.85, 95% CI [2.85, 4.00]), F(1, 74) = 8.05, p = .006, ηp² = .098. The two-way interaction was not significant, F(1, 74) = 0.36 and p = .549, ηp² = .005.

Oral sexual debut. The pattern of results for oral sexual debut mirrored that for vaginal/anal sexual debut. Women (M = 4.37 months, SD = 1.93, 95% CI [3.80, 4.94]) reported longer relationships than men (M = 3.24 months, SD = 1.57, 95% CI [2.63, 3.85]), F(1, 71) = 7.31, p = .009, ηp² = .093. Consistent with our hypothesis, participants whose sexual debut within the relationship occurred after 2 months (M = 4.72 months, SD = 1.76, 95% CI [4.04, 5.41]) reported longer relationships than participants whose sexual debut had occurred earlier (M = 3.49 months, SD = 1.80, 95% CI [2.96, 4.02]), F(1, 71) = 5.48, p = .022, ηp² = .072. The two-way interaction was not significant, F(1, 71) = 1.01, p = .318, ηp² = .014.

Discussion
The purpose of the present study was to compare associations between sexual debut within a
Sexual Debut and Romantic Satisfaction | Vancour and Fallon

romantic relationship, relationship satisfaction, sexual satisfaction, and relationship duration in undergraduate men and women involved in dating relationships. We found that young adults who delayed sexual intercourse (both vaginal/anal sex and oral sex) within their relationships reported higher relationship satisfaction than those who had sexual intercourse within 2 months. These findings are consistent with previous research (Busby et al., 2010; Peplau et al., 1977; Willoughby et al., 2014) and with the sexual restraint theory (Metts, 2004), which posits that early sexual debut compromises the development of emotional intimacy and, hence, relationship satisfaction.

The effect size of the association between sexual debut and relationship satisfaction was small in the current study (.03), which is consistent with previous research. Busby et al. (2010) surveyed only married couples and also reported a small effect for the association between sexual debut and relationship satisfaction (.02). Willoughby and colleagues (2014) also reported a very small effect size (.01) in young adults who were in dating relationships. Peplau et al. (1977), who also sampled young adults in dating relationships, did not report effect size or dispersion statistics. Nevertheless, it appears that the effect size was slightly stronger in Peplau et al. (1977) than in the current study, which is likely due to the increased sensitivity from interviewing couples within the same relationship. Taken together, these findings suggest that sexual debut exerts small effects on relationship satisfaction in dating couples as well as married and cohabiting couples. Although the design of the present study did not allow for conclusions regarding the causal direction of the association between sexual debut and relationship satisfaction, previous research using structural equation modeling has suggested that sexual debut influences relationship satisfaction (Busby et al., 2010).

Contrary to prediction, we observed no statistically significant association between sexual debut and sexual satisfaction across men and women. Busby and colleagues (2010), who found that early sexual debut was associated with lower sexual quality, quantified sexual quality using three questions, one of which involved the frequency of intercourse. By contrast, the ISS (Fischer & Corcoran, 1994) used in the present study is a 25-item measure that does not include frequency of intercourse. Thus, our unexpected null results regarding sexual satisfaction may be due to qualitatively different operationalizations of sexual satisfaction. Future research should examine this possibility.

Associations between sexual debut and relationship satisfaction were comparable across men and women in the present study, which is contrary to some previous research. Sassler et al. (2012) found that later sexual debut produced a 6% increase in women’s, but not men’s, relationship satisfaction. By contrast, Busby et al. (2010) and Willoughby et al. (2014) found that sex did not moderate the association between sexual debut and relationship satisfaction. This discrepancy may be due to differences in the definition of late sexual debut. Sassler et al. (2012) considered late sexual debut as 6 months after the start of the relationship, compared to 2 months in the present study and other research (Busby et al., 2010; Willoughby et al., 2014).

Peplau et al. (1977) found that women with later sexual debut reported less sexual satisfaction than women with earlier sexual debut. Women with later sexual debut also noted greater guilt about sexual intercourse than women with earlier debut. However, consistent with Busby et al. (2010), we found that sex did not moderate the association between sexual debut and sexual satisfaction. This difference may reflect a shift in women’s enjoyment of sex. Overall, views in the United States about sex have changed drastically since the 1970s. As “hook-up” culture has become more acceptable (Bogle, 2007), sexual satisfaction may be dependent on the couple’s sexual debut. In this hook-up culture, contemporary young adults are likely to have more experience with sex outside of their current relationship and less guilt about having sex within the relationship. Future research should examine whether young adults—and especially women’s—feelings of guilt over sexual intercourse within a romantic relationship have changed since the 1970s.

Previous research and the current study testing the sexual restraint model (e.g., Busby et al., 2010) assumes that sexual debut predicts subsequent relationship satisfaction. However, the relationship between sexual debut and relationship satisfaction could be bidirectional. To our knowledge, no research has longitudinally followed newly dating couples to track their satisfaction before and after sexual debut. Such an approach would prove challenging given that approximately 40% of dating couples have sexual intercourse within the first 2 months of dating (e.g., Willoughby et al., 2014). Further, this approach would preclude couples who have sexual intercourse casually before entering
into a romantic relationship.

In our study, the relationship between sexual debut within a relationship may be, in part, due to relationship duration. Indeed, the median duration of relationship for participants reporting early debut was under 1 year, whereas the median for those who prolonged sexual debut was 2 years. Previous research that controlled for relationship duration is mixed. Busby et al. (2010) found comparable associations between sexual debut and satisfaction after controlling for relationship duration. Willoughby et al. (2014), who tested “serious” dating couples, found that relationship duration did moderate the association between sexual debut and relationship satisfaction such that there was no effect of sexual debut within the first year of dating. Although we did not have the statistical power within our study to properly control for relationship duration, we ran exploratory 2 x 2 x 2 ANOVAs with relationship duration (2014), young adults in the current study with later sexual debut tended to have longer relationships and satisfaction regardless of relationship status in the first year of dating. Although we did not have the statistical power within our study to properly control for relationship duration, we ran exploratory 2 x 2 x 2 ANOVAs with relationship duration (=< 12 months, > 12 months), sexual debut, and relationship status as between-participants factors and found no interactions between sexual debut and relationship duration. Therefore, we cautiously conclude that the associations we observed between sexual debut and relationship satisfaction are not due to relationship duration. Nevertheless, future research should further examine this possibility.

The present study was the first to differentiate between oral and vaginal/anal sexual debut. We found that the association between early sexual debut and lower relationship satisfaction was comparable for vaginal/anal and oral intercourse. Thus, our findings demonstrated that any early sexual intercourse is associated with lower relationship satisfaction. Most college students do not consider oral sex to be sexual intercourse (Byers et al., 2009). As such, the present results could help sex educators and clinicians counsel young people who are considering becoming sexually active within their romantic relationships.

The association between sexual debut and relationship satisfaction may be small, but it is pervasive. Neither sex nor type of intercourse moderated the association. Further, early sexual debut was associated with poorer relationship satisfaction regardless of relationship status in the current study. Consistent with Willoughby et al. (2014), young adults in the current study with later sexual debut tended to have longer relationships than participants with earlier sexual debut. Further, participants in failed relationships reported significantly earlier vaginal/anal and oral sexual debut within a relationship than students in intact relationships. Of note, single participants in the current study were comparable to participants in romantic relationships across all demographic characteristics we gathered with the exception of their age during their first oral sexual experience. Although it is possible that the association between sexual debut and relationship duration we observed in single participants was due to substantial qualitative differences between these participants and students in a relationship at the time of testing, we have little evidence to suggest that is the case.

Unlike Le et al. (2010), we did not find a positive association between relationship satisfaction and relationship duration. Perhaps our sample did not have a sufficient range of longer term relationships for this association to emerge. However, single participants’ previous relationships ranged from 1 to 66 months, with nearly 20% of participants having been in relationships spanning more than 2.5 years. A more likely explanation involves the retrospective nature of participants’ report. Le et al. (2010) studied relationships longitudinally, so their participants in failed relationships rated relationship satisfaction while still in their relationship. Not surprisingly, in the current study, young adults in failed relationships reported significantly poorer relationship and sexual satisfaction than their counterparts in intact relationships. Our findings could reflect hindsight bias, when people feel as though they knew something was going to happen only after that event occurred (Roese & Vohs, 2012). Thus, if people already knew their relationship was destined to fail, they would report less satisfaction, regardless of the duration of their relationship.

There are some demographic limitations of the current study. Most studies that compare sexual debut and sexual and relationship satisfaction have been conducted with heterosexual adults. Our study is no exception. Although we included participants who did not identify as heterosexual within our sample, the number was too small to permit comparisons across sexual orientation. In addition, our sample was largely Caucasian and under 20 years of age. Relationship between sexual debut, sexual satisfaction, and relationship satisfaction may be different across racial and ethnic groups and in older young adults (e.g., 30-year-olds) who likely have more experience with sexual intercourse and in romantic relationships. Future research should examine whether the current findings extend to more diverse samples.
The present study produced four important findings. First, early sexual debut was associated with less relationship satisfaction, but not less sexual satisfaction in young adults in dating relationships. Second, associations between sexual debut and satisfaction were comparable regardless of the type of sexual intercourse (anal/vaginal/oral). Third, neither sex nor current relationship status moderated associations between sexual debut and relationship satisfaction. Fourth, earlier sexual debut was associated with shorter durations of failed relationships. Taken together, these findings support the sexual restraint model (Busby et al., 2010; Metts, 2004). Delaying sexual intercourse may provide young adults with more time to develop emotional connection with their partners, thereby increasing relationship satisfaction and motivation to remain in the relationship. Based on these findings, young adults who desire a more satisfying and longer romantic relationship may choose to delay their first sexual intercourse. The sex will still be just as satisfying.

References


Author Note. Jaclyn M. Vancour, Department of Psychology, University of New Haven. Marianne Fallon, Department of Psychology, Central Connecticut State University.

Portions of this manuscript have been presented at local undergraduate conferences and at the Eastern Psychological Association’s annual convention, and have appeared within Jaclyn M. Vancour’s Honors Thesis. Special thanks to all Psi Chi Journal reviewers for their support.

Correspondence concerning this article should be addressed to Jaclyn M. Vancour, e/o Dr. Marianne Fallon, Department of Psychological Science, Central Connecticut State University, 1613 Stanley Street, New Britain, CT 06050. E-mail: jvanc2@unh.newhaven.edu
The Effects of Parental Support and Self-Esteem on Internalizing Symptoms in Emerging Adulthood

Lewie E. Moore II
Madelynn D. Shell*
The University of Virginia’s College at Wise

ABSTRACT. This study examined social and individual predictors of internalizing symptoms in college students, and in particular explored the indirect effects of mother and father support through self-esteem. A total of 123 college students completed self-reported online surveys measuring mother and father support, self-esteem, and internalizing symptoms (depression and anxiety, withdrawal, and somatic symptoms). Students reported greater support from mothers compared to fathers, $t(114) = 5.84, p < .001$. On average, higher maternal (but not paternal) support ($\beta = -0.25, p = .006$) and self-esteem ($\beta = -0.61, p < .001$) were associated with lower internalizing symptoms. For women, self-esteem mediated the relationship between maternal support and internalizing symptoms, indicating that greater maternal support was associated with greater self-esteem, which in turn was associated with lower internalizing symptoms ($F = 41.98, p < .001$). However, this pathway was not significant for men. These results highlight the importance of exploring the influence of different sources of support separately, and suggest that improving self-esteem for both men and women, and improving maternal support for women may decrease risk for internalizing symptoms in emerging adulthood.

Internalizing disorders such as depression and anxiety are often a concern on college campuses, and even subclinical internalizing symptoms have the potential to significantly impair students’ social, emotional, and academic well-being (Eisenberg, Gollust, Golberstein, & Hefner, 2007; Kessler, Foster, Saunders, & Stang, 1995; Weitzman, 2004). Internalizing symptoms often begin around the start of puberty and increase across adolescence and emerging adulthood (Bhatia & Bhatia, 2007; Crawford, Cohen, Midlarsky, & Brook, 2001; Schraedley, Gotlib, & Hayward, 1999). Identifying factors that predict internalizing symptoms, particularly at developmental turning points such as the college transition, during which these symptoms may be more likely to change, is crucial to preventing them. Emerging adults who transition to college experience dramatic shifts in sources and types of social support, therefore it is important to explore how social relationships can impact the development of psychopathology during this time. Parental support, both early in childhood and during college, may protect students from internalizing symptoms by providing them with the emotional tools such as self-esteem to help them cope with challenges on their own (Chao, 2012; Colarossi & Eccles, 2000; Lee, Dickson, Conley, & Holmbeck, 2014; Rueger, Malecki, & Demaray, 2010). Furthermore, the support received and the impact of this support may differ for men and women (Rueger et al., 2010). Thus, this investigation will explore the effects of parental support, self-esteem, and gender on the development of internalizing symptoms in college students.

*Faculty mentor
Internalizing Symptoms Emerging Adulthood

Emerging Adulthood
Emerging adulthood is a developmental period, between ages 18 to 25, in which individuals rely less heavily on their parents compared to adolescence, but have not yet achieved adult milestones such as marriage or a stable career (Arnett, 2000, 2007). One of the biggest sources of change in emerging adult development is education, and college is often a time for identity exploration (Chiang & Hawley, 2013; Germejs, Luyckx, Notelaers, Goossens, & Verschueren, 2012). However, with this exploration may come stressors and challenges such as increased independence in financial, social, and educational decisions. When this independence is combined with less parental guidance, it could lead to the development of internalizing symptoms (Lee et al., 2014; Mattanah, Brooks, Brand, Quimby, & Ayers, 2010). Identifying predictors of such internalizing symptoms could help improve emerging adult well-being by easing the transition to college and adulthood. Although not all emerging adults enter college, over 65% of high school graduates go on to college, so college students are representative of a large portion of early emerging adults (Bureau of Labor Statistics and U.S. Department of Labor, 2016; U.S. Department of Education and National Center for Education Statistics, 2016). Furthermore, these processes of increased independence may be similar for emerging adults who do not attend college.

Internalizing Symptoms
Internalizing symptoms include depression, anxiety, withdrawal, and somatic symptoms that can, directly or indirectly, affect individuals’ well-being (Lee et al., 2014; Telzer & Fuligni, 2013). Internalizing symptoms may affect well-being even if the individual does not meet diagnostic criteria for specific anxiety or depressive disorders, thus it is important to understand internalizing symptoms along a continuum. Previous evidence has suggested that, during emerging adulthood, internalizing symptoms are often associated with negative outcomes such as alcohol abuse, poor academic performance, and poorer social adjustment (Eisenberg et al., 2007; Kessler et al., 1995; Weitzman, 2004). Thus, it is important to identify factors that may contribute to internalizing symptoms.

Starting in puberty and continuing into adulthood, many studies have found that girls and women experience higher levels of internalizing symptoms than boys and men (Bosacki, Dane, Marini, & Youth Lifestyle Choices-Community University Research Alliance, 2007; Crawford et al., 2001; Li, Albert, & Dwelle, 2014; Rueger et al., 2010). Although some have posited that this gender gap develops as a result of hormonal differences at puberty (Bhatia & Bhatia, 2007), there is also evidence that environmental changes may be more strongly associated with internalizing symptoms in women than men. For example, girls experienced higher levels of psychological distress in response to the middle school transition (Chung, Elias, & Schneider, 1998), and were more likely than boys to experience increases in depression (Hirsch & Rapkin, 1987). Likewise, women may be more likely to experience increases in internalizing symptoms during the transition to college and may be particularly affected by external factors such as social support (Li et al., 2014; Schraedley et al., 1999). Although internalizing symptoms may have a significant effect on the well-being of both men and women, gender may play an important role in the rates and predictors of such symptoms.

Parental Support
Parental support can come in many forms including emotional support expressed through positive affection and instrumental support in the form of financial or other assistance (Chao, 2012; Lee et al., 2014). Parental support may influence how students cope with stressors associated with college. Many previous studies have combined parental support with other sources of support such as peers and romantic partners, and in general, evidence has suggested that social support decreases internalizing symptoms (Lee et al., 2014; Schraedley et al., 1999; Zhao, Kong, & Wang, 2013; Zhou, Zhu, Zhang, & Cai, 2013). In addition, evidence has suggested that, in adolescence, decreased parental support may be a stronger predictor of internalizing symptoms than changes in other sources of support (Hughes & Gullone, 2008; Rueger et al., 2010). Because parental relationships are particularly likely to change in emerging adulthood as children become more independent, it is important to explore the independent effects of parental support during college.

In addition to separating parents from other sources of support, it may be beneficial to explore the differences in the amount and function of support from mothers versus fathers. Evidence has suggested that mothers often provide higher levels of support compared to fathers (Furman & Buhrmester, 1992; Phares, Renk, Duhig, Fields, & Sly, 2009; see Colarossi & Eccles, 2000 for an
exception) and that maternal versus paternal support may play a more important role in protecting children from internalizing symptoms (Anderson, Salk, & Hyde, 2015). In addition, there may be differences in the levels of parental support perceived by men versus women. Although some studies have found no such gender differences (Phares et al., 2009; Rueger et al., 2010), others have found that adolescent boys perceive more father support and girls perceive more mother support (Robinson, 1995) and report having both more positive and more negative interactions with family members (Telzer & Fuligni, 2013). Perhaps as a result of differing interaction quality, parental support may have a greater impact on women’s versus men’s internalizing symptoms (Schraedley et al., 1999). Thus, mother versus father support may uniquely impact internalizing symptoms, and these effects could differ for women versus men.

**Mediating Effects of Self-Esteem**

In addition to parental support directly impacting internalizing symptoms, it may also indirectly affect internalizing symptoms by helping emerging adults develop personal characteristics that decrease the likelihood that stressors will lead to internalizing symptoms. One such characteristic is self-esteem, or positive evaluations of one’s abilities and worth (Geng & Jiang, 2013). Such positive self-evaluations may help emerging adults continue to persist in the face of social or academic stressors, thus be associated with lower internalizing symptoms (Orth, Robbins, Meier, & Conger, 2016). Because self-esteem may develop as a result of parental support (Rueger et al., 2010), it may be the mechanism by which parental support reduces internalizing symptoms. Thus, self-esteem may be a mediating variable, which helps to explain how external events (parental support) take on psychological significance (internalizing symptoms, Baron & Kenny, 1986). Consistent with this, Lee and colleagues (2014) identified that low social support from family, friends, and significant others in the beginning of college was associated with lower self-esteem, which in turn led to increased depressive symptoms at the end of the first year of college. In addition, friendship quality and insecure adult romantic attachment have been associated with lower self-esteem, which is subsequently associated with higher internalizing symptoms (Bosacki et al., 2007; Roberts, Gotlib, & Kassel, 1996). Similarly, self-esteem may mediate the relationship between parental support and internalizing symptoms.

**The Present Study**

This study explored the role of maternal and paternal support on college students’ internalizing symptoms (i.e., depression and anxiety, somatic symptoms, and withdrawal) and the indirect influences of gender and self-esteem on this relationship. In contrast to previous research, this study explored the separate influences of maternal and paternal support on internalizing symptoms. Furthermore, this study added to previous research by exploring separate paths to internalizing symptoms for men versus women. Self-report surveys assessed maternal and paternal support, self-esteem, and internalizing symptoms among students at a small liberal arts college. It was hypothesized that maternal support would be more predictive of internalizing symptoms than paternal support. Furthermore, it was expected that students with greater maternal support would have higher self-esteem and fewer internalizing symptoms, and that self-esteem would mediate the relationship between maternal support and internalizing symptoms. However, these pathways may differ between men and women given that women are more likely to experience internalizing symptoms.

**Method**

**Participants**

An initial group of 123 college students (71 women, 58%; 52 men, 42%) enrolled in psychology courses at a small, public liberal arts college completed the survey. In exchange for their participation, the students received extra credit in one course. Participants ranged in age from 18 to 27 (Mage = 19.88, SD = 1.64) and class year (40.7% first-year students, 22.8% sophomores, 23.6% juniors, and 13% seniors). Participants were 76% European American, 17% African American, 2% Latino, 2% Asian American, and 2% other. Although this sample contained slightly more women and first-year students than the general campus population, the ethnicity distribution was approximately equivalent to the campus as a whole. Out of 123 students who initiated the survey, 112 completed the three questionnaires used in this study. The final participants did not differ significantly from the initial sample in terms of age, t(118) = -0.91, p = .36, college year, X²(3) = 4.20, p = .24, or race/ethnicity, X²(4) = 7.09, p = .13. However, women were significantly more likely than men to have completed all three surveys, X²(1) = 4.59, p = .03.

All study measures and procedures were reviewed and approved by the Institutional Review Board of...
the University of Virginia’s College at Wise.

**Measures**

**Parental support.** The Network of Relationships Inventory was used to measure parental support (Furman & Buhrmester, 1985). Individuals indicated their primary mother and father figure (biological/adoptive, step, or other), then indicated how true each statement was for both mother and father, from 1 (*little or none*) to 5 (*the most*). All participants with complete data identified a mother figure (95% biological or adopted, 2.5% stepmother, 2.5% other) and father figure (88% biological or adopted, 7% stepfather, 5% other). Maternal and paternal support composites were computed from an average of the following subscales (with three questions each): Instrumental Aid (e.g., “How much does this person help you when you need to get something done?”), Emotional Support (e.g., “How often do you turn to this person for support with personal problems?”), Reassurance of Worth (e.g., “How much does this person treat you like you’re admired and respected?”), Seeking and Providing Safe Haven (e.g., “How much do you seek out this person when you’re upset?”), and Seeking and Providing Secure Base (e.g., “How much does this person show support for your activities?”).

**Self-esteem.** The Self-Perception Profile for College Students (Neeman & Harter, 2012) was used to measure global self-esteem on a six-item scale. The self-esteem subscale was developed based on Rosenberg’s (1979) construct of self-esteem. However, it was designed to specifically assess college student self-esteem (Neeman & Harter, 2012). This measure has been used to assess self-esteem in adolescence and emerging adulthood, and previous publications have reported that the measure has acceptable reliability (Onwuegbuzie, 2000; Pittman & Richmond, 2008; Wichstrøm, 1995).

**Internalizing symptoms.** The Adult Self-Report (Achenbach & Rescorla, 2003) was used to measure internalizing symptoms. Participants indicated how well each item described them in the past six months on a scale of 0 (*not true*), 1 (*somewhat true*), 2 (*very true*). Participants answered 17 questions that assessed depression and anxiety (e.g., “There is very little that I enjoy” and “I worry a lot”), nine that assessed withdrawal (e.g., “I would rather be alone than with others”), and 12 that assessed somatic symptoms (e.g., “I feel dizzy or lightheaded”). The mean of the three subscales was used to obtain the total internalizing symptoms score.

**Procedure**

Participants completed a series of online surveys. Students who expressed interest following in-class announcements were e-mailed the link to an online survey and were asked to complete the questionnaires in one sitting in a quiet environment. The questionnaires took approximately 1 hour, and students were given 2 weeks to complete the survey.

**Results**

First, descriptive statistics and correlations were conducted for all variables (see Table 1). Next, relationships between maternal and paternal support were explored. Then, hierarchical linear regression analyses identified predictors of internalizing symptoms. Finally, regressions were used to
assess whether self-esteem mediated the relationship between parental support and internalizing symptoms.

Descriptive statistics and tests for internal consistency were computed for each scale (see Table 1). For the Network of Relationships Inventory, higher scores indicated higher levels of perceived support (Maternal: $M = 3.55$, $SD = 1.08$, Cronbach’s $\alpha = .94$; Paternal: $M = 2.91$, $SD = 1.13$, $\alpha = .94$). For the Self-Perception Profile for College Students, composites were calculated such that higher scores indicated higher self-esteem ($M = 3.13$, $SD = 0.65$, $\alpha = .83$). Finally, for the Adult Self-Report, high scores indicated higher levels of internalizing symptoms ($M = 0.40$, $SD = 0.31$, $\alpha = .74$).

A series of analyses were conducted to explore differences between maternal and paternal support. First, gender differences in the level of support were explored. There were no differences between men and women in perceived maternal support ($M_{men} = 3.57$, $M_{women} = 3.54$, $t(114) = -1.67$, $p = .87$), or paternal support ($M_{men} = 2.78$, $M_{women} = 3.10$, $t(114) = -1.47$, $p = .15$).

Next, the relationship between maternal versus paternal support was explored. Maternal support and paternal support were positively correlated ($r = 0.43$, $p < 0.001$) indicating that higher perceived maternal support was associated with higher paternal support. However, paired-sample $t$ tests indicated that perceptions of maternal support were significantly higher than paternal support ($M_{maternal} = 3.55$, $M_{paternal} = 2.91$, $t(114) = 5.84$, $p < .001$).

Next, hierarchical linear regressions were conducted to identify significant predictors of internalizing symptoms (see Table 2), and test whether self-esteem mediated the relation between parental support and internalizing symptoms. According to Baron and Kenny (1986), mediation requires that three criteria are met. First, the independent variable (parental support) must directly predict the mediator (self-esteem, Path A). Next, the mediator (self-esteem) must directly predict the dependent variable (internalizing symptoms, Path B). Finally, when Paths A and B are taken into account, a previously significant relation between the independent variable (parental support) and dependent variable (internalizing symptoms) is no longer significant (Path C).

First, consistent with expectations, women

### TABLE 2

<table>
<thead>
<tr>
<th>Variable</th>
<th>Combined Model</th>
<th>Women</th>
<th>Men</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>$-0.09$</td>
<td>$-0.29$</td>
<td>$-0.30^{**}$</td>
</tr>
<tr>
<td>Step 2</td>
<td>$-0.07$</td>
<td>$-0.25^{**}$</td>
<td>$-0.08$</td>
</tr>
<tr>
<td>Maternal Support</td>
<td>$-0.02$</td>
<td>$-0.02$</td>
<td>$-0.07$</td>
</tr>
<tr>
<td>Self-esteem</td>
<td>$-0.29$</td>
<td>$-0.61^{**}$</td>
<td>$-0.37$</td>
</tr>
<tr>
<td>Gender x Self-esteem</td>
<td>$0.10$</td>
<td>$1.02^{**}$</td>
<td>$-0.18$</td>
</tr>
</tbody>
</table>

**Note.** $*p < .05$, $^{**}p < .01$, $^{***}p < .001$
reported greater levels of internalizing symptoms than men ($\beta = -0.30$, $p < .001$), thus gender was retained in the subsequent models. At Step 2, the effects of maternal and paternal support were added to the model. Maternal, but not paternal, support was significantly associated with lower internalizing symptoms ($\beta = -0.25$, $p = .006$), so paternal support was dropped from subsequent analyses. In addition, to test for mediation, the relationship between maternal support and self-esteem was explored. As expected, college students with higher maternal support had higher self-esteem ($\beta = 0.30$, $p = .001$, Path A). Thus, in Step 3, self-esteem was added to the model predicting internalizing symptoms. In the resulting model, self-esteem ($\beta = -0.61$, $p < .001$) was associated with significantly lower internalizing symptoms (Path B), but maternal support no longer significantly predicted internalizing symptoms ($\beta = -0.07$, $p = .32$, Path C). This indicated that the link between maternal support and internalizing symptoms was partially accounted for by increased self-esteem. Finally, to assess whether gender influenced this indirect relationship, interactions between all variables were tested in Step 4. A significant Gender × Self-Esteem interaction emerged ($\beta = 1.02$, $p = .002$). This indicated a moderated mediation effect (final $R^2$ for full model = 0.53, $F = 30.52$, $p < .001$), suggesting that the mediational effects of self-esteem vary by gender (Baron & Kenny, 1986).

To probe the Gender × Self-Esteem interaction, analyses were re-run separately for men and women. For women, maternal support was associated with lower internalizing symptoms ($\beta = -0.28$, $p = .02$). However, once self-esteem was added to the model, the pathway between maternal support and internalizing symptoms was no longer significant ($\beta = -0.09$, $p = .28$), indicating that self-esteem mediated the relationship between maternal support and internalizing symptoms for women (see Figure 1, Sobel test = 2.31, $p = .02$). In contrast, for men, maternal support was not a significant predictor of internalizing symptoms ($\beta = -0.21$, $p = .17$) and remained nonsignificant when self-esteem was entered in the model, thus conditions for mediation were not met ($\beta = -0.02$, $p = .91$; see Figure 1, Sobel test = 1.32, $p = .19$). These models explained significant portions of the variability in internalizing symptoms, particularly for women (final $R^2$ for women = 0.56, $F = 41.98$, $p < .001$; for men = 0.27, $F = 7.57$, $p = .002$).

**Discussion**

This study provided important information about factors that contribute to internalizing symptoms during college. In particular, it highlighted the importance of separating different sources of social support and identified the essential role that mothers play in the transition to college. College students reported more support from mothers versus fathers, although mother and father support were correlated. Women, in general, had higher levels of internalizing symptoms and, consistent with hypotheses, mother (but not father) support was associated with lower internalizing symptoms in women. However, this relationship was mediated by self-esteem, indicating that maternal support may have protected women from internalizing symptoms through high self-esteem. In contrast, neither maternal nor paternal support predicted internalizing symptoms in men.
Parental Support During College
Overall, the evidence demonstrated some differences between mother and father support. Consistent with much of previous evidence, there were no differences between men and women in the amount of perceived parental support (Phares et al., 2009; Rueger et al., 2010). Nonetheless, college students consistently reported that mothers provided more emotional, instrumental, and relational support than fathers (consistent with Furman & Buhrmester, 1992; Phares et al., 2009). This may be the long-term outcome of early attachment relationships that often form with mothers first (Lamb, 2002; Lewis, Feiring, & Weinraub, 1981). Despite reporting more support from mothers, there was also a strong correlation between maternal and paternal support, indicating that, if students received support from one parent, they often also received it from the other (Lewis et al., 1981). This evidence suggests that students who come from a more supportive home will experience greater support from both mothers and fathers, thus parents may have related but independent roles in emerging adult development.

Predictors of Internalizing Symptoms During College
Several social and individual factors emerged as important predictors of internalizing symptoms. First, gender had a significant impact on the pathways to internalizing symptoms. Consistent with previous research, women reported more internalizing symptoms compared to men (Bosacki et al., 2007; Crawford et al., 2001; Li et al., 2014; Rueger et al., 2010). In emerging adulthood, as in adolescence, men may be more likely to experience externalizing, rather than internalizing symptoms as a result of stress (Hicks et al., 2007; Leadbeater, Kuperminc, Blatt, & Hertzog, 1999). Furthermore, both social factors (maternal support) and individual factors (self-esteem) contributed to internalizing symptoms for women, although these same pathways were not significant for men.

Parental support and internalizing symptoms.
Results demonstrated that maternal support, but not paternal support, was associated with increased internalizing symptoms in college students, consistent with previous work (Anderson et al., 2015). Mothers are often more involved in day-to-day activities throughout development (Lamb, 2002; Lewis et al., 1981). During emerging adulthood, this may take the form of more contact, reassurance, and encouragement, which may help mothers prevent, identify, or intervene should students develop internalizing symptoms. Alternatively, mothers and fathers may provide different types of support (e.g., mothers may provide more emotional aid whereas fathers provide more instrumental aid; Lewis et al., 1981) and the type of support that mothers provide may be particularly protective for internalizing symptoms. However, father support could be just as (or more) important for other aspects of well-being. Nonetheless, identifying the influence of mothers, but not fathers, highlights the importance of exploring these sources of support separately, rather than combined as they have been in previous research.

This evidence suggested that, in general, maternal support was positive. However, other evidence has found that it is possible to experience too much support. For example, helicopter parenting, when parents become overly involved in their children’s lives, is associated with increased depression and less life satisfaction in college students (Schiffrin et al., 2014). One explanation for this discrepancy could be that some components of parental support have positive impacts, while others have negative impacts (Larose & Boivin, 1998; Roberts et al., 1996). Thus, it is necessary to explore how different components of parental support contribute to internalizing symptoms.

An additional variable that could influence the relationship between parental support and internalizing symptoms is the context. It is expected that these findings generalize to emerging adults who are not in a college setting; however, this would need to be explicitly explored. Because emerging adults not enrolled in college do not have the institutional supports that are available on a college campus, parental support could be even more important for these emerging adults.

The mediating role of self-esteem. Consistent with previous research, high self-esteem was associated with fewer internalizing symptoms in both men and women, indicating that positive self-perceptions may serve to protect college students from psychopathology (Li et al., 2014; Orth et al., 2016; Rosenberg, Schooler, & Schoenbach, 1989). Self-esteem may play a particularly important role during college as students are facing increased academic and social stressors with less structured social support from family, teachers, and peers. Those who have high self-esteem may be more likely to have the confidence to persist in the face of failure, rather than experiencing feelings of helplessness and hopelessness that can lead to
internalizing symptoms (Nolen-Hoeksema, Girgs, & Seligman, 1986).

Furthermore, evidence suggested that higher self-esteem could explain the relationship between maternal support and lower internalizing symptoms for women. Thus, women with greater maternal support were more likely to have higher self-esteem, which was associated with lower levels of internalizing symptoms. In contrast, women who had lower self-esteem, possibly as a result of poor maternal support, may have been less able to cope with stressors and challenges associated with college life, and therefore experienced more internalizing symptoms. This gender difference is consistent with previous work identifying that stress and social support were particularly important in predicting depression in adolescent girls compared to boys (Schraedley et al., 1999), however it demonstrates that these patterns can continue into emerging adulthood. Women may experience greater benefits (protection from internalizing symptoms) from social support compared to men.

In contrast, for men, maternal support was not directly or indirectly associated with internalizing symptoms. Although no gender differences were found in the amount of parental support in this study, previous research has identified that men report fewer positive and negative interactions with their parents (Telzer & Fuligni, 2013), potentially resulting in a lower impact of parental interactions on men’s well-being. Instead, other sources of support, such as peers may become more important than parents during emerging adulthood for men (Rueger et al., 2010), and therefore may be more predictive of men's internalizing symptoms. Furthermore, because low self-esteem, but not maternal support, was associated with increased internalizing symptoms for men, it is possible that other sources of support may influence the development of self-esteem for men. In addition, other factors such as genetic predispositions or personality traits may predict men versus women's internalizing symptoms (Kendler, Gardner, & Prescott, 2006).

Contributions and Limitations

This study highlighted the importance of exploring both social and individual factors in predicting internalizing symptoms in college students. The different effects of maternal versus paternal support demonstrated the necessity of separating individual types of support, rather than combining them as in previous research (Chao, 2012; Lee et al., 2014; Schraedley et al., 1999; Zhao et al., 2013; Zhou et al., 2013), and highlighted the importance of maternal support in particular. In addition, this study demonstrated the importance of self-esteem in predicting internalizing symptoms for both men and women, suggesting that positive self-image may serve a buffering role against negative outcomes that could result from many stressors that emerging adults face during college. Finally, gender differences in this study highlight the importance of exploring individual trajectories of internalizing symptoms. Although these results were specific to emerging adults currently enrolled in college, the same processes may occur for emerging adults who do not attend college.

Despite these strengths, there are several limitations to these findings. First, all data was collected at one point in time, but in order to test for complete mediation, the temporal order of effect needs to be established. Thus, maternal support would need to be tested first, then self-esteem, and then internalizing symptoms. In addition, assessing these constructs longitudinally would allow for an exploration of whether these processes work in the same way immediately after the transition to college (when students may be relying more heavily on parent support) versus later in their college careers when they also have established peer support networks. A second limitation is that the study relied exclusively on self-reports, thus shared-method variance could have affected the findings such that students who experienced more internalizing symptoms might have perceived lower social support as a result of distorted perceptions. Adding parent-reports could further corroborate these findings. A third limitation is that data was collected online, and only from students in psychology courses on a small campus. Greater control over the response setting, samples from more diverse campuses, and the inclusion of emerging adults not attending college could have made the data more generalizable. Finally, although a clear pathway emerged for women, further exploration is needed to identify variables that may predict internalizing symptoms for men.

Despite limitations, these findings suggest two potentially important points of intervention to decrease internalizing symptoms in emerging adulthood. First, interventions that improve self-esteem may decrease internalizing symptoms for both men and women. Although it appears that women develop high self-esteem as a result of positive maternal support, an increase in self-esteem for any reason could decrease internalizing symptoms.
Programs at the college- or high school-level that promote positive self-evaluation may result in students being less likely to experience anxiety, depression, or withdrawal in the face of stressors. In addition, programs designed to increase maternal support (particularly for women) may help to increase self-esteem, and as a result decrease internalizing symptoms in emerging adulthood. Teaching mothers about providing age-appropriate support may help them develop positive relationships with their daughters, and as a result, students may be better able to cope with the challenges associated with college. Because internalizing symptoms can negatively influence academic, social, and emotional adjustment in emerging adulthood (Eisenberg et al., 2007; Kessler et al., 1995; Weitzman, 2004), it is necessary to continue to identify how to minimize the development of such symptoms. College administrators may use these findings to develop programs to help identify students at risk for internalizing symptoms and create intervention programs based on risk factors.

References

COPYRIGHT 2017 BY PSI CHI, THE INTERNATIONAL HONOR SOCIETY IN PSYCHOLOGY (VOL. 22, NO. 2/ ISSN 2325-7342)
Internalizing Symptoms Emerging Adulthood | Moore II and Shell


Author’s Note. Lewie Moore II and Madelynn D. Shell, University of Virginia’s College at Wise.

This research was funded through a Slemp Foundation Grant, which supports The University of Virginia’s College at Wise Summer Scholars Program. Thank you to the students who participated in the project. Special thanks to Psi Chi Journal reviewers for their support.

Correspondence should be addressed to Madelynn D. Shell, 1 College Avenue, Wise, VA 24293, (276) 376-3421. E-mail: mjs5ma@uvawise.edu
Learning through song is a commonly used strategy—after all, many children are taught lyrics with melody to learn the alphabet, days of the week, the 50 states, and more. Music-based learning can also be effective for more dense material and difficult facts. For example, there are songs to support the learning of statistics, the periodic table of elements, and foreign languages. The current study replicated and extended prior research on learning through music, with the goal of establishing more definitively the benefit of music as a mnemonic tool for college-aged younger adults.

Mnemonics are memory strategies that organize to-be-learned information in ways that enable encoding and facilitate later retrieval. Various types of verbal mnemonics (e.g., acronyms, acrostics) and visual-spatial mnemonics (e.g., keyword, method of loci, pegword) have been shown to improve memory for longer term retention of information (Bellezza, 1996). Music is promising as an effective mnemonic because it offers unique chunking and associative strategies based on rhythm, melody, and pacing. Relevant to pacing, temporal or time-based aspects of songs have been considered especially helpful when organizing information in memory (see Yalch, 1991, for a review). Additionally, evidence that people integrate the tune and the lyrics in the resulting mental representation is consistent with the idea that musical elements
support memory for lyrics (Morrongiello & Roes, 1990; Serafine, Crowder, & Repp, 1984).

Prior research on the use of music as a mnemonic has generally supported the current hypothesis that setting lyrics to music will improve recall. Several studies have focused on music and memory in children, finding an advantage for music as a mnemonic aid (Campabellino, De Carlo, O’Neil, & Vacek, 2002; Gingold & Abravanel, 1987). With regard to the population of interest for the current study, college-aged adults, Chazin and Neuschatz (1990) compared the effectiveness of learning scientific information through a familiar melody versus through lecture. Results showed that song-based information produced a higher recall over lecture-based information (see also commentary by Scruggs & Brigham, 1991). In another study that utilized familiar melodies, Rainey and Larsen (2002) found no effect of music on initial learning, but did find faster relearning in participants who heard the sung version of the information.

However, other research has failed to show a memory impact for musical presentation of information. Moore, Peterson, O’Shea, McIntosh, and Thaut (2008) studied music as a mnemonic in people with multiple sclerosis and healthy controls. The results did not show significant differences in recognition memory between groups or between music and spoken conditions. Moore et al. discussed that these null results could have been due to differences in the format of learning and testing conditions, and also insufficient time for encoding. Deason et al. (2012) examined recognition memory performance in patients with Alzheimer’s and healthy older adults. Both groups listened to lyrics accompanied by music and lyrics spoken without music, while having the lyrics in front of them. There was no significant difference in memory performance, including at a 1-week delay, between sung versus spoken lyrics in the healthy older adult group. Although on the surface these studies do not appear support our hypothesis about music benefitting memory, it is important to note the differences between their studies and the current experiment—namely, the focus on comparing preexisting participant groups, and the choice of stimuli and procedures.

Some research has even suggested superior memory for spoken words compared to song conditions. For example, Calvert and Billingsley (1998) compared children’s memory of their phone numbers presented in a song versus spoken condition. The results indicated that the children recited the spoken condition better than the song. Researchers suggested that songs offer children stimuli that were too challenging to their minimal attention and processing resources, a stark contrast to our study that examined adults with better ability to handle task load.

In addition to examining the effectiveness of music as a mnemonic device, the current study examined whether music “experts” (i.e., musicians) can use the musical structure of a song to better or more efficiently memorize lyrics compared to nonmusicians, perhaps due to enhanced chunking. Some research has suggested an advantage for musicians when learning from musical stimuli, whereas other studies have failed to show an advantage for those with musical training. Ginsborg (2007) had singers perform an a capella song from memory after they had intentionally learned and memorized the lyrics and melody, both independently and together. Musicians performed more accurately and articulately than participants who had less musical expertise, but this happened only when they memorized the lyrics and melody together. Ginsborg argued that lyrics and melody are recollected together, such that recalling one allows the recall of the other. These findings support the hypothesis that memorizing lyrics and melody together can be a successful strategy, especially for musicians.

Furthermore, Silverman (2007) studied the effect of paired pitch, rhythm, and speech on college students’ memory, using sequential digit recall as the dependent measure. Participants listened to four counterbalanced conditions, each consisting of nine randomized one-syllable digits paired with speech, pitch, rhythm, and the combination of pitch and rhythm. Participants recalled digits from the rhythm condition most precisely, and recalled digits from the speech and pitch only conditions the least precisely. Most important for the current study, the music majors scored significantly higher than the nonmusic majors.

Conversely, some research has shown no effect of music on memory recall for musicians. In Racette and Peretz’s (2007) study, participants learned an unfamiliar song in three conditions: sung-sung (the song being learned was sung, and the response was sung), sung-spoken (the response was spoken), and divided-spoken (the presented lyrics were accompanied by music and the response were spoken). Fewer words were recalled when singing than when speaking in both the musician and nonmusician groups. Taken together, these
conflicting results speak to the need for further research on the effect of music on memory recall, specifically in musicians.

As a means of understanding why musical experts may experience a stronger benefit of music as a mnemonic, and to better understand the broad theoretical frame regarding the role of musicians versus nonmusicians in the current study, it is helpful to consider classic and contemporary research on chess players, specifically focusing on the role of chunking as a hallmark of expertise. Chase and Simon (1973) found that master chess players more readily memorized a midgame configuration of chess pieces compared to chess novices; due to their extensive prior knowledge about chess, they perceived and encoded groups (or chunks) of pieces in meaningful sets. However, this memory advantage did not appear when memorizing chess pieces in random placements. Thus, experts may have a specific advantage in meaningfully chunking information in their domain of expertise. More recently, Bilalić, Langer, Erb, and Grodd (2010) showed that it took expert chess players half the time to count knights and bishops in a chess task compared with novices, particularly in the condition that was identical to a midgame scenario; yet contrary to Chase and Simon’s result, experts still had an advantage over novices for randomly placed positions in the chess task. Other researchers have used neurophysiological measures to look at chunking. Amidzic, Riehle, and Elbert (2006) found that the location of brain activity is different in experts, who show activation that is consistent with the retrieval of chunks, compared to nonexperts, who show activation in areas for memory formation. The results from these expertise and chunking studies support the general idea that musicians (as experts in the domain of music) may show a larger advantage for music-based encoding compared to nonmusicians.

The current study offered a partial replication and extension of two prior studies that used existing but unfamiliar songs as stimuli, and that found superior memory for musical encoding conditions in younger adults. First, Wallace (1994) presented verses of 80 to 85 words, with and without accompanying melody; results showed that, as long as the stimulus was repeated sufficiently for initial learning, there was an advantage for sung lyrics over read-aloud lyrics. Second, McElhinney and Annett (1996) compared the impact of hearing sung lyrics to read-aloud lyrics; each presentation was followed by a lyric recall test. Results showed that sung lyrics were recalled at a higher level in the second and third rounds of recall, and also that recall patterns for songs showed more evidence of chunking (i.e., recall in groups or phrases versus single words).

The purpose of the present study was to investigate music as a mnemonic device using elements of experimental design to improve internal validity (e.g., use of two counterbalanced songs/lyrics as stimuli, delayed recall tests). The two main research questions were as follows: (a) Is memory recall better for lyrics set to music compared to lyrics read without music? (b) Do musicians and nonmusicians benefit differently from music as a mnemonic device? Although prior research on this topic has been mixed, we hypothesized that the recall of musical lyrics would overall be higher than the recall of nonmusical lyrics at each time point. Further, we predicted an interaction between encoding condition and musicianship, such that musicians would have a specific advantage in recalling lyrics set to music.

In addition to analyzing the objective measure of recall, we also aimed to focus on subjective measures of task experience. We included questions about level of task load experienced during the recall task, with the hypothesis that learning lyrics with music may impact phenomenological perceptions of mental activity, time pressure, and/or task success; and that musicians may experience music-based learning differently than nonmusicians.

Method
Design
The experiment was a mixed-factor 2 (encoding condition: music, nonmusic) x 2 (expertise: musician, nonmusician) x 3 (Time: Recall 1, Recall 2, Recall 3) design. Encoding condition and time were manipulated within-subjects, and expertise was a between-subjects factor.

Participants
One hundred fifteen undergraduate students were recruited from various psychology courses and music groups at a private liberal arts college. Twenty students failed to return for the second session, and one student indicated familiarity with recall stimuli used; this resulted in the omission of 21 participants from the data set, leaving a final sample size of n = 94. Students were offered extra credit in some psychology courses, or entry in a gift card raffle in exchange for participation. The sample had a mean age of 19.79 (SD = 1.39) and
included 24.2% men, 72.6% women, and 2.1% other/preferred not to disclose.

Musicianship of participants was determined based on a survey about past musical experience, resulting in a reasonably equal comparison of \( n = 44 \) (47%) musicians and \( n = 50 \) (53%) nonmusicians. Using responses from this survey, researchers classified each participant as a musician or as a nonmusician. Two researchers separately reviewed each survey and decided whether each participant would fit the description of a musician in the context of the current study, based on the extent of their musical training. A third researcher mediated any discrepancies between the judgments of the two researchers.

Materials
Two sets of lyrics were chosen for the study from pre-existing songs. The musical stimuli, presented to participants in the music condition, consisted of portions from two songs that were delivered in a moderate musical pacing (mean BPM = 110) and were both sung by men. Portions of the songs were chosen to exclude any repeated choruses or lines, so there was very little lyrical repetition. In addition, both portions portray a narrative storyline, in order to open the potential of examining not only word recall, but a more general idea recall as well. The songs were “Taxi” by Harry Chapin and “Cold Missouri Waters” by Cry Cry Cry. The selection of lyrics from “Taxi (Lyric A)” included 147 words and 21 lines; the music clip was 1 minute 54 seconds long, and the nonmusic clip was 1 minute 3 seconds long. The selection of lyrics from “Cold Missouri Waters” (Lyric B) included 148 words and 17 lines; the music clip was 1 minute 53 seconds long, and the nonmusic clip was 1 minute 8 seconds long. In the nonmusic condition, participants were presented with a prerecorded reading of the lyrics. Each reading was done in a male voice, and each followed the pace of the original song. Participants were asked about their prior familiarity with the songs at the conclusion of each session; as noted above, the one participant who indicated familiarity was not included in the data set.

In addition to a sheet containing the printed lyrics, each participant was given a testing packet, which contained a blank sheet of paper for each free recall test, and an inventory to measure task load. The inventory contained six modified questions from the NASA-Task Load Index (Hart & Staveland, 1988), which were answered on a 7-point Likert-type scale. Questions were chosen and modified from the original scale to pertain to the task of lyric encoding and recall (see Table 1 for the modified questions). This inventory allows for items to be analyzed individually, each as a separate indicator of task load.

Testing packets for Session 2 contained a survey with demographic and musical background questions. In addition to asking participants to indicate whether they self-identify as musicians, the musical experience survey inquired about the participants’ level, type, and length (number of years) of musical experience. Finally, to collect further data regarding musicianship, the survey asked participants two questions on a 5-point Likert-type scale ranging from 1 (definitely no) to 5 (definitely yes): “Can you sight-read an easy piece of music?” and “If a melody was sung to you, could you sing it back accurately?”

Procedure
Following institutional review board approval (2014124R), testing sessions were conducted in groups of 1 to 10 participants. Counterbalancing was achieved by using two different lyric sets and two different encoding conditions (music and nonmusic), forming four possible conditions: musical presentation of Lyric A (MA), nonmusical (spoken) presentation of Lyric A (NMA), musical presentation of Lyric B (MB), and nonmusical (spoken) of Lyric B (NMB). Based on the assigned condition of Session 1, participants received the alternate pairings of each of the variables in Session 2. Thus, the four possible counterbalancing conditions are as follows (given in the format of Session 1-Session 2): MA-NMB (\( n = 22 \)), MB-NMA (\( n = 23 \)), NMA-MB (\( n = 23 \)), and NMB-MA (\( n = 26 \)). Groups of participants were initially assigned to a counterbalancing condition with the goal of achieving an equal distribution of participants in each condition. Results showed that there were no differences in memory recall performance across counterbalancing orders, so we will not discuss this variable further.

Upon arrival to the testing room, participants read and signed the informed consent form provided, then were each given a testing packet and a sheet of paper containing the lyrics of the audio clip to be presented during that session. They were told that they would listen to an audio clip three separate times, and that they would be asked to recall as much as they could after each presentation. Participants were asked to silently read along with the lyrics as they heard them. Upon the conclusion of the audio clip, they put the lyric
sheet away, turned to a blank page in their testing packets, and wrote down as many of the lyrics as they could remember including words, phrases, or general ideas, in a 3-minute period. The audio clip was then repeated, and they again completed the free recall test for 3 minutes. After the audio clip was played a third and final time, participants were given a 5-minute distractor task (a paper packet with several different puzzles) before completing the third free recall test. Participants then turned to the final page in their testing packet to complete the task load questionnaire (see Table 1 for items).

The second session, exactly 1 week later, was identical to the first, except for the encoding condition and lyric set (see counterbalancing description above), and the fourth free recall test at the end of the session. The fourth recall test was given to test longer term memory for the lyrics presented in the first session. Thus, approximately half of the participants performed a 1-week-delayed recall task from material in the music condition, and the other half recalled from the nonmusic condition. Unlike the other recall variables, which were manipulated within-subjects, the measure of delayed recall of lyrics from Session 1 was a between-subjects variable. Finally, at the conclusion of Session 2, participants were asked to complete a questionnaire to indicate demographic information and the nature of their musical experience (see Figure 1 for an overview of the study design). In summary, participants completed three recalls of one set of lyrics during one session; then, one week later, they completed three recalls of another set of lyrics, along with one recall for the lyrics from the previous week.

Free recall tests were scored for exact (verbatim) word recall and more leniently for line recall. Line recall was implemented to measure memory for general ideas, and used a point system in which one point was awarded for every line of the lyrics for which the scorer judged that the participant accurately expresses the idea or gist, regardless of exact word recall. Final proportion scores for both words and lines were calculated by dividing the units (words or lines) recalled by the maximum number of units in the given set of lyrics. Each response was scored separately by two different researchers, and any discrepancies were mediated by a third researcher, in order to ensure consistency.

Scoring for evidence of chunking was also done by a point system, based on clusters, which are operationally defined as a group of three to five words in a participant’s response that were recalled...
in the same order in which they were presented in the original stimulus (Meijs, Hurks, Rozendaal, & Jolles, 2013). For example, with the original lyrics reading, “she just looked out the window…” a participant would receive one point for “looked out the window.” Given the similar number of words in each set of lyrics, we did not compute proportion correct scores for Lyric A and Lyric B; instead, we report absolute number of clusters below.

Results

Musicianship Ratings

Pearson correlation analyses among pairs of items on the musicianship survey (see details above in Method section) showed strong positive correlations ($p$s < .001) for all variable pairs. Given that researcher-determined and self-reported musicianship ratings were strongly correlated, $r(92) = .730$, $p$s < .001, the researcher-determined ratings are used to classify musicianship as a dichotomous variable (i.e., musician, nonmusician) in the remainder of the analyses. Interrater reliability was determined using Cohen’s $\kappa$, which showed an almost perfect agreement between researchers’ judgments of musicianship, $\kappa = .914$, $p < .001$.

Lyric Recall

The alpha level was set at $p = .05$. We report $\eta_p^2$ as a measure of effect size: $\eta_p^2 = .01$ represents a small effect; $\eta_p^2 = .06$, a medium effect; and $\eta_p^2 = .14$, a large effect (Cohen, 1988). An outlier analysis of each recall variable was conducted based on converting raw accuracy scores to $z$ scores, then evaluating against the commonly used criterion of $z = 3.29$. Scores that exceed this criterion would be considered outliers, falling in the top or bottom 0.1% of the distribution. These analyses revealed no outlying scores in any of the variables.

Proportion of words recalled. A $2 \times 3 \times 2$ mixed-factor Analysis of Variance (ANOVA) was conducted to analyze the effect of encoding condition, time, and musicianship on proportion of words recalled (see Table 2 and Figure 2). Results showed a significant increase in word recall across each of three time periods, $F(2, 188) = 438.65$, $p < .001$, $\eta_p^2 = .828$, with all contrasts being significantly different, $p$s < .001 (see Table 2 for all descriptive statistics). Word recall was also significantly different between encoding conditions, $F(1, 92) = 15.86$, $p < .001$, $\eta_p^2 = .148$, with higher scores in the music condition ($M = 36$, $SD = 0.16$) than in the nonmusic condition ($M = 0.32$, $SD = 0.15$). In regard to musicianship, word recall for musicians

| TABLE 2 |

| Means and Standard Deviations for Recall in Music and Nonmusic Encoding Conditions for Musicians and Nonmusicians |
|-----------------|-----------------|-----------------|
|                 | Musician         | Nonmusician     | Overall |
| Word R1         | 0.25 (0.12)      | 0.23 (0.12)     | 0.24 (0.12) |
| Word R2         | 0.42 (0.15)      | 0.38 (0.15)     | 0.40 (0.15) |
| Word R3         | 0.48 (0.15)      | 0.42 (0.16)     | 0.45 (0.16) |
| Overall         | 0.38 (0.14)      | 0.34 (0.14)     | 0.36 (0.17) |
| Delayed Word    | 0.38 (0.18)      | 0.24 (0.14)     | 0.31 (0.17) |
| Line R1         | 0.44 (0.16)      | 0.44 (0.17)     | 0.44 (0.16) |
| Line R2         | 0.62 (0.13)      | 0.61 (0.16)     | 0.62 (0.14) |
| Line R3         | 0.68 (0.16)      | 0.65 (0.16)     | 0.67 (0.16) |
| Overall         | 0.58 (0.15)      | 0.57 (0.16)     | 0.57 (0.18) |
| Delayed Line    | 0.59 (0.20)      | 0.52 (0.18)     | 0.55 (0.19) |

Note: Recall 1 (R1) and Recall 2 (R2) each occurred after an opportunity to encode the lyrics while listening and looking at them. Recall 3 (R3) occurred after a filled delay. Delayed recall occurred one week after the initial session. Word recall represents verbatim memory for lyrics, and line recall represents memory for general ideas in lines of the song.
(M = 0.36, SD = 0.17) was not significantly higher than that for nonmusicians (M = 0.32, SD = 0.16), F(1, 92) = 2.79, p = .099, η² = .030. No interactions were found between any of the variables.

Next, to examine the 1-week-delayed recall test for lyrics presented in the first session, a 2 (encoding condition) x 2 (musicianship) ANOVA was conducted on word recall. A significant difference was found in relation to encoding condition, F(1, 92) = 4.96, p = .029, η² = .058, with higher scores in the music condition (M = 0.30, SD = 0.18) than in the nonmusic condition (M = 0.23, SD = 0.16). A significant difference was also found in relation to musicianship, F(1, 92) = 5.82, p = .018, η² = .067, with musicians having higher scores (M = .31, SD = .17) than nonmusicians (M = .23, SD = .16). There was no interaction, F(1, 92) = 2.65, p = .107, η² = .032 (see Figure 3).

**Proportion of lines recalled.** Similar results were found in regard to line recall (see Table 2 and Figure 2). A 2 x 3 x 2 ANOVA showed a significant increase in line recall scores in relation to time of recall, F(2, 188) = 227.23, p < .001, η² = .710, with all contrasts being significantly different, ps < .001. A significant difference in line recall was also found in relation to encoding condition, F(1, 92) = 4.09, p = .046, η² = .043, with higher scores in the music condition (M = 0.58, SD = 0.14) than in the nonmusic condition (M = 0.55, SD = 0.18). Line recall for musicians (M = 0.43, SD = 0.16) was not significantly higher than that for nonmusicians (M = 0.41, SD = 0.16), F(1, 92) = 1.25, p = .267, η² = .014, and no interactions were found between any of the variables.

Turning to results for 1-week-delayed line recall, a significant difference was found in relation to musicianship, F(1, 92) = 4.25, p = .042, η² = .050, with musicians showing significantly higher scores (M = 0.55, SD = 0.19) than nonmusicians (M = 0.46, SD = 0.22). No significant differences were found in relation to encoding condition, F(1, 92) = 0.23, p = .631, η² = .003; further, there was no interaction.

Cluster scores. Cluster scores produced similar patterns to that of line and word recall. Scores significantly increased with each recall, F(2, 188) = 217.92, p < .001, η² = .710, with all contrasts being significantly different, ps < .001. In addition, there was a significant difference in relation to encoding condition, F(1, 92) = 16.32, p < .001, η² = .155, with higher scores in the music condition (M = 11.80, SD = 7.76) than in the nonmusic condition (M = 9.71, SD = 6.35). For the 1-week-delayed recall, scores were also significantly higher in the music (M = 9.03, SD = 7.34) than in the nonmusic condition (M = 5.42, SD = 4.80), F(1, 92) = 7.40, p = .008, η² = .086. No significant differences were found in relation to musicianship. Not surprisingly, cluster scores were positively correlated with word and line recall scores in the corresponding encoding conditions, ps < .001.

**Task Load**

A series of 2 (encoding condition) x 2 (musicianship) mixed-factor ANOVAs was used for the task load questions (see Table 1). For the question, “How much mental activity did you feel due to pacing of the tasks,” there were no main effects (ps > .05), but there was a significant interaction, F(1, 92) = 4.22, p = .043, η² = .045. This interaction was driven by musicians reporting higher mental activity required for the task in the nonmusic condition, whereas nonmusicians reported higher mental activity in the music condition.

Analyses of the question, “How much time pressure did you feel due to pacing of the tasks,” showed a significant main effect of encoding condition, F(1, 92) = 4.09, p = .046, η² = .043; participants reported higher time pressure in the music condition (M = 4.69, SD = 1.66) than in the nonmusic condition (M = 4.38, SD = 1.47). No other effects were significant (ps > .05).

For the question, “How successful do you think you were in accomplishing the goals of the tasks set by the experimenter,” there was a main effect of encoding condition, F(1, 92) = 10.34, p = .002, η² = .102. Participants reported significantly higher perceived success in the music condition (M = 4.19, SD = 1.21) compared to the nonmusic condition (M = 3.81, SD = 1.11). We found no significant findings in relation to the remainder of the task load questions (ps > .05).

**Discussion**

The current study examined the effect of music on recall. Participants listened to lyrics set to music (or read aloud without music) three times, each followed by a free recall test, with a 5-minute delay period preceding the third test. One week later, participants repeated the procedure in the other encoding condition (music, nonmusic). They then completed a free recall test on the lyrics learned in the testing session 1 week prior. Recall was measured by percentage of words recalled, percentage of lines recalled, and number of clusters in each response. After both testing sessions, participants completed a task load inventory to measure the
subjective mental demands of the recall tasks.

As predicted, music-based encoding produced overall higher recall than nonmusic encoding. Participants showed higher word recall, line recall, and cluster scores across all three same-session recall tests when learning lyrics as part of the original song; this pattern was also evident on the 1-week-delayed recall test, for words and clusters. Music-based encoding, therefore, showed an advantage in long-term memory. These results not only provide support for the original hypothesis, but they are consistent with the previous literature regarding the effectiveness of music as a mnemonic device to improve recall of information (Campabell et al., 2002; Chazin & Neuschatz, 1990; Gingold & Abravanel, 1987). The finding that music encoding had an advantage in recall of specific verbatim information, ideas or gist, and chunks of information, has implications for educational settings: Learning information set to music may be useful in situations ranging from detailed memorization to the learning of general concepts and ideas.

The hypothesis stated that musical expertise may play a role in learning through song, based on findings of a music-based memory advantage for musicians (Ginsborg, 2007; Silverman, 2007) and extrapolated from the idea that experts have a chunking advantage in their domain of expertise (Chase and Simon, 1973). Contrary to this hypothesis, in the current study, there was no unique recall advantage for musicians in the music condition. Interestingly, however, musicians had significantly higher word and line recall scores on the 1-week-delayed recall test, regardless of encoding condition. Beyond the scope of this study, it is possible that the superior scores of the musicians reflected an overall advantage in general mental abilities as a result of musical training, and/or due to pre-existing characteristics (e.g., working memory capacity, verbal recall; Brandler & Ramm-sayer, 2003).

Results from the task load inventory provided a lens into the subjective experience of learning with and without music. The modified NASA-Task Load Index (Hart & Staveland, 1988) was administered at the end of both testing sessions to examine whether music-based encoding would change the perception of mental activity, effort, time, or task success. An interesting pattern was found in regard to the question, “How much mental activity was required for the tasks?” Whereas musicians reported higher mental activity required for the task in the nonmusic condition, nonmusicians reported higher mental activity required for the task in the music condition. Because this trend does not parallel the objective recall scores across conditions and participant groups, it is perhaps a reflection of an increased comfort level for musicians when experiencing a memory task in the context of their domain of expertise.

Additionally, participants (regardless of musical expertise) reported feeling higher time pressure from the task in the music condition, compared to the nonmusic condition. It can be speculated that the explicitly rhythmic and paced nature of songs, with accompanying melody and singing, might have caused participants to feel more rushed during the encoding task, in comparison to hearing the lyrics spoken aloud without music. It is also possible that these ratings were affected by the level of recall on the subsequent memory tests; given that the task load questions were administered at the very end of the session, the fact that participants were able to recall more words in the music condition might have contributed to their retroactive perceptions of a higher time constraint on encoding.

Finally, participants reported significantly higher perceived success in the music condition compared to the nonmusic condition. This is in accordance with the finding that participants had higher recall scores in the music condition. As such, participants seemed to be metacognitively aware that they were learning more of the lyrics when presented as part of a song.

The results overall suggest that music can aid in easier encoding and subsequent recall of information and that musicians may have an overall advantage in delayed recall of both sung and read lyrics. Results also suggest that music-based learning may cause participants to experience higher time pressure and also higher task success, and that musicians may experience a lower mental activity load when learning with music, as compared to nonmusicians.

Limitations and Future Directions

Several limitations should be taken into account when considering the results of the current study. To begin, even with equivalent pacing of songs and spoken lyrics, musical structure does add multiple layers of information to the encoding task, and the current design did not allow an examination of how those elements impact memory. It would be interesting to include as an additional comparison condition a set of lyrics that are rhythmically paced but not musical with regard to pitch.
In addition, seven participants’ data had to be omitted from the 1-week-delay recall, due to the researchers’ knowledge that the participants had sought out and practiced the musical stimuli from the first session. Some of the borderline-significant results might have passed the critical threshold for significance if these data could have been included. In general, there is a limitation that the lyrics were pre-existing and publicly available, and therefore no control existed regarding whether other participants had listened to the musical stimuli in between the first and second session. It is important to note that this limitation would only be relevant to data from the 1-week-delayed recall test. This concern would be completely alleviated by using original lyrics and songs, though this was addressed in the current study by omitting anyone who indicated prior knowledge of the song(s).

The results highlight several related areas that could be extended with further research. To begin, it would be beneficial to enhance understanding of the relationship between musical experience—and expertise in general—and memory tasks when performed in or outside the domain of expertise. This could also be expanded to investigate whether experts in various areas tend to spontaneously use their domains of expertise as a basis for chunking. Experts in various areas tend to spontaneously use their domains of expertise as a basis for chunking (mnemonic) strategies when learning new material. Although in the current study we did not find higher clustering scores in the music condition for musicians compared to nonmusicians, more research is needed to investigate this topic.

The motivation behind this study was to determine whether music provided effective memory support for encoding new information, using a research design that improved on prior research with regard to internal validity (e.g., a controlled laboratory setting, counterbalancing conditions, a mainly within-subjects design with repeated recall tests). As a result, this study provides stronger evidence for the advantage of music as a support for encoding new information.

A next step is to examine implications for education. Music is already being used to teach lower level educational content and is also being applied creatively to higher level content. For example, VanVoorhis (2002) demonstrated better learning and student enjoyment in a psychological statistics course when students learned statistical information through “stat jingles.” Future research should investigate the advantages and boundary conditions of musical learning. With specific consideration to the fact that much of the material covered in educational settings might not fit neatly into a narrative presentation, the current study could be replicated using non-narrative material as stimuli in order to explore differences in recall or experience that may accompany recall for a set of declarations (e.g., facts, formulas) rather than a story. Recent evidence-based initiatives directed to educators have provided suggestions about the types of low-cost, high-impact learning strategies that should be implemented in the classroom (e.g., Roediger & Pyc, 2012). Educators should not minimize the importance of also choosing strategies that may feel more successful—and even enjoyable—to the learners. Learning through music may be in this category of strategies that both enhance objective measures of memory, and also enhance motivation to learn.

References
Lyrical Memory | Lummis, McCabe, Sickles, Byler, Hochberg, Eckart, and Kahler


Author Note. Sarah N. Lummis, Jennifer A. McCabe, Abigail L. Sickles, Rebecca A. Byler, Sarah A. Hochberg, Sarah E. Eckart, and Corinne E. Kahler, Goucher College.

Thanks to Bianca Z. Stern, Alyssa J. Hauptman, and Talya Stern, for assistance with editing and data scoring and entry, and Brandon Meyers-Orr for assistance with designing the study. Special thanks to all Psi Chi Journal reviewers for their support.

Correspondence concerning this article should be addressed to Jennifer A. McCabe, Center for Psychology, Goucher College, 1021 Dulaney Valley Road, Baltimore, MD 21204, USA. E-mail: Jennifer.McCabe@goucher.edu
OUR JOURNAL
Advertising in Psi Chi Journal allows you to connect with established psychology researchers and mentors, as well as undergraduates and graduate students striving to build a career in one of the many areas of research. People regularly visit our journal online to
- view current and past issues,
- submit their research for publication,
- learn about reviewing for Psi Chi Journal,
- share invited editorials as teaching tools in the classroom.

All issues and advertisements are permanently free online to both members and nonmembers alike. During the 2015–16 fiscal year, psichi.org received 1.67 million page views, ensuring that high-achieving students and professionals will see your content for years to come.

To further enhance the visibility of our journal, latest issues and calls for submissions are regularly featured in Psi Chi Digest e-mails (180,000+ subscribers) and on our three social media platforms.
- Facebook (22,500+ followers)
- Twitter (4,700+ followers)
- LinkedIn (10,200+ followers)

Articles are also indexed in PsycINFO, EBSCO, and Crossref databases—essential tools that researchers use to search for millions of psychology-related articles. This makes Psi Chi Journal a key place to communicate your message with our Professional Organization’s 700,000+ lifetime members and far beyond.

AD SPECIFICATIONS
Digital format: PDF, EPS, and TIFF
Resolution: 300 dpi | B&W line art—1,200 dpi Black & white ads (no RGB or 4-color process)
PDF settings: Press quality, embed all fonts

DEADLINE/BILLING
Payment due upon receipt of invoice.

CONTACT
Submit contract by e-mail to
Susan Iles
Advertising Sales
Psi Chi Central Office
E-mail: psichieye@psichi.org
Phone: 423-771-9964


All advertisements must be scholarly and professional in nature, and Psi Chi reserves the right to reject (or cancel) any ads that are not in the best interest of the Organization or consistent with the Society’s mission.

ADV-J8-06 (05-2017)
Publish Your Research in *Psi Chi Journal*

Undergraduate, graduate, and faculty submissions are welcome year round. Only the first author is required to be a Psi Chi member. All submissions are free. Reasons to submit include

- a unique, doctoral-level, peer-review process
- indexing in PsycINFO, EBSCO, and Crossref databases
- free access of all articles at psichi.org
- our efficient online submissions portal

View Submission Guidelines and submit your research at [www.psichi.org/?page=JN_Submissions](http://www.psichi.org/?page=JN_Submissions)

---

**Become a Journal Reviewer**

Doctoral-level faculty in psychology and related fields who are passionate about educating others on conducting and reporting quality empirical research are invited to become reviewers for *Psi Chi Journal*. Our editorial team is uniquely dedicated to mentorship and promoting professional development of our authors—Please join us!

To become a reviewer, visit [www.psichi.org/page/JN_BecomeAReviewer](http://www.psichi.org/page/JN_BecomeAReviewer)

---

**Resources for Student Research**

Looking for solid examples of student manuscripts and educational editorials about conducting psychological research? Download as many free articles to share in your classrooms as you would like.

Search past issues, or articles by subject area or author at [www.psichi.org/?page=journal_main](http://www.psichi.org/?page=journal_main)

---

**Add Our Journal to Your Library**

Ask your librarian to store *Psi Chi Journal* issues in a database at your local institution. Librarians may also e-mail to request notifications when new issues are released.

Contact PsiChiJournal@psichi.org for more information.

---

Register an account:

[http://pcj.msubmit.net/cgi-bin/main.plex](http://pcj.msubmit.net/cgi-bin/main.plex)