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 2164-8204) 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.

Postmaster: Send address changes to the address above.

Print subscriptions are available on a calendar-year basis only (Spring–Winter). U.S. rates are as follows (four issues): Individual $20; Institution $40.

For international rates or other information, contact Psi Chi Central Office. Printed in the USA. Periodicals postage paid at Chattanooga, TN, and additional mailing offices.

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. Psi Chi does not accept paid advertising for Psi Chi Journal of Psychological Research.
Egocentric Inferences About the Values of Terrorist Groups
Nina Combs and Brian Collisson
Marian University

Young Adults’ Expectations About the Values of Religious and Spiritual People
Matthew Kobza, Montclair State University
Nicholas P. Salter, Ramapo College of New Jersey

Cross-Cultural Globalization of Advertisements
Julian B. Allen, Kimberly K. Lee, and Elena Escalera
Saint Mary’s College of California

Relationship Between Physical Activity and Cognition in Adults With and Without Alzheimer’s Disease
Briana N. Sprague, Amber S. Watts, and Jeffrey M. Burns
University of Kansas

Predictors of Psychological Outcomes in Nonheterosexual Individuals
Emily K. Sanders and Holly McCartney Chalk
McDaniel College

The Relationship Between Corumination and Health in Young Women: Evidence from Facebook Communications
CaSandra Swearingen, Jennifer Byrd-Craven, and Shelia M. Kennison
Oklahoma State University

The Effects of Empathy on Disparagement Humor
Peter Bui, Maria Kalpidou, Lauren DeVito, and Todd Greene
Assumption College

Steven V. Rouse
Pepperdine University

Journal Submissions Overview
Egocentric Inferences About the Values of Terrorist Groups
Nina Combs and Brian Collisson*
Marian University

ABSTRACT. In the wake of a terrorist attack, people question terrorists’ motives and values. In the current research, we hypothesized that people may egocentrically infer that terrorists oppose values that are important and relevant to themselves more than values that are personally unimportant or are relevant to someone else. To test these hypotheses, participants inferred the values of a hypothetical terrorist group. They inferred the terrorist group’s position regarding values that were relevant/irrelevant and important/unimportant to themselves. As expected, people inferred that the terrorist group was more opposed to their own values than the values of others, $F(1, 65) = 5.66, p = .02, \eta^2_p = .08$. People also inferred that the terrorist group was more opposed to important values than unimportant values, $F(1, 65) = 7.67, p = .009, \eta^2_p = .10$. These findings suggested that egocentrism may guide people’s inferences about terrorists when little information is available. Implications regarding public perceptions of real terrorist groups are discussed.

In the aftermath of the terrorist attacks on September 11, 2001, Americans experienced many different emotions. They felt fear, surprise, pain, and confusion (Silver, 2011). It is understandable that people were confused because, in the early stages after the attack, little information was available regarding what motivated a terrorist group to attack the United States. Given the little information available initially, U.S. citizens and political commentators often inferred whether the terrorist groups’ motives were economic, political, or morally based (Silver & Matthew, 2008).

Within psychology, there is a rich literature regarding how people explain, predict, and understand the behavior of others. When information about others is limited, people tend to egocentrically use information about themselves as a reference for making inferences about others (Tamir & Mitchell, 2013). Stated differently, people use their own preferences and values as an anchor to base their inferences about other people including strangers (Tamir & Mitchell, 2013), friends (Epley, Keysar, Van Boven, & Gilovich, 2004), and even members of rival groups (Chambers & Melnyk, 2006).

When people infer the values of rival groups, they tend to perceive greater conflict and disagreement than truly exists (Chambers & Melnyk, 2006). Such misperceptions are caused because people egocentrically focus on their own values more than their rival’s values (Chambers, Baron, & Inman, 2006). This egocentric focus leads people to overestimate the degree to which rival groups are opposed to their own most important values. To illustrate this point, consider the rival groups, prochoice and prolife advocates. People who are prochoice focus on their most important values (i.e., women’s reproductive rights). Similarly, people who are prolife focus on their own most important values (i.e., the value of human life). When prochoice advocates infer the values of their prolife rivals, they may egocentrically infer that their prolife rivals are more opposed to their own most important values (i.e., women’s reproductive rights) than they truly
are (Chambers et al., 2006). Similarly, when prolife people infer the values of their prochoice rivals, they may egocentrically infer that prochoice people are more opposed to their own most important values (i.e., the value of human life) than they truly are (Chambers et al., 2006).

In the current research, we extended the previous findings on egocentric social inferences to the realm of terrorism. Because prochoice and prolife advocates inferred greater opposition toward their own values, we hypothesized that participants would egocentrically infer that terrorist groups are opposed to their own values more than the values of others. Given the extreme nature of a terrorist attack, we also hypothesized that participants would infer that terrorist groups are opposed to personally important values more than personally unimportant values. To test these hypotheses, participants inferred the degree to which a hypothetical terrorist group opposed values that were self-relevant, other-relevant, important, and unimportant.

Method

Participants

Sixty-seven participants ($M_{age} = 19.64, SD_{age} = 4.31$) were recruited from lower level, undergraduate psychology courses at a midwestern university. All participants received a small amount of course credit for participating. The sample consisted of 37 women, 28 men, and two participants who chose not to identify their sex. Participants identified as European American (80%), Hispanic (8%), African American (8%), two or more races (3%) and other (unspecified; 1%). On a spectrum from 1 (liberal) to 5 (conservative), the average participants’ political beliefs were a 3.22 ($SD = 0.91$). In regard to political party affiliation, nearly 45% of the sample identified as Republican, 35% did not identify with any party, 12% identified as Democrat, and less than 3% identified with the Libertarian or Green Party.

Materials and Procedure

At the beginning of the study, participants were presented with an IRB approved informed consent form that explained participants’ rights during the study. All participants indicated that they read the informed consent form and voluntarily agreed to participate in the study. Afterward, participants were given a survey titled “Judgments About a Terrorist Group.” At the beginning of the survey, participants were led to believe that a new, relatively unknown domestic terrorist group had emerged in the United States. Participants were told that, “According to the U.S. National Security Agency, a domestic terrorist group is beginning to emerge that refers to itself as ‘The BCNC.’ Little information has been disclosed about The BCNC, so you may or may not have heard about it.” To avoid any demand characteristics, The BCNC was said to be supportive and opposed to many different values that may or not be related to participants’ personal values.

Next, participants inferred the values of the BCNC Group on a 7-point Likert-type scale, ranging from 1 (not at all opposed) to 7 (extremely opposed). To manipulate the domain relevance (self-relevant, other-relevant) and importance (important, unimportant) of the value, four questions were presented to participants. The first question asked, “To what extent do you think the values of the BCNC are opposed to your most important values?” The second question asked, “To what extent do you think the values of the BCNC are opposed to values that are unimportant to you?” The third question asked, “To what extent do you think the values of the BCNC are opposed to the most important values of someone else?” The fourth question asked, “To what extent do you think the values of the BCNC are opposed to values that are unimportant to someone else?” It is important to note that participants did not state what their personally important or unimportant values were nor what the important or unimportant values were for someone else.

After participants completed the value inference survey, they indicated the sex and ethnicity they identified most strongly with. Response options included man and woman for sex and European American, African American, Hispanic, Asian, two or more races, or other for ethnicity. Participants also indicated their age, political beliefs, and political party identification. Political beliefs were assessed by a single item, ranging from 1 (extremely liberal) to 5 (extremely conservative). Political parties included Republican, Democrat, Green, Libertarian, Other, or No Party Affiliation.

At the conclusion of the study, participants were presented with a one-page debriefing form. They were made aware of the deception used in the study, specifically that the BCNC was a fictitious terrorist group. Participants also learned the importance of the study and its implications for the public’s response to real-world terrorist attacks. Finally, the contact information of the principal
investigators, the IRB chair, and the student counseling center was provided to participants. Participants were granted extra credit for participating and dismissed.

Results

A repeated measures 2 X 2 factorial Analysis of Variance (ANOVA) was used to test whether people’s inferences regarding the values of a terrorist group varied as a function of domain-relevance (self-relevant, other-relevant) and importance (important, unimportant). As predicted, there was a main effect of domain-relevance. As seen in Figure 1, people inferred that the terrorist group was more opposed to their own values than the values of others, \( F(1, 65) = 5.66, p = .02, \eta^2_p = .08 \). There was also a main effect of importance. People inferred that a terrorist group was more opposed to important values than unimportant values, \( F(1, 65) = 7.67, p = .009, \eta^2_p = .10 \). There was no interaction between domain-relevance and importance in regard to people’s inferences of the terrorist group’s values, \( F(1, 65) = 1.28, p = .26, \eta^2_p = .02 \).

In an exploratory fashion, we also included participants’ political beliefs (grand mean centered) into the 2 (domain relevance: self-relevant, other-relevant) x 2 (importance: important, unimportant) repeated measures ANOVA model. There was no significant three-way interaction between political beliefs, domain relevance, and importance, \( F(4, 60) = 0.63, p = .84, \eta^2_p = .02 \). The main effects of domain relevance, \( F(1, 60) = 1.36, p = .25, \eta^2_p = .02 \), and importance, \( F(1, 60) = 1.59, p = .21, \eta^2_p = .03 \), were no longer significant, suggesting a small or underpowered effect. Furthermore, similar results were found when political party identification was used in the model instead of political beliefs. There was no three-way interaction between political party identification, domain relevance, and importance, \( F(5, 59) = 0.36, p = .88, \eta^2_p = .03 \). Main effects of domain relevance, \( F(1, 59) = 0.62, p = .43, \eta^2_p = .01 \), and importance, \( F(1, 59) = 1.12, p = .29, \eta^2_p = .02 \), were not significant.

Discussion

In the current study, we tested whether people’s inferences about the values of a terrorist group depend on the self-relevance and importance of the value. As predicted, both hypotheses were confirmed. People egocentrically inferred that a hypothetical terrorist group was more opposed to their own values than the values of someone else. Furthermore, participants inferred that a terrorist group was more opposed to important, rather than unimportant, values.

Implications and Future Directions

Although the current study assessed inferences of a hypothetical terrorist group, the results may have implications for inferences of real terrorist groups. Presumably, the average American knows very little about real terrorist groups. Indeed, the exact number of domestic and foreign terrorist groups is unknown. Therefore, people may be left to infer a real terrorist group’s motives and values after an attack occurs domestically (e.g., threatening political figures by the “Sovereign Citizens”) or abroad (e.g., beheading of journalists by “ISIS”). The extent to which egocentric thinking guides people’s inferences of real-world terrorist groups is a logical extension of the current research.

If egocentrism affects people’s inferences about terrorist groups, different people may have different inferences about the same terrorist group. For instance, Republicans may egocentrically infer that a domestic terrorist group is opposed to conservative, rather than liberal, values. On the other hand, Democrats may egocentrically infer that the same terrorist group is opposed to liberal, rather than conservative, values. Presumably, both Republicans and Democrats egocentrically think about terrorist groups in relation to their own values. Although the current research was not adequately powered to answer these questions, future studies may address the link between personal political beliefs and political value inferences.

Finally, the current findings have implications for the broader public. In the aftermath of...
a terrorist attack, people may need to adjust their inferences of the terrorist group to account for their own egocentric biases. Presumably, once people learn about the terrorist group’s values, they should no longer need to base their inferences upon their personal values. However, it is unclear to what extent people correct their inferences of terrorists’ values. In people’s search for answers, they may selectively seek out information that confirms their inferences (Swann, Stein-Seroussi, & Giesler, 1992). Indeed, people may misconstrue the values of terrorist groups and maintain inaccurate perceptions until they are corrected by unbiased, objective facts. It is also possible that such beliefs persevere and remain unchanged in light of evidence (Ross, Lepper, & Hubbard, 1975). Future research can better identify the extent to which people’s perceptions of terrorist groups are accurate.

**Limitations**

Given the lack of previous research regarding the inferred values of terrorist groups, we asked participants to infer values quite generally. In the study, participants were asked to infer the extent to which a terrorist group was opposed to their own, or another person’s, important or unimportant values. We did not specify which values (e.g., imperialism, honesty) participants should make inferences about. Therefore, the broad and general wording of our survey might have introduced demand characteristics. It is possible that participants felt compelled to infer that the terrorist group was opposed to important, rather than unimportant, values. If demand characteristics caused people to infer greater value opposition from terrorists generally, then people would infer that the terrorist group was equally opposed to important and self-relevant values more than unimportant and other-relevant values, respectively. Given the emergence of radical groups throughout the world, this research is applicable and timely. It appears that, when little or no information is available about a group, the term terrorist may affect people’s inferences about the group’s values.

**Conclusion**

The current study demonstrated that people egocentrically infer that a terrorist group is opposed to important and self-relevant values more than unimportant and other-relevant values, respectively. The results better fit an egocentrism account because participants felt compelled to infer that the terrorist group was equally opposed to their own values rather than another person’s values. Regardless, future studies that better control for demand characteristics are a useful extension of the current study. For example, using a deceptive cover story regarding the true purpose of the study may reduce potential demand characteristics (Swann, Stein-Seroussi, & Giesler, 1992). Indeed, people may misconstrue the values of terrorist groups and maintain inaccurate perceptions until they are corrected by unbiased, objective facts. It is also possible that such beliefs persevere and remain unchanged in light of evidence (Ross, Lepper, & Hubbard, 1975). Future research can better identify the extent to which people’s perceptions of terrorist groups are accurate.

**References**


**Author Note.** Nina M. Combs and Brian Collisson, Department of Psychological Sciences, Marian University, Indianapolis, IN.

Correspondence concerning this article should be addressed to Brian Collisson, Department of Psychological Sciences, Marian University, 3200 Cold Spring Road, Indianapolis, IN 46222. E-mail: bcollisson@marian.edu.
The belief in a higher power has been a topic for consideration and debate since very early in human history. For instance, ancient Greek philosopher Aristotle (1999) described God as the unmoved or prime mover, or the force that causes everything in the universe. Because this concept is considered a core component of the human condition, it became a topic of interest to social scientists in the early 1900s. One of the most influential researchers was William James, whose 1902 book *The Varieties of Religious Experience* was foundational in the fields of psychology and religion (James & Richardson, 2010). In this book, James explored the idea of religion as an experience someone has, rather than just an organizational affiliation.

In later decades, social scientists began to discuss spirituality as well as religion. For instance, Maslow (1971) believed spirituality was a vital aspect of the human experience. He viewed it as the process of achieving a person’s highest personal potential through self-actualization. Similarly, Rogers (1961) greatly valued spirituality. As a therapist, he encouraged his clients to attend to their spiritual health and saw it as crucial to their mental health. Rogers argued that spirituality could help replace negative feelings with more positive emotions.

Although many saw religion and spirituality as intertwined, Hill et al. (2000) argued that the terms *religion* and *spirituality* were different concepts. Much of Western society now views religion and spirituality as two separate entities (Marler & Hadaway, 2002; Pargament, 1999; Selvam, 2013). However, this goes against what many scholars have suggested, who saw religion and spirituality as overlapping concepts (Hill & Pargament, 2003; Hill et al., 2000; Zinnbauer et al., 1997).

Although many have suggested that religion and spirituality are seen as different in general,
little empirical research has looked at what specific values and characteristics people view as different between the two. Therefore, the goal of the current study was to analyze the values people expect of religious people in comparison to what people expect of spiritual people. This way, we could see in what ways these groups are seen as different as well as how their values are seen as overlapping.

Religion and Spirituality
Although heavily studied, there is not one definition of religion or spirituality that is agreed upon by all researchers. Scholars have argued that the definitions are unclear (Zinnbauer & Pargament, 2013; Zinnbauer et al., 1997). This may be due to disagreement among scholars in the field, but it may also be due to the multidimensional nature of the concepts. In other words, both religion and spirituality encompass a number of different factors and characteristics (Smith, 2007). For this reason, Hill et al. (2000) advised against determining one definitive definition and instead suggested focusing on the multidimensionality of the constructs.

When discussing the different dimensions of religion and spirituality, it becomes obvious that there is overlap between these two concepts. For instance, both religious and spiritual people value a connection to God or a higher power (Lun, 2015; Smith, 2007; Zinnbauer & Pargament, 2013; Zinnbauer, Pargament, & Scott, 1999). In addition to having an internal relationship with God or a higher power, both religion and spirituality also place an emphasis on personal beliefs and inner experiences (Donahue, 1985; Smith, 2007; Zinnbauer et al., 1997). But besides these internal aspects, both also place an emphasis on connections and relationships with other people (Burkhart & Solarí-Twadell, 2001; Smith, 2007; Zinnbauer & Pargament, 2013) such as looking for harmony with other people, and seeking meaning and purpose in their communities. In addition to overlap in the values of religious and spiritual people, there can also be overlap in how people self-identify. In other words, many people see themselves as both spiritual and religious (Marler & Hadaway, 2002; Shahabi et al., 2002; Zinnbauer et al., 1997). People are not forced to choose one over the other.

Despite the overlap, religion and spirituality are often seen as distinct from each other (Hill & Pargament, 2003; Marler & Hadaway, 2002; Smith, 2007; Zinnbauer et al., 1997). Although some researchers have argued that religion and spirituality could be distinct concepts (Del Rio & White, 2012; Helminiak, 2001; Legere, 1984), many have argued that seeing them as separate can cause problems in operationalizing their definitions (Hill et al., 2000; Smith, 2007). Selvam (2013) suggested that viewing these two concepts as separate is particularly predominant in Western cultures; in other cultures, the concepts of religion and spirituality are the same.

Some experts have studied the stereotypes and beliefs about religious individuals (Galen, Williams, & Ver Wey, 2014; Zafar & Ross, 2015) as well as beliefs about different religions (Hodge, Baughman, & Cummings, 2006). Additionally, researchers have studied what people expect of nonreligious individuals (Harper, 2007; Saroglou, Yzerbyt, & Kaschten, 2011). Less research has looked at the stereotypes and expectations people hold about spiritual individuals. Some have looked at beliefs about spirituality in relation to something else such as how spirituality can play a role in counseling (Cornish, Wade, & Post 2012; Henriksen, Polonyi, Bornsheuer-Boswell, Greger, & Watts, 2015). Overall, most of the research on stereotypes has focused on who is viewed favorably and who is viewed unfavorably.

Few have studied stereotyped expectations of religion and spirituality in the same study. Hodge et al. (2006) studied the stereotypes of religion and spirituality in social work textbooks, but did not differentiate between the two. This may have been because the social work textbooks endorse the idea that religion and spirituality are defined as the same (Marler & Hadaway, 2002; Shahabi et al., 2002; Zinnbauer et al., 1997). Lun (2015) analyzed the qualitative responses of 28 students about how they perceived the concept of religion and spirituality. They found many trends in the responses. For example, some respondents noted that Christianity is associated with icons such as the cross and Bible. Both religion and spirituality were associated with giving meaning to life. Similar to Hodge et al. (2006), Lun (2015) seemed to treat religion and spirituality somewhat interchangeably and not as separate concepts.

Although some research has looked at stereotypes of religious and spiritual people, less research has analyzed the specific stereotypes of what religious and spiritual people value. However, there is one trend that emerged. Religion is often viewed as an external phenomenon, but spirituality is viewed as an internal phenomenon (Hyman & Handal, 2006; Smith, 2007; Zinnbauer et al., 1997). In other words, religion is often associated with external
activities such as church attendance, but spirituality is thought of as an internal experience. Lun’s (2015) research did look at some characteristics of religiosity and spirituality because she examined expectations of both groups in her study. Although Lun’s (2015) qualitative research design gave an in-depth picture of how these people are viewed, it did not allow for strong statements as to how these groups are seen as similar and different. The question of how religion and spirituality are expected is especially important to ask among young adults because many have become interested in spirituality in addition to traditional religion (Zinnbauer et al., 1997). Therefore, the current study aimed to examine the expectations of what young adults believed spiritual and religious people valued.

Values and Characteristics People Expect of Religious and Spiritual People

Religion is often expected to be a stereotypically external phenomenon (Hyman & Handal, 2006; Smith, 2007; Zinnbauer et al., 1997), and one common aspect frequently associated with religiosity is church or temple attendance. For instance, in studies of self-identified religious and/or spiritual people, individuals have been asked to define what religion and spirituality mean to them (Shahabi et al., 2002; Zinnbauer et al., 1997). In the studies, participants included church attendance in their personal definitions of religion, but not in their definitions of spirituality. Similarly, Hyman and Handal (2006) asked religious professionals to define the two terms, and they associated church attendance with religiosity more so than with spirituality. Finally, in a review of how the research literature defines the terms, church attendance is more often associated with religion than with spirituality (Smith, 2007).

There are many reasons why attendance at church or temple is associated more with religion and less with spirituality. Attendance at some kind of church or temple is part of most major world religions. For instance, meeting with others is discussed in the Bible (Hebrews 10:25 New International Version). Attendance at services is a way to externally demonstrate an individual’s commitment to religion. It is also a way to build a religious community, which provides social support (Burkhart & Solari-Twadell, 2001; Hummer, Rogers, Nam, & Ellison, 1999; Smith, 2007; Zinnbauer & Pargament, 2013). Spirituality, on the other hand, is often thought of as a private and individual experience (Helminiak, 2001; Richards & Bergin, 1997). This may lead to people not associating spirituality with attending church. People who view spirituality as an internal phenomenon may not believe there is a need for spiritual individuals to attend church. Therefore, the first hypothesis of the current study was that religious individuals would be expected to attend church services or temple on a regular basis more than spiritual individuals.

A logical extension to church attendance is the beliefs and doctrine that go along with church. Because of this, many have associated religiosity with an organized set of doctrine and beliefs (Del Rio & White, 2012; Meagher & Kenny, 2013; Smith, 2007). Similar to church attendance, many studies have also found that people generally include doctrine and beliefs in their definitions of religion as opposed to their definitions of spirituality (Hyman & Handal, 2006; Zinnbauer et al., 1997). One study even found that people who were connected to a church valued obedience to doctrine and beliefs over many other factors including intellectual autonomy (Ellison & Sherkat, 1993).

Although religions often have a set of doctrine associated with them, the concept of spirituality does not have one set of beliefs associated with it. Koenig, McCullough, and Larson (2001) suggested that spirituality is more subjective than religiosity. Instead of there being only one objective form, spirituality can be seen as the subjective beliefs of the individual person. This can lead to multiple expectations about spirituality among different types of people. Therefore, for our second hypothesis, we suggested that religious individuals would be expected to place importance on adhering to doctrine or rules more than spiritual individuals.

Prayer is defined as connecting and communicating with God or some sort of higher being (Gubi, 2001) and is often associated with religion. For instance, people often include prayer in their definitions of religion but not spirituality (Shahabi et al., 2002; Zinnbauer et al., 1997). This association between religion and prayer is so common that it is often used to measure people’s religiosity (Ap Siôn & Nash, 2013; Breslin & Lewis, 2008; Meraviglia, 1999).

There are reasons why people may expect prayer to be associated with religion more than spirituality. Prayer is an external action often, though not exclusively, performed as part of church or temple ceremonies (Lewis & Malbty, 1996; Zinnbauer et al., 1997), which as discussed, are often associated with religion (Hyman & Handal,
2006; Shahabi et al., 2002; Smith, 2007). Spiritual people may also be expected to pray (Meraviglia, 1999), but the word prayer is often less associated with spirituality. Instead, the word meditation is more commonly associated with spirituality (Wachholtz & Pargament, 2005; Zumeta, 1993). Although there is much overlap between prayer and meditation, the stereotype about the word prayer is still more closely associated with religiosity. Therefore, the third hypothesis of the current study was that religious individuals would be expected to place importance on prayer more than spiritual individuals.

Spirituality is often discussed in relation to the sacred, which is the pursuit of a divine or supernatural power (Hill et al., 2000; Pargament, 1999). Many have argued that connection with God is a core component of spirituality (Del Rio & White, 2012; Meraviglia, 1999). In Hyman and Handal’s (2006) study of religious professionals, they included “connection to God” in their definitions of spirituality, but not religiosity. In Smith’s (2007) review of how the research literature defines the terms, connectedness, which may include both a connection to God as well as a connection to others, is associated with spirituality. However, no equivalent term is associated with religion. Hill and Pargament (2003) listed four scales that measure “closeness to God.” Interestingly, three of the scales had spirituality in the title but only one measure had religion in the title. Although it is likely that these authors saw an overlap between religion and spirituality, it does suggest that there is an association between spirituality and a connection to God or higher power.

However, this association is not exclusively with spirituality; many have linked religion to a connection with God as well. Allport and Ross (1967) discussed the importance of internal religiousness, which refers to personal beliefs about God. This was compared to external religiousness, or the institutional aspects of religion, and suggested that a connection with God or a higher power is an important part of religiousity. Also, in Zinnbauer et al.’s (1997) study of how people defined religion and spirituality, connection to God was associated with both. Therefore, there have been arguments that closeness to God is part of both religion and spirituality. However, because of the stereotype that religion is more external than internal (Hyman & Handal, 2006; Smith, 2007; Zinnbauer et al., 1997), it is likely that people will expect a connection to God to be more associated with spirituality than with religion. Therefore, for our fourth hypothesis, we predicted that spiritual individuals would be expected to place importance on a close connection to God more than religious individuals.

Thus far, we have discussed four important values associated with religion and spirituality: church/service attendance, adherence to doctrine, prayer, and connection to God. However, there are many other values associated with these topics. For instance, previous research has highlighted conversion and sharing one’s beliefs (Rambo & Farhadian, 2014), regret and repentance (Randall, 2013), and altruism (Huber & MacDonald, 2012). Religious education (and more generally, gaining knowledge and wisdom about one’s beliefs) is important enough to warrant its own association (“Religious Education Association,” n.d.). Although these values are often associated with religion and/or spirituality, there is no rationale to suggest that any of them would be associated more so with one concept over the other. Therefore, as exploratory hypotheses, we also examined the expectations participants had about religious and spiritual people with regard to these other variables.

The Current Study
The goal of the current study was to measure the expectations young adults have about values of religious people and spiritual people. Participants randomly completed one of two surveys. In one survey, participants characterized religious individuals, whereas in the other survey, participants characterized spiritual individuals. In addition to expectations about the hypothesized variables, we also studied exploratory variables to develop a broader understanding of what young adults expected about these two groups.

Method
Participants
Two hundred forty-four participants were recruited. Nineteen participants were removed from the sample due to missing data, resulting in 225 participants in the final sample. Sixty-nine percent of participants identified as women, 26.5% identified as men, and 4.5% marked “other/prefer not to answer.” The sample was primarily White (66.4%), followed by Hispanic (9.3%), Asian (8%), Black (5.3%), mixed race/other (7.1%), Middle Eastern (1.3%), Pacific Islander (1.3%), and 1.3% who chose not to respond. The mean age was 19.54 (SD = 2.32). Participants identified as Christian...
Religion and Spirituality Among Young Adults | Kobza and Salter

(70.4%), spiritual (5.3%), nonaffiliated (4.4%), other (4.4%), Atheist (4%), Jewish (3.5%), Agnostic (3.5%), Muslim (1.8%), Buddhist (1.3%), and Hindu (0.4%). Some chose not to identify their religious or spiritual affiliation (0.9%). The political identity of participants was also collected, which included liberal (33.7%), “in-the-middle” (29.6%), other (23%), conservative (13.3%), and 0.4% who chose not to respond.

Materials
The current study was a between-subjects design. The independent variable was the usage of the word religious or the usage of the word spiritual. The first survey asked participants what they believed to be the values of religious individuals. The second survey asked to characterize spiritual people. One hundred twelve participants took the religious people survey and the other 113 took the spiritual people survey. See Table 1 for the 11 Likert-type scale questions related to the four main study hypotheses as well as the exploratory hypotheses. These questions were derived from research on the topic as previously discussed, and also resulted from a consultation with a local religious authority (N. Grevas, personal communication, March 1, 2014). He discussed the survey with us, suggested what values were important, and generally gave input on how to best analyze religion and spirituality. For each of the 11 questions, participants were asked “How important is each of the following for a (religious/spiritual) person?” On the response scale, 1 indicated strongly disagree and 7 indicated strongly agree. In addition to the close-ended questions, an open-ended question asked, “In the box below, please describe what it means to be a religious/spiritual person. What are the characteristics and behaviors of a religious/spiritual person? Feel free to be as brief or descriptive as you desire.” The end of the survey asked demographic questions regarding participant’s sex, race, age, religious affiliation, and political ideology.

| TABLE 1 |
| Differences in Expected Values of Religious and Spiritual Individuals |
| Statement | Religious individuals | Spiritual individuals | t test | 95% CI of the difference | Effect size |
| | M | SD | M | SD | | |
| Should attend church service or temple on a regular basis | 4.41 | 1.78 | 3.58 | 1.80 | t(223) = 3.47 | p = .001 | [0.38, 1.23] | d = .46 |
| Adhere to the doctrine/rules of whatever the beliefs the person is most affiliated with | 5.16 | 1.35 | 4.82 | 1.56 | t(223) = 1.73 | p = .08 | [-0.05, 0.72] | d = .23 |
| Prayer | 5.47 | 1.15 | 4.59 | 1.69 | t(222) = 4.53 | p < .001 | [0.50, 1.26] | d = .61 |
| Be exposed to some kind of (religious/spiritual) education | 5.00 | 1.45 | 4.53 | 1.51 | t(221) = 2.36 | p = .02 | [0.08, 0.86] | d = .32 |
| Perform good deeds and altruistic actions | 5.04 | 1.39 | 4.63 | 1.57 | t(223) = 2.11 | p = .04 | [0.03, 0.81] | d = .28 |
| Repent | 5.14 | 1.36 | 4.84 | 1.51 | t(223) = 1.58 | p = .12 | [-0.08, 0.68] | d = .21 |
| Believe they have a close connection to God | 5.02 | 1.54 | 4.35 | 1.91 | t(223) = 2.91 | p = .004 | [0.22, 1.23] | d = .39 |
| Feel regret for wrongdoings | 5.08 | 1.48 | 5.04 | 1.45 | t(223) = 0.23 | p = .82 | [-0.34, 0.43] | d = .03 |
| Convert other people to their (religious/spiritual) affiliation | 2.46 | 1.52 | 2.33 | 1.53 | t(223) = 0.63 | p = .53 | [-0.27, 0.53] | d = .08 |
| Gain knowledge and wisdom | 5.35 | 1.24 | 5.49 | 1.30 | t(221) = 0.78 | p = .44 | [-0.47, 0.20] | d = .10 |
| Share views | 3.22 | 1.67 | 3.50 | 1.74 | t(223) = 1.20 | p = .23 | [-0.72, 0.18] | d = .16 |

Note: Participants were asked “How important is each of the following for a (religious/spiritual) person?”
Procedure
Institutional review board approval was received for the current study (IRB #184). To recruit participants, students in Introductory Psychology courses received academic credit for participation using the Psychology Subject Pool. The surveys were completed electronically using the online survey website Qualtrics®. Participants were not required to complete the survey in the lab but were instead allowed to complete it wherever they wished. The survey began with an informed consent form explaining the study to participants. Next, participants were randomly assigned to one of the two surveys. Their survey included all the questions stated above in the Materials section. Finally, they were debriefed and thanked.

To receive academic credit, participants were directed to a second survey automatically after they finished the survey. This second survey asked their name, what psychology class they were taking, and what instructor taught the course. When the respondent finished the second survey, they were able to receive academic credit. Because this information was in a separate survey, the participants’ names were not attached to the survey that asked the main study questions. Therefore, we could not identify individual responses, ensuring the participants’ survey responses remained confidential.

Results
Before testing the hypotheses, we compared the two conditions (i.e., being presented with the word religious or being presented with the word spiritual) to see if they differed in any demographics. There were no differences across the two conditions in sex, race, age, religious affiliation, or political ideology. Therefore, this suggested that the conditions were equivalent in these factors (and therefore the only difference was likely the independent variable).

Table 1 shows the t-test results for all values examined in the present study. Due to the high number of statistical tests conducted, a lower p value (.01) was used to determine significance. Hypothesis 1, which suggested that religious individuals would be expected to attend church services or temple on a regular basis more than spiritual individuals, was supported. Hypothesis 2 was not supported. There was no difference in the expected importance of adhering to doctrine or rules between religious and spiritual individuals. Hypothesis 3 was supported. Religious individuals were expected to place importance on prayer more than spiritual individuals. Finally, Hypothesis 4 was not supported. We hypothesized that spiritual individuals would be expected to believe they have a close connection to God more than religious individuals. Instead, this t test was significant in the opposite direction of what was hypothesized. Religious individuals were expected to place significantly more importance on a connection to God than spiritual individuals.

Table 1 also shows the results of the exploratory hypotheses. As stated, these hypotheses explored values often associated with religion and/or spirituality, but previous literature does not indicate that one concept would value these traits over the other. Most of these values did not result in significant differences across the two conditions. The exceptions were the value of altruism. For both of these variables, participants expected religious people to value them more than they expected spiritual people to do so.

Discussion
The purpose of the present study was to see if young adults expected the core values of religious people to be different than the core values of spiritual people. In particular, this study looked at how religiosity and spirituality was associated with the following: church attendance, adherence to religious doctrine, prayer, and connection to God or a higher power. Three out of the four hypotheses resulted in statistically significant differences, although some of the results were unexpected. In addition, exploratory hypotheses found that there are many values expected to be similar in both religious and spiritual people. Therefore, these findings could lead to some insight regarding the expectations of religion and spirituality.

Study Findings
Some of our hypotheses were supported. Hypothesis 1 was supported, which stated that religious people would be expected to attend church services or temple on a regular basis more than spiritual people. Hypothesis 3 was also supported, which stated that religious people would be expected to place importance on prayer more than spiritual individuals. These findings supported previous research, which suggested that religion is often associated with church attendance (Hyman & Handal, 2006; Shahabi et al., 2002; Smith, 2007; Zinnbauer et al., 1997) as well as prayer (Ap Siôn & Nash, 2013; Breslin & Lewis, 2008; Meraviglia,
Religion and Spirituality Among Young Adults | Kobza and Salter

Although religious individuals were expected to place more importance on church attendance and prayer than spiritual individuals, the current study did not claim that spirituality was expected to be unrelated to church attendance or prayer. In our study, we found that respondents did expect these activities to be important to spiritual people, but less than for religious people. This supported the suggestion that religion and spirituality may have similar central traits of characterization (Hill & Pargament, 2003; Hill et al., 2000; Zinnbauer et al., 1997).

However, not all our hypotheses were supported. For instance, Hypothesis 2 suggested that religious individuals would be expected to place more importance on adhering to doctrine or rules than spiritual individuals, but our results found no difference between the two groups. We predicted this because many researchers have argued that adherence to tradition or an external belief system is a crucial element to religiosity (Del Rio & White, 2012; Ellison & Sherkat, 1993; Hyman & Handal, 2006; Meagher & Kenny, 2013; Smith, 2007; Zinnbauer et al., 1997). Researchers often associate spirituality more with pursuing the sacred and less on doctrines and beliefs (Hyman & Handal, 2006; Zinnbauer et al., 1997). When participants were asked to characterize spirituality, some expected spiritual people to have less structure in their belief system. Specifically, one said, “To me, anyone who believes in something as vague as a ‘force’ like fate or karma is spiritual.” Another participant noted that for spirituality “there are no rules.”

However, the prediction about doctrine and beliefs was not supported; the two groups were seen as placing the same amount of importance on this characteristic. This could be for several reasons. Just because some have suggested that spirituality is somewhat subjective (Koenig et al., 2001) does not mean that spiritual people are expected to have no beliefs. As stated, spirituality is often thought of as private and individualistic (Helminiak, 2001; Richards & Bergin, 1997). Perhaps spiritual people are expected to have their own private and individually defined set of beliefs and adhere to them to the same degree that religious people adhere to the beliefs of their church or temple. Another possible reason why the two groups were seen as placing the same amount of importance on doctrine could be a wording issue. Doctrine is a word not commonly used among young adults, and some participants mentioned that this word was confusing. Perhaps if a different word was used, the results would be different.

Hypothesis 4 was also not supported. We predicted that participants would expect spiritual individuals to have a closer connection to God than religious individuals. However, the opposite was found. Religious individuals were expected to have a close connection to God significantly more than spiritual individuals. This hypothesis was based on the notion that connecting with some kind of sacred or higher power is often seen as a crucial aspect of spirituality (Emmons & Crumpler, 1999; Hill et al., 2000; Marler & Hadaway, 2002). In the open-ended question that asked participants to characterize spirituality, many participants viewed spirituality as a connection to some general higher force. In fact, one respondent noted that a spiritual person “believes and feels the presence of an unseen being or life form that he or she holds to some sacred regard.”

However, as discussed, many have argued that a connection to God is also important to religious people (Allport & Ross, 1967; Zinnbauer et al., 1997). In our findings, some respondents said they characterized religious people as having a connection to God. One responded reported “I feel that a religious person is one who has a personal relationship with God.” It appears that, in our sample, the association between religion and a connection to God was stronger than the association between spirituality and a connection to God. Another possible reason for this might have been a wording issue. In our survey, we asked about a connection to “God.” This is a phrase that research has connected with spirituality and religion, but our participants might have associated “God” with “religion.” We might have found different results if we had asked about a connection to “God or a higher power.”

Our exploratory hypotheses also yielded interesting results. For instance, there were no differences in expectations of how religious and spiritual people were characterized with the following: valuing regret of wrongdoing, sharing views, converting others, gaining knowledge, or repenting of wrongdoings. This supported the idea that religion and spirituality are not always seen as separate from each other. This was not surprising because many researchers have argued that they are not completely different from each other (Hill & Pargament, 2003; Hill et al., 2000; Zinnbauer et al., 1997), and many people identify as both (Marler & Hadaway, 2002; Shahabi et al., 2002; Zinnbauer et al., 1997).
Ultimately, these hypotheses help reinforce several conclusions. First, religious and spiritual people can be seen as having overlapping characterizations. In our study, many values were seen as important to both religious and spiritual people. Many have argued that society views these concepts as separate from each other (Marler & Hadaway, 2002; Pargament, 1999; Selvam, 2013). Our results seemed to suggest that people may agree with previous scholars who believe there is overlap in the definitions of these concepts (Hill & Pargament, 2003; Hill et al., 2000; Zinnbauer et al., 1997).

Another conclusion of our study was that religiosity can be seen as having both external and internal values. As stated, religion is often stereotyped as something that is external, and spirituality is often stereotyped as something that is internal (Hyman & Handal, 2006; Smith, 2007; Zinnbauer et al., 1997). In our study, we found these stereotypes to be invalid. For example, in our study, religion was more strongly associated with a connection to God, which could be argued to be more internal than external, than spirituality. Some have argued that religiosity can be associated with internal beliefs (Allport & Ross, 1967; Donahue, 1985; Emmons & Crumpler, 1999) and that attending church, which could be argued to be more external than internal, can be a spiritual experience (Hill et al., 2000). Therefore, our study supported researchers who have suggested that both religion and spirituality can have internal and external qualities.

A final conclusion of our study was about the relationship between religion and spirituality. Some have noted a hierarchy regarding religion and spirituality (Hill & Pargament, 2003; Selvam, 2013; Zinnbauer & Pargament, 2013). Specifically, some have thought that spirituality is superior to religiosity because spirituality is tied with the internal (Hyman & Handal, 2006; Smith, 2007; Zinnbauer et al., 1997) and is considered superior to the external, which is associated with religion. They cautioned that this idea was biased, a value judgment, and not accurate. The findings of our study seemed to show that both religiosity and spirituality are expected to have important values. In fact, religious people were expected to place greater importance on the hypothesized values than spiritual people. We argue that these results reinforced the notion that one concept is in no way superior to the other.

Our findings were also interesting because the sample was young adults. We believe it is possible that young adults today feel differently about these topics than the young adults of previous years. The present study added to the research on the topic, and will possibly inspire future researchers to analyze the topics of religion and spirituality in young adults. According to Smith et al. (2003), many assumptions about young adults are not based on empirical data. We hoped to help bridge the gap between what researchers think about religion and spirituality, and what young adults think about the topic.

Limitations and Future Research

There were limitations to our study, which suggest ideas for future research. First, our study measured how people characterized religious or spiritual people, but it did not examine how young adults characterized people who are religious and spiritual. As previously discussed, many people personally identify as both (Marler & Hadaway, 2002; Shahabi et al., 2002; Zinnbauer et al., 1997). When answering the questions in our study, some might have thought of the two as the same. For instance, one participant noted, “The word spiritual always meant church and religion.” However, not everyone felt this way. Another participant said, “To be spiritual and religious are two very different things.” Therefore, a study that focuses on expectations of people who are both religious and spiritual would be beneficial to the research literature.

The outcome variables were also limitations to the current study. In this survey, only single-item responses were used to gauge the characteristics of religion and spirituality. This was done to quickly gain a broad understanding of what young adults believed religious and spiritual people valued. However, single-item measures can lead to limited or questionable reliability (Diamantopoulos, Sarstedt, Fuchs, Wilczynski, & Kaiser, 2012; Fisher, Matthews, & Gibbons, 2016). In the future, a study with multiple-item dependent variables would be beneficial.

Finally, another avenue for future research is to continue to study this topic qualitatively such as was conducted by Lun (2015). Although the present study had one qualitative question, the bulk of the research was a quantitative assessment. According to Smith (2010), it is important for researchers to contribute data that is a “sheer, accurate, rich description of social life” (p. 585). This means that future studies need to “go deeper” with how people self-identify or characterize religion and spirituality. In our study, one open-ended question
was not enough to give a detailed account of how young adults characterize religion and spirituality. In the future, focus groups or surveys that ask multiple open-ended questions could be used to accomplish this goal.

**Conclusion**

The definitions of religion and spirituality are complex, multidimensional, and overlapping (Hill et al., 2000; Smith, 2007). Our study found that people expect religious and spiritual people to have similarities and differences in what they value. This suggests that many people may not view religion and spirituality as completely distinct concepts. As discussed by previous researchers, many argue that people can be religious and spiritual. This study hopes to encourage future scholars to analyze specific traits related to religion and spirituality among young adults, along with analyzing specific stereotypes and expectations about these concepts.

**References**


Kobza and Salter | Religion and Spirituality Among Young Adults


Author Note. Matthew Kobza and Nicholas P. Salter, Department of Psychology, Ramapo College of New Jersey.

Matthew Kobza is now at Department of Counseling and Educational Leadership, Montclair State University.

Correspondence concerning this article should be addressed to Matthew Kobza, Department of Counseling and Educational Leadership, Montclair State University, Montclair, NJ 07043. E-mail: Kobzam1@montclair.edu
The term globalization has become increasingly popular over the past decade. It now represents the increasing interconnectedness of societies (Steger, 2009). Countries’ connectivity has been amplified through changes in economic trends, consumer behavior, trade, capital flows, migration, and technology (Jordá & Sarabia, 2014). As globalization brings about increased change and opportunities, it also leads to increased risk and instability (Lee & McNulty, 2003). Companies attempt to advertise their products to an entire society, but as global connectivity increases, the probability of relating to every culture becomes more difficult. Within the realm of consumer culture and advertising strategies, opinions vary on the best method of connecting to a greater population. Some believe that globalization has created Global Consumer Culture Positioning (GCCP), a shared global view toward products (Alden, Steenkamp, & Batra, 1999; Jiang & Wei, 2012). As such, this type of advertising strategy highlights brands by using globally understood meanings and cultural values. Others have focused on an increasing shift toward Western behavior and values (Gupta & De, 2007; Lee & McNulty, 2003). The current study analyzed the changes in the cultural values expressed through advertisement over time in relation to increases in globalization.

GCCP, in accordance to Jiang and Wei’s (2012) findings, is associated with the global view of advertisements. Alden et al. (1999) found support that GCCP is a useful global positioning tool for marketers to present their products as being
“global.” The main goal for positioning a brand globally is to have it be viewed as a membership into an elite club (Hannerz, 1990). Advertising in this way fosters power and value toward a brand (Shocker, Srivastava, & Reukert, 1994). GCCP is used as an indirect, soft-sell approach and also focuses on high technology items, popular to the younger modern population (Alden et al., 1999). Therefore, the presentation of indirect youthful ideas in advertisements displays GCCP’s emphasis on a global consumer culture. As such, positioning advertisements according to GCCP allows for a larger audience in that the messages conveyed are similar cross-culturally. Overall, in relation to multiple cultures, GCCP has been found in 26.5% of advertisements in other countries, although only 5.5% of U.S. advertisements display GCCP (Alden et al., 1999). Focusing on different cultures, Jiang and Wei (2012) found an increase in the similarity of advertisements cross-culturally in North America, Europe, and Asia supporting the globalization of ads. The research further focused on the time period prior to the 2008 economic crash because it was influential on the global perspective of world market connectivity (Jiang & Wei, 2012). The results indicated an increasing trend of similar values being expressed internationally in relation to global consumer culture. There were no differences in the values presented in China or in U.S. advertisements. Further, a specific shift toward GCCP was evident through the representation of similar values across different cultures in advertisements.

Instead of concentrating on the positioning of advertisements, others have studied the shifts in cultural variables, specifically collectivism and individualism. Collectivism is the increased focus on the group over the individual (Hofstede, 1997). On the other hand, individualism is represented through individuals’ greater degree of separation between self-perceptions and their culture (Hofstede, 1997). Hofstede (1997), along with Franke, Hofstede, and Bond’s (1991) research from 53 countries, found that national wealth is highly correlated with individualism. As such, it can be understood that, as economic growth increases, so does individualism. Hofstede (1997) predicted that, if any cultural dimension were to change, it would be individualism, because it has a direct relationship with the wealth of a country. As an example, Lee and McNulty (2003) found a major shift in Korea’s cultural and consumer values after their economic crash of 1997. In their study, they analyzed the reasoning behind the downturn and Korea’s ability to overcome it. They found that Korea’s collectivist approach to business was impeding their global economic growth into an economy influenced by Western individualism. Simultaneously, Korea shifting their business approach toward individualistic values, the Korean economic market quickly recovered. This suggested a connection between an individualistic approach and economic success (Lee & McNulty, 2003). Further research by Gupta and De (2007) revealed a similar shift toward individualism in India during a peak in globalization. They found an increase in the importance of the individual through 1947 to 2005.

Based on the evidence, there is a discrepancy in the way GCCP and Western individualism would advertise a product. GCCP would emphasize products using globally shared ideas, while still maintaining culture specificity through differences for collectivistic and individualistic cultures. Opposing this, a shift toward Western values would focus exclusively on individualistic advertisement strategies. To test this hypothesis in the present study, a sample of both U.S. and Chinese television commercials were chosen. It is possible to develop a comparative analysis focusing on these two leading countries due to their differences in cultural values, the leading ideals for thought and action within a society (Srikandath, 1991). The United States is primarily represented as an individualistic country, and China as a collectivist country (Hofstede, 1997). According to the Hofstede Center, the United States scored 91 out of 100 on individualism and China scored 20. However, focusing on advertising strategies in accordance with GCCP and shifts in individualism, traditional, and cultural values can fluctuate.

Cultural Values and Advertising
A substantial amount of research has focused on the cultural values present in advertisements. Mueller (1987) outlined 10 traditional appeals that express culture through ads: group/consensus versus individual/independence, soft-sell versus hard-sell, veneration of older adults/traditional versus modernity/youth, status appeal perspective, product merit perspective, oneness with nature appeal, or time-orientation appeal. On the basis of these appeals, it is possible to measure cultural variations in advertisements, and their shifts over time. Multiple researchers have identified culture specific indicators of Mueller’s appeals in both Western and East Asian countries. For example, Lin...
Cross-Cultural Globalization of Advertisements

Allen, Lee, and Escalera

(2001) studied Mueller’s traditional appeals present in U.S. and Chinese advertisements. Lin’s work demonstrated subtle changes in cultural values and advertisement strategies in 1998; the current study aimed to extend Lin’s work to contemporary times.

**Group/consensus appeal.** In accordance with Mueller’s (1987) definition, group/consensus appeal focuses on the individual as being part of the whole and conforming to the group. While comparing Japanese (a collective culture) and U.S. (an individualistic culture) commercials, Lin (1993) argued against the complete standardization of advertisements, because she identified a more collective approach in Japan. In other words, group/consensus appeal was stronger in Japanese culture. Similar to Lin’s study, Maynard and Taylor (1999) further found that Eastern cultural values represented collectivism.

**Individual/independence appeal.** This appeal shifts the focus away from the group and toward the individual (Mueller, 1987). Research has identified that this appeal is more common in Western and individualistic cultures (Lin, 1993; Mueller, 1987). In a recent study of 2,158 print advertisements, Okigbo, Martin, and Amienyi (2005) found that American magazines demonstrate U.S. cultural values, the most common cultural values being individualism, low context communication, and product merit. With regard to both appeals, evidence has shown that there is a difference between Western and Eastern advertisements. To solidify this, Kalliny and Ghanem (2009) surveyed the Middle East (a collective culture) and U.S. advertising agencies to test for the approach taken on advertisement creation. They found that these agencies use culture specific appeals because one focuses on group/consensus and the other on individual/independence.

**Soft-sell appeal and hard-sell appeal.** Soft-sell appeal is characterized by ads that set a tranquil mood and atmosphere by emphasizing the storyline instead of the products. On the other hand, hard-sell appeal focuses on branding, merit, and recommendation of the product (Mueller, 1987). Early findings on these opposing appeals have shown that hard-sell appeal is seen more frequently in an individualistic market, whereas soft-sell appeal is more common in collectivistic cultures (Lin, 2001; Maynard & Taylor, 1999; Mueller, 1987). More recent research has suggested that there has been a shift toward a global perspective on hard-sell and soft-sell. Okazaki, Mueller, and Diehl (2013) examined the effectiveness of these appeals in a global market between France, Italy, Japan, Germany, and the United States. They found that the soft-sell appeal is more effective and tends to induce less irritation compared to hard-sell commercials.

**Veneration of older adults/tradition appeal.** Mueller (1987) defined this appeal as a respect toward older adults’ wisdom and traditional values. Previous research has found that tradition is a key factor in many Asian advertisements. Belk and Pol- lay (1985), as well as Mueller (1992), later found that, although Japan appeared to become more “Americanized” between 1953 and 1988, the importance of traditional values expressed through their ads was still strong and even increasing. Similarly, research by Lin (2001) showed that, compared to the United States, Chinese ads displayed more veneration of older adults. Current research has recognized a change in Chinese commercials over time. Zhang and Harwood (2004) identified five dominant themes in Chinese commercials: product quality/effectiveness, family, modernity, beauty/youth, and pleasure. They showed that the importance of family was still a leading traditional appeal (Zhang & Harwood, 2004). However, there has been a shift toward utility, youth, and modernity. This is consistent with the continual growth and ever changing global market.

**Modernity/youth appeal.** Modernity/youth appeal represents a younger generation through the emphasis on contemporariness and antiaging benefits (Mueller, 1987). Lin (1993) found that a younger and more modern appeal is typical of Western countries, representing a more individualistic culture. However, Lin (2001) found no significant difference between Chinese and U.S. use of modernity/youth appeal.

**Status appeal and oneness with nature appeal.** Status appeal suggests that individuals can increase their position or rank within the opinion of others (Mueller, 1987). Mueller (1987) demonstrated that Japan, being a collective culture, had a greater inclination to appeals on social status and oneness with nature than the United States. The oneness with nature appeal was defined by Mueller (1987) as the importance of the interaction between people and nature. Similarly, Cheng and Schweitzer (1996) identified that advertisements in China used traditional appeals like status and oneness with nature more readily compared to the United States. They also found that the United States being more individualistic reflected manipulation and control over nature in their ads. However, Lin (2001) found support only for the fact that Chinese commercials
used more oneness with nature appeal, but not status appeal compared to the United States.

**Product merit appeal.** This appeal is characterized by an ad that describes a product and its merit in depth (Mueller, 1987). China has an inclination toward a soft-sell approach, thus they also often place less information in their ads (Lin, 1993). Lin (1993) concluded that this was because China’s traditional values consider it impolite to be direct or “boastful.” Similarly, Lin (2001) found that U.S. commercials use more product merit appeals than China.

**Time-orientation appeal.** Time-orientation appeal refers to advertisements that focus on the importance of time and how the product can save or generate more time (Mueller, 1987). Similar to the concept of economic growth and shifts in individualism, Bruneau (1979) found that time perception and orientation was highly influenced by the amount of industrialization in a culture. Furthermore, Lin (2001) identified that U.S. commercials were characterized by more time-orientation than China.

**The Current Study**

Similar to the 1997 Korean crisis (Lee & McNulty, 2003), the 2008 global economic downturn demonstrated the world’s connectedness. Prasad (2009) revealed that the Chinese and U.S. economies had a close relationship. He predicted that the 2008 crisis, which brought the United States and China under the spotlight, would increase the two countries’ close relationship (Prasad, 2009). This was demonstrated in 2010 when China became the second largest global economy, surpassing Japan and trailing the United States (Wang & Praet, 2012). Because of the exponential changes in both the Chinese and U.S. global markets, our study was conducted to replicate Lin’s (2001) research. Her research examined a cultural comparison between Chinese and U.S. cultural values as reflected through advertising appeals. Lin’s (2001) main finding was that both Chinese and U.S. advertisements remained consistent in their traditional portrayal of cultural values. In taking Lin’s findings and our current findings, we tracked the changes and drew a comparison over time.

Lin (2001) identified and tested nine hypotheses to measure the main differences in cultural values expressed in advertisements in China and the United States, all of which were tested again in the present study. Specifically, we expected that (H1) U.S. commercials would have a higher ratio of hard-sell to soft-sell appeals than their Chinese counterparts, (H2) Chinese commercials would portray more veneration of older people models than their U.S. counterparts, (H3) U.S. commercials would use more time-oriented advertisements than their Chinese counterparts, (H4) Chinese commercials would use more group consensus appeals than their U.S. counterparts, (H5) U.S. commercials would use more individual/independence appeals than their Chinese counterparts, (H6) U.S. commercials would use more modernity/youth appeals than their Chinese counterparts, (H7) U.S. commercials would use more product merit appeals than their Chinese counterparts, (H8) Chinese commercials would use more status appeals than their U.S. counterparts, and (H9) Chinese commercials were more likely to reflect oneness with nature appeals than their U.S. counterparts. Out of the nine hypotheses, Lin found support for seven. There was no support for H6 or H8.

In addition, the present study examined whether there has been a shift in cultural values either toward Western values or a global consumer culture. Thus, we hypothesized (H10) that a change in U.S. and Chinese commercials did not occur between 1998 and 2014 in shared and personal product categories but that a change did occur in the representation of cultural values in advertising. In addition, we hypothesized (H11) that, between 1998 and 2014, there would be a greater shift in cultural advertisement appeals toward Western individualistic values than toward GCCP. A shift toward individualism would be characterized by less usage of a group consensus appeal and greater use of individual/independent appeal. Further, this may be identified if Chinese advertising appeals exhibit cultural norms typically seen in U.S. ads, and U.S. advertising stays consistent in their current use of appeals. On the other hand, a shift toward GCCP could be identified through a greater use of the soft-sell appeal and the modernity/youth appeal. In addition, in GCCP, an ad from one country would convey cultural values in a similar way to an ad in another. The fact that two countries’ cultures are moving toward one another and away from their own cultural advertising norms would be representative of the shift toward GCCP. Therefore, in the current study, if the United States demonstrated advertising appeals used in Chinese culture and vice versa, this would exhibit GCCP.
Cross-Cultural Globalization of Advertisements | Allen, Lee, and Escalera

**Method**

**Sampling Method**

For the purpose of the present study, Lin’s (2001) criteria and sampling procedures were replicated for the collection of advertisements. In Lin’s (2001) study, content in the areas of sports, news, entertainment, and cultural programs that aired during primetime were taken into consideration. To draw a comparison between Lin’s (2001) results and current television advertisements, the same U.S. and Chinese television channels were chosen for the present study. Channel One (CCTV1) focuses on news, sports, serials, and entertainment programs; Channel Two (CCTV2) focuses on news, sports, arts and music, entertainment, and children’s programs. Lin chose these two channels because of the similarity to the United States’ audience population and program genres. A third channel chosen by Lin was the Beijing Television Station (BTV), which covers local news, serials, entertainment programs, and sports. BTV provided Lin’s study with a wider variety of program genres, apart from those in CCTV1 and CCTV2. To represent the United States in Lin’s study, the National Broadcasting Company (NBC), Entertainment and Sports Programming Network (ESPN), and Arts and Entertainment (A&E) were chosen.

Our sampling procedure consisted of recording live primetime television (8–11 p.m.) during a random week in June 2014 (Sunday 22–Saturday 28). To further replicate Lin’s (2001) study, six separate days were observed rather than collecting network recordings on the same day of the week. First, the networks were randomly assigned to the days of the week. Randomization was done by writing all the days of the week and the networks on individual slips of paper. The days of the week were put into one container and the networks into another. A researcher then picked a slip of paper from the days of the week container, and a slip of paper from the network container, resulting in a network/weekday schedule. The weekday and network pairings were as follows: Sunday/ESPN, Monday/BTV, Tuesday/NBC, Thursday/A&E, Friday/CCTV1, and Saturday/CCTV2.

A total of 572 advertisements were gathered: 370 U.S. ads and 202 Chinese ads. Duplicate commercials were eliminated from the study, as well as promotional advertisements for the station, the network, or for a program. This was done in order to prevent potentially skewed data. One hundred seventy-two commercials were rejected from the U.S. group and 55 from the Chinese group. This left a total of 198 commercials for the U.S. sample and 147 for the Chinese sample.

**Measurement**

Advertisements were classified into 10 categories: auto, beauty and personal care, clothing, food and drink, household appliances, medicine, travel, services, industrial products, and miscellaneous. These categories were taken from Katz and Lee (1992), which were also used by Lin in 2001.

The cultural values evaluated in the present study and Lin’s (2001) were adopted from Mueller (1987) because they reflected a comparison between Western and Eastern values. Mueller’s conception of cultural values echoed the central difference between Hořestede’s (1984) definition of independent and collectivist cultures. The distinct cultural values adopted were: group consensus, individual/independence, soft-sell appeal, hard-sell appeal, veneration of older adults/tradition, modernity/youth, status, product merit, oneness with nature, and time orientation (Mueller, 1987). Advertisements from the sample were coded in terms of all cultural values on a scale ranging from 1 (weak) to 3 (strong).

As a precaution, advertisements were also coded into two categories: shared or personal products using Han and Shavitt’s (1994) product categories. Personal products were determined through the sum of the following categories: auto, beauty and personal care, clothing, food and drink, and medicine. Shared products were the sum of the following categories: household appliances, travel, services, industrial products, and miscellaneous. Out of 401 advertisements, 54% were classified as personal and 46% as shared for the 1998 sample. In 2014, out of 207 advertisements 57% were classified as personal and 43% as shared. Shared products were defined as those that required more than one individual to make the decision of purchase and usage. Inversely, personal products were those that required one individual to make the decision for usage and purchase (Han & Shavitt, 1994). The coding of personal and shared products was done in the current study as a cautionary measure for the identification of either a cultural shift or a shift in product categories. Therefore, a comparative analysis was conducted between the shared and personal samples of 1998 and 2014. Han and Shavitt (1994) identified that personal products tended to be advertised as being more individualistic, and shared as more collectivist. As such, the product type, instead of a shift in consumer culture, may...
represent a change in cultural values.

Coding Procedures
The samples were collected and coded based on product type and cultural values as described previously. Chinese ads were translated by native Mandarin speakers into English. The two coders were undergraduate students, one Canadian and one Chinese American. The coders were also the primary and secondary investigators; thus, they previously read the background research and were familiar with the background information needed to code the advertisements. They coded 50% of the ads together, 25% of the U.S. and 25% of the Chinese advertisements. Reliability was checked with this sample and coders had an additional training session to discuss and code misunderstood items. The rest of the ads were coded separately. Differences in coding between both individuals resulted in a discussion and were then settled upon. Percent agreement for the U.S. sample was 80% and for the Chinese sample 91%.

Results
Results for the 11 hypotheses are addressed in Tables 1 to 4. Table 1 addresses the current product distribution between U.S. and Chinese commercials. The results in Table 1 reflect that the U.S. 2014 sample had a greater number of ads in the auto, medicine, travel, and services categories compared to the Chinese 2014 sample. On the other hand, the Chinese 2014 sample consisted of a greater number of commercials in the beauty and personal care, food and drink, and miscellaneous categories. The distribution for clothing, household appliances, and industrial products was similar across both the U.S. and Chinese 2014 samples.

Table 2 displays the comparison of product category distribution between 1998 and 2014. Product categories from Table 1 were summed together based on Han and Shavitt (1994) to create the personal and shared categories. This was used as a cautionary measure to examine whether changes in product categories had influenced the changes in advertising appeals. Table 2 exhibits that for both the U.S. and Chinese personal category the results of the analyses were nonsignificant, with moderately low effect size differences. The U.S. shared category results were similar. However, although the China shared category results were nonsignificant as well, the effect size difference was moderately high. These analyses partially supported H10, demonstrating that there were no changes in product categories between 1998 and 2014. In addition, to identify a comparison between 1998 and 2014 and to further answer H10, Tables 3 and 4 were created.

Table 3 displays the cultural values reflected in both the U.S. and Chinese advertisements. Alpha was set at .01, as a precautionary measure for Type 1 error that may occur through the use of multiple

<table>
<thead>
<tr>
<th>TABLE 1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Product Category Distribution in U.S. and Chinese Sample</strong></td>
</tr>
<tr>
<td><strong>1998 Sample</strong></td>
</tr>
<tr>
<td>Product Category</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Auto</td>
</tr>
<tr>
<td>Beauty and personal care</td>
</tr>
<tr>
<td>Clothing</td>
</tr>
<tr>
<td>Food and drink</td>
</tr>
<tr>
<td>Household appliances</td>
</tr>
<tr>
<td>Medicine</td>
</tr>
<tr>
<td>Travel</td>
</tr>
<tr>
<td>Services</td>
</tr>
<tr>
<td>Industrial products</td>
</tr>
<tr>
<td>Misc</td>
</tr>
<tr>
<td><strong>2014 Sample</strong></td>
</tr>
<tr>
<td>Product Category</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Auto</td>
</tr>
<tr>
<td>Beauty and personal care</td>
</tr>
<tr>
<td>Clothing</td>
</tr>
<tr>
<td>Food and drink</td>
</tr>
<tr>
<td>Household appliances</td>
</tr>
<tr>
<td>Medicine</td>
</tr>
<tr>
<td>Travel</td>
</tr>
<tr>
<td>Services</td>
</tr>
<tr>
<td>Industrial products</td>
</tr>
<tr>
<td>Misc</td>
</tr>
</tbody>
</table>

TABLE 2

<table>
<thead>
<tr>
<th>Change in Product Category Distribution Between 1998 and 2014</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Product Category by Country</strong></td>
</tr>
<tr>
<td>U.S. Personal</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>China Personal</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>U.S. Shared</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>China Shared</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

*Note. Personal defined by the sum of (auto, beauty & personal care, clothing, food & drink, and medicine). Shared defined by the sum of (household appliances, travel, services, industrial products, and miscellaneous). Alpha = .01.*

Correspondingly, veneration of older adults and traditional appeal was consistent between 1998 and 2014 in both countries. On the other hand, the results in 2014 were significant in that modernity youth appeal was viewed more in the Chinese sample than the U.S. sample.

Status appeal remained the same throughout 1998 and 2014, not showing any significant difference. Conversely, product merit appeal use reversed between 1998 and 2014. In 1998, the U.S. sample demonstrated a greater use of product merit appeal compared to the Chinese sample (Lin, 2001). In 2014, the Chinese sample showed a strong increase of its use in product merit appeal compared to the U.S. sample.

Both oneness with nature and time-orientation appeals displayed similar means between the 1998 and 2014 studies. Oneness with nature appeal in the 1998 study was more prevalent in China compared to the United States (Lin, 2001). Time-orientation appeal in the 1998 study was more prevalent in the United States compared to China (Lin, 2001). However, in the 2014 study, both of these appeals were similarly used across samples.

Table 4 shows that, although H11 was not supported, there were shifts in U.S. and Chinese cultural values between the 1998 and 2014 studies that indicated a change toward GCCP. Focusing on Table 4, it is possible to identify a new profile in line with the emerging use of GCCP. The 2014 Chinese ads increased in the use of individual/
independence appeal, and the 2014 U.S. sample increased in group/consensus appeal. In addition, the 2014 Chinese sample deviated from the advertising norm in the 1998 sample by increasing the use of the hard-sell appeal product merit appeal, and time orientation appeal. These findings suggested a shift toward increasing globalization of cultural values. The shift can also be seen through the 2014 U.S. sample’s increased use of the veneration of older adults/traditional appeal and oneness with nature appeal, as well as a decreased use of time orientation appeal. Furthermore, GCCP was evident in the 2014 sample through the increased use of modernity/youth appeal in Chinese advertising. It was also displayed through the increased use of a soft-sell approach in 2014 U.S. ads sampled.

**Discussion**

The findings for the present study were surprising. We initially expected a shift toward Western individualistic values. Our initial belief derived from the positive correlation between personal wealth and individualism (Franke et al., 1991; Hofstede, 1997). Focusing on this relationship, a shift toward individualism may still occur. However, the 2008 global economic crash may still be affecting individuals’ opinions regarding their current and expected personal wealth. Similarly, Jiang and Wei (2012) predicted that the 2008 economic crash would be influential on the global perspective of the connectivity of the world market. Therefore, in the future, there may be a shift toward Western individualism. However, although there was not a shift in individualism, the results of the present study demonstrated a shift toward GCCP. We found support for only one out of the nine hypotheses (H5) although Lin (2001) supported seven. The hypotheses focused on the differences between the United States and China, stating that one would be greater or lower in distinct cultural values. Looking at our results for Lin’s hypotheses, it is likely that China and the United States have become

| Table 3: Cultural Values Reflected by Advertising Appeals in U.S. and Chinese Samples |
|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
|                     | 1998 Sample         |                      | 2014 Sample         |                      |
| Country             | n    | M    | SD  | t    | d    | p   | Country             | n    | M    | SD  | t    | d    | p   |
| Group/consensus     | U.S.  | 206  | 1.21| 1.06| -3.77| 0.38| <.001 | Group/consensus     | U.S.  | 198  | 1.92| 0.64| 3.38| 0.64| .001 |
|                     | China | 195  | 1.61| 1.05|     |     |      |                     | China | 147  | 1.51| 0.59|     |     |      |
| Individual/independence | U.S.  | 206  | 1.54| 0.99| 10.99| 1.09| <.001 | Individual/independence | U.S.  | 198  | 1.79| 0.65| 8.42| 1.04| <.001 |
|                     | China | 195  | 0.58| 0.74|     |     |      |                     | China | 147  | 1.20| 0.43|     |     |      |
| Soft-sell           | U.S.  | 206  | 0.74| 0.93| -14.40| 1.44| <.001 | Soft-sell           | U.S.  | 198  | 2.08| 0.72| 6.58| 0.84| <.001 |
|                     | China | 195  | 2.04| 0.87|     |     |      |                     | China | 147  | 1.52| 0.54|     |     |      |
| Hard-sell           | U.S.  | 206  | 1.80| 1.00| 10.14| 1.49| <.001 | Hard-sell           | U.S.  | 198  | 1.97| 0.77| 1.43| 0.18| .018 |
|                     | China | 195  | 0.83| 0.91|     |     |      |                     | China | 147  | 1.83| 0.60|     |     |      |
| Veneration of older adults/traditional | U.S.  | 206  | 0.50| 0.77| -10.93| 1.08| <.001 | Veneration of older adults/traditional | U.S.  | 198  | 1.59| 0.59| 1.24| 0.38| .901 |
|                     | China | 195  | 1.52| 1.09|     |     |      |                     | China | 147  | 1.35| 0.67|     |     |      |
| Modernity/youth     | U.S.  | 206  | 1.13| 1.09| -0.17| 0.02| N.A. | Modernity/youth     | U.S.  | 198  | 1.33| 0.55| -7.56| 0.92| <.001 |
|                     | China | 195  | 1.15| 0.98|     |     |      |                     | China | 147  | 1.80| 0.50|     |     |      |
| Status appeal       | U.S.  | 206  | 1.52| 0.81| 0.89 | 0.85| N.A. | Status appeal       | U.S.  | 198  | 1.41| 0.61| 0.70| 0.19| .143 |
|                     | China | 195  | 1.44| 1.06|     |     |      |                     | China | 147  | 1.25| 0.45|     |     |      |
| Product merit       | U.S.  | 206  | 1.77| 0.83| 2.70 | 0.28| .007 | Product merit       | U.S.  | 198  | 1.94| 0.75| -3.39| 0.40| .001 |
|                     | China | 195  | 1.53| 0.90|     |     |      |                     | China | 147  | 2.22| 0.56|     |     |      |
| Oneness with nature | U.S.  | 206  | 0.84| 0.93| -3.38| 0.34| .001 | Oneness with nature | U.S.  | 198  | 1.68| 0.74| -0.32| 0.10| .549 |
|                     | China | 195  | 1.17| 1.02|     |     |      |                     | China | 147  | 1.76| 0.67|     |     |      |
| Time-oriented        | U.S.  | 206  | 1.69| 1.11| 10.50| 0.80| <.001 | Time-oriented        | U.S.  | 198  | 1.42| 0.66| 0.78| 0.15| .044 |
|                     | China | 195  | 0.69| 0.76|     |     |      |                     | China | 147  | 1.31| 0.47|     |     |      |

Note: N.A. = Not available. Alpha = .01.
The main change that we identified was the use of GCCP. A change was evident through both the U.S. and Chinese sample’s shifts away from Lin’s (2001) original findings of typical Chinese and American appeals. Supporting the shift was the increased use of individual/independence appeal, modernity/youth appeal, hard-sell appeal, product merit appeal, and time orientation appeal in the Chinese advertising sample. The U.S. sample changes also displayed the shift through an increased use of the group/consensus appeal, veneration of older adults/traditional appeal, and oneness of nature appeal, as well as decrease in time orientation appeal. These changes supported the idea that a product is consumed globally, and depicts similarity between attitude, lifestyles, and aspirations worldwide (Alden et al., 1999). This new profile of advertisements is surrounded by specific cultural values.

The current study indicated that the United States may implement individual/independent and group/consensus appeal more than China. Still focusing on the shift between 1998 and 2014 studies, a trend toward globalization can be identified. The United States might have increased in its use of group/consensus appeal, which has previously been associated with collectivistic Eastern cultural values (Lin, 2001; Maynard & Taylor, 1999). Consequently, the Chinese sample also indicated an increase in China’s use of the individual/independence appeal between 1998 and 2014. This finding can be explained through the greater emphasis the Chinese society currently places on modern and constantly changing styles (Zhang & Harwood, 2004). Lin (2001) predicted that a younger millennial generation influenced by a search for a unique fashion could increase individual values. Both of these shifts, which can be identified in Table 4, relate to the growing and assimilation of a global market, with the possibility of representing products universally.

Consistent with previous research, the use

<table>
<thead>
<tr>
<th>Country</th>
<th>n</th>
<th>M</th>
<th>SD</th>
<th>t</th>
<th>d</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group/consensus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U.S. 1998</td>
<td>206</td>
<td>1.21</td>
<td>1.06</td>
<td>8.19</td>
<td>0.81</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>U.S. 2014</td>
<td>198</td>
<td>1.92</td>
<td>0.64</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>China 1998</td>
<td>195</td>
<td>1.61</td>
<td>1.05</td>
<td>-1.04</td>
<td>0.12</td>
<td>.849</td>
</tr>
<tr>
<td>China 2014</td>
<td>147</td>
<td>1.51</td>
<td>0.59</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Individual/independence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U.S. 1998</td>
<td>206</td>
<td>1.54</td>
<td>0.99</td>
<td>2.99</td>
<td>0.21</td>
<td>.002</td>
</tr>
<tr>
<td>U.S. 2014</td>
<td>198</td>
<td>1.79</td>
<td>0.65</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>China 1998</td>
<td>195</td>
<td>0.58</td>
<td>0.74</td>
<td>9.07</td>
<td>1.02</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>China 2014</td>
<td>147</td>
<td>1.20</td>
<td>0.43</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Soft-sell</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U.S. 1998</td>
<td>206</td>
<td>0.74</td>
<td>0.93</td>
<td>16.15</td>
<td>1.61</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>U.S. 2014</td>
<td>198</td>
<td>2.08</td>
<td>0.72</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>China 1998</td>
<td>195</td>
<td>2.04</td>
<td>0.87</td>
<td>-6.38</td>
<td>0.72</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>China 2014</td>
<td>147</td>
<td>1.52</td>
<td>0.54</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hard-sell</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U.S. 1998</td>
<td>206</td>
<td>1.80</td>
<td>1.00</td>
<td>1.91</td>
<td>0.19</td>
<td>.028</td>
</tr>
<tr>
<td>U.S. 2014</td>
<td>198</td>
<td>1.97</td>
<td>0.77</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>China 1998</td>
<td>195</td>
<td>0.83</td>
<td>0.91</td>
<td>11.56</td>
<td>1.30</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>China 2014</td>
<td>147</td>
<td>1.83</td>
<td>0.60</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Veneration of older adults/traditional</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U.S. 1998</td>
<td>206</td>
<td>0.50</td>
<td>0.77</td>
<td>15.93</td>
<td>1.59</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>U.S. 2014</td>
<td>198</td>
<td>1.59</td>
<td>0.59</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>China 1998</td>
<td>195</td>
<td>1.52</td>
<td>1.09</td>
<td>-1.67</td>
<td>0.19</td>
<td>.952</td>
</tr>
<tr>
<td>China 2014</td>
<td>147</td>
<td>1.35</td>
<td>0.67</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Modernity/youth</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U.S. 1998</td>
<td>206</td>
<td>1.13</td>
<td>1.09</td>
<td>2.31</td>
<td>0.23</td>
<td>.110</td>
</tr>
<tr>
<td>U.S. 2014</td>
<td>198</td>
<td>1.33</td>
<td>0.55</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>China 1998</td>
<td>195</td>
<td>1.15</td>
<td>0.98</td>
<td>7.35</td>
<td>0.84</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>China 2014</td>
<td>147</td>
<td>1.80</td>
<td>0.50</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Status appeal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U.S. 1998</td>
<td>206</td>
<td>1.52</td>
<td>0.81</td>
<td>1.54</td>
<td>0.15</td>
<td>.937</td>
</tr>
<tr>
<td>U.S. 2014</td>
<td>198</td>
<td>1.41</td>
<td>0.61</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>China 1998</td>
<td>195</td>
<td>1.44</td>
<td>1.06</td>
<td>-2.04</td>
<td>0.23</td>
<td>.979</td>
</tr>
<tr>
<td>China 2014</td>
<td>147</td>
<td>1.25</td>
<td>0.45</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Product merit</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U.S. 1998</td>
<td>206</td>
<td>1.77</td>
<td>0.83</td>
<td>1.61</td>
<td>0.25</td>
<td>.016</td>
</tr>
<tr>
<td>U.S. 2014</td>
<td>198</td>
<td>1.94</td>
<td>0.75</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>China 1998</td>
<td>195</td>
<td>1.53</td>
<td>0.90</td>
<td>8.18</td>
<td>0.92</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>China 2014</td>
<td>147</td>
<td>2.22</td>
<td>0.56</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oneness with nature</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U.S. 1998</td>
<td>206</td>
<td>0.84</td>
<td>0.93</td>
<td>10.02</td>
<td>0.99</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>U.S. 2014</td>
<td>198</td>
<td>1.68</td>
<td>0.74</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>China 1998</td>
<td>195</td>
<td>1.17</td>
<td>1.02</td>
<td>6.09</td>
<td>0.68</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>China 2014</td>
<td>147</td>
<td>1.76</td>
<td>0.67</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time-oriented</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U.S. 1998</td>
<td>206</td>
<td>1.69</td>
<td>1.11</td>
<td>-2.96</td>
<td>0.30</td>
<td>.002</td>
</tr>
<tr>
<td>U.S. 2014</td>
<td>198</td>
<td>1.42</td>
<td>0.66</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>China 1998</td>
<td>195</td>
<td>0.69</td>
<td>0.76</td>
<td>8.71</td>
<td>0.98</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>China 2014</td>
<td>147</td>
<td>1.31</td>
<td>0.47</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Alpha = .01.
of a tranquil soft-sell approach dominated commercials. Between the 1998 and 2014 samples, the United States increased in its use of the soft-sell appeal. In addition, there was an increase in the Chinese sample’s use of the branding and merit of a forceful hard-sell approach. In the past, these opposing appeals have been found to dominate in one specific culture, with the United States focusing on hard-sell appeal and China on soft-sell appeal (Lin, 2001; Maynard & Taylor, 1999; Mueller, 1987). More current research has demonstrated that a soft-sell approach is more effective and tends to stimulate less irritation than hard-sell appeal (Okazaki et al., 2013). This finding, in relation to our own, demonstrates the current shift toward GCCP in the United States. To adapt to a constantly changing global market and implement the universal view of products, ads need to be subtle, image-based, and indirect. In accordance with Alden et al. (1999), a direct hard-sell approach limits the ability to portray a product’s symbols globally. Therefore, the emotional-and image-based appeal of the soft-sell approach is more conducive toward GCCP.

A third shift that occurred was the movement away from veneration of older adults/traditional appeal and toward modernity/youth appeal in the Chinese samples from 1998 to 2014. Previous research has found that tradition is a key feature in many Asian advertisements (Belk & Pollay, 1985; Lin, 2001). However, consistent with our results, there has been evidence for a growing younger and more status-conscious generation in China (Lin, 2001). These changes in population could account for the increased focus on modernity/youth appeal and the Chinese use of GCCP in the present study. The Chinese focus on this appeal in the present study revealed the shift toward contemporariness and a more individualistic ideal. The emphasis on product benefits to younger generations demonstrated GCCP through the changing market. More evidence for this is China’s increasing use of product merit over the United States in 2014. This can be represented as a shift away from status and toward the purchase of a product for the benefits of the item. These changes can be equated to China’s shift toward global individual values. In addition, although the United States does not show a similar increase in modernity/youth appeal compared to 1998, it does demonstrate a significant increased shift in veneration of older adults between 1998 and 2014. Hence, the changes in both appeals exhibit increased similarity between both societies’ advertisements.

The findings from the present study supported the globalization strategy of advertisements. The changes that have occurred have shifted the power away from manufacturers toward consumers. As technology has allowed consumers to purchase items in a smart and more informed basis, advertising must be used more effectively to capture the purchaser. It is important to remember that these changes do not indicate a change in culture itself, but instead advocate for shifts in advertising and marketing strategies. In conclusion, with the changes that have occurred in both China and the United States, it is important to recognize the shift that both countries have made. For instance in the present study, China is implementing a global view of products in a soft-sell, modern, and product merit style, and the United States is now shifting toward a more tranquil collectivist approach.

With the analysis of these results, it is important to remember the limitations that may inhibit their accuracy. In the present study, we only analyzed one week of data for 2014, instead of developing a longitudinal study, which could develop more consistent results. Another limitation was that we studied each network on a different day of the week and that only three channels were selected per country. This was done to replicate Lin’s work. However, advertisements tend to change daily and between channels with programming. Future studies should focus on developing a more efficient way of gathering data by recording all networks on the same day of the week. Additionally, the small cell size for some product categories and the error that can occur with selective coding schemes should be noted. To further develop the accuracy of these results, replication with a larger sample size is needed.

The present study identified a shift between 1998 and 2014 toward the representation of similar cultural values in accordance with GCCP, for both U.S. and Chinese advertisements. Further research should focus on the development of merging cultures through advertisements and the connection to individual’s personal wealth and spending habits. Advertising trends can change rapidly. Therefore, a future comparative analysis over a longer time frame could develop a better understanding of global media flows and cultural values expressed within commercials.

References
Cross-Cultural Globalization of Advertisements | Allen, Lee, and Escalera


Author Note: Julian B. Allen, Kimberly K. Lee, and Elena Escalera, Department of Psychology, Saint Mary’s College of California.

Correspondence concerning this article should be addressed to Julian B. Allen, Department of Psychology, Saint Mary’s College of California, Moraga, CA, 94575. E-mail: julianba@me.com.
Physical activity has many health benefits including maintaining or improving cognitive functioning in healthy older adults (Barnes, Yaffe, Satariano, & Tager, 2003; Colcombe & Kramer, 2003; Colcombe et al., 2003; Kramer, Bherer, Colcombe, Dong, & Greenough, 2004). Past studies have looked at aerobic exercise interventions in healthy but physically inactive older adults with mixed results. For example, Hill, Storan, and Malley (1993) found little improvement in cognition, but Dustman et al. (1984) found that physical activity improved cognitive function. A meta-analysis by Colcombe and Kramer (2003) found that overall cognitive functioning was positively influenced by physical activity in healthy older adults. The role of physical activity in neurological disorders such as Alzheimer’s disease (AD) is less well-known (Burns, Cronk, et al., 2008). Currently, AD, whether it is mixed with other dementias or is “pure,” is one of the most common forms of dementia (Jellinger & Attems, 2010), and research has suggested that modifying health behaviors could decrease the prevalence worldwide (Barnes & Yaffe, 2011). Because of the expected increase in cases and costs associated with AD, it is important to find ways to prevent, delay, or alleviate its symptoms. Because of the relative affordability and modifiability of physical activity, it is important to explore whether engagement in physical activity can provide cognitive maintenance or gains in AD. With regard to research, examining a self-reported measure of physical activity in an AD-specific sample could illustrate whether the measure is appropriate to use on this sample or whether there are issues like floor effects. If this measure is not appropriate, new physical activity measures could eventually be constructed to capture which activities individuals with AD are likely to engage in.

**ABSTRACT.** Physical activity is believed to improve cognition, particularly executive function and working memory, in older adults. The current study investigated whether self-reported physical activity as measured by the Physical Activity Scale for the Elderly (PASE) and an objective fitness measure of submaximal oxygen volume intake, VO$_2$$_{peak}$, predicted performance on tests of executive function and working memory in older adults with and without Alzheimer’s disease (AD). In a sample of 74 healthy older adults, we found that, of the individual PASE items, the walking-related question was the single best indicator of performance on executive function (verbal fluency-animals $\beta = .30, p = .01$) and working memory (digit span forward $\beta = .31, p = .04$). In the 72 participants with AD, the overall PASE score significantly predicted executive function performance (verbal fluency sum $\beta = .29, p = .02$), but neither the PASE individual items nor the VO$_2$$_{peak}$ significantly predicted the working memory tasks. Future studies should analyze longitudinal data to determine the relationship between physical activity and cognition over time.
Physical Activity and Cognition

Physical Activity and Executive Function

Physical activity’s role on cognition for healthy older adults (i.e., no neurological or physical disorder) is fairly widespread across multiple domains (Colcombe et al., 2003). One domain that receives significant benefit from physical activity is executive function, a “higher order” function that regulates “lower order” processes (Alvarez & Emory, 2006, p. 17). That is, executive function is conceptualized as the cognitive domain responsible for organizing and executing complex thoughts and behaviors such as balancing a checkbook. For a more in-depth review of physical activity and executive function, see Etnier and Chang (2009). A meta-analysis of randomized fitness intervention trials exploring the effect of physical activity on cognition found that the biggest gains from physical activity in healthy older adults are in executive function, although other cognitive domains (e.g., spatial) significantly improved as well (Colcombe & Kramer, 2003). Executive function is closely associated with the frontal cortex (Stuss & Alexander, 2000). In typical aging, the grey matter density associated with executive function declines most severely in the prefrontal, superior parietal, and middle/inferior temporal cortices; coincidently, these areas receive the greatest benefits of aerobic fitness as well (Colcombe et al., 2003, Colcombe et al., 2006). It is possible that physical activity would decrease atrophy in regions associated with executive function, which may lead to improved performance or smaller observed deficits in executive function.

In mild AD, research has shown that executive function deficits exist, albeit not uniformly across various tests (Lafleche & Albert, 1995). Some studies have found improvement in executive function with exercise in older adults with AD (Yu, Kolanowski, Strumpf, & Eslinger, 2006). Clues to understanding the relationship between physical activity and executive function in individuals with Alzheimer’s disease may be found in the neural makeup of AD. In AD, amyloid-β, an abnormal protein, is linked to neuritic plaque formation (Pérez & Cancella Carral, 2008). Some research with mice has shown that, after 5 months of voluntary exercise, there was a decrease in extracellular amyloid-β plaques in the frontal cortex, among other places (Adlard, Perreau, Pop, & Cotman, 2005), likely due to limited ability to measure plaques in living humans. However, as new technologies develop to examine the brains of AD patients in vivo (e.g., Pittsburgh Compound B), future studies will likely address this issue.

Physical Activity and Working Memory

Another cognitive domain affected by physical activity and dementia is working memory (Belleville, Peretz, & Malenfant, 1996), defined as “the online storage and manipulation of information for a short period of time” (Kensinger, Shearer, Locascio, Growdon, & Corkin, 2003, p. 230). Working memory, like executive function, is associated with the prefrontal lobe (Funahashi & Kubota, 1994). Like executive function, it may receive great benefit from aerobic fitness (Colcombe et al., 2003; Colcombe et al., 2006). In Newson and Kemps’ (2006) cross-sectional study of healthy individuals, processing resources including working memory were found to account for more variance between cardiorespiratory fitness and cognitive functioning than other measures.

In dementia (especially AD), the possible relationship between working memory and physical activity may be apolipoprotein E (APOE; Deeny et al., 2008), a plasma protein in the tangles and plaques of people with AD (Lehtovirta et al., 1995). APOE is a genetically inherited risk factor with three possible allele types: ε2, ε3, and ε4 (Rosen, Berges, Putnam, Harwell, & Sunderland, 2002). The APOE allele associated with increased risk of AD is ε4 (Bondi, Salmon, Galasko, Thomas, & Thal, 1999). A study examining the relationship found that ε4 carriers were less able to divide their attention than non-ε4 carriers, suggesting that APOE may affect the central executive component of working memory (Rosen et al., 2002).

Physical activity appears to have benefits for ε4 carriers. A study by Deeny et al. (2008) found that physically active ε4 carriers fared better on various working memory tasks than physically inactive ε4 carriers. Physical activity also appears to have a positive effect for ε4 carriers with other memory tasks; it appears to increase the levels of brain-derived neurotrophic factor in the hippocampus, an important component in memory (Nichol, Deeny, Self, Camaclang, & Cotman, 2009). However, Nichol and colleagues’ finding was discovered in mice and has not yet been generalized to a human population. Because of the emerging support for physical activity’s effect on the brain, it is possible that the brain changes could impact cognition as well, especially cognitive domains like executive function and working memory, brain regions that are impacted by physical activity.
Research Questions and Hypotheses
The present study aimed to examine the relationship between self-reported and objective measures of physical activity with executive function and working memory in both healthy older adults and older adults with early stage AD. Currently, no studies have examined the appropriateness of administering older adult-specific physical activity questionnaires to individuals with AD. The current study examined whether a self-reported physical activity questionnaire and an objective fitness measure predict similar cognitive outcomes for both individuals with and without AD. Given that findings from previous studies support the relationship between physical activity and cognition, we expected to find a significant relationship between physical activity (both self-reported and objective) and executive function and working memory in healthy older adults, specifically that higher physical activity engagement would be associated with better performance on the cognitive tasks. For older adults with AD, we expected to find a significant relationship between objective physical activity and executive function, where higher levels of physical activity engagement would be associated with better performance on executive function tasks.

Method
Sample and Participant Selection
Healthy older adults (Clinical Dementia Rating, CDR, 0; n = 74) and early stage AD participants (CDR 0.5, n = 58; and CDR 1, n = 14; total N = 72) aged 65 and over were enrolled in an ongoing study, the Brain Aging Project, at a large midwestern university with a medical center. There were 139 European American (95%), six African American (4%), and one Native American (1%) participants in the study. The University of Kansas Institutional Review Board provided approval for the Brain Aging Project (Clinical Trials Identifier number: NCT00267124). Participants were recruited from a referral-based memory clinic and by media appeals. Study exclusions included neuropsychologic disease other than AD, diabetes mellitus (defined as a clinical diagnosis and use of an antidiabetic agent), history of ischemic heart disease (acute coronary artery event, angina), schizophrenia, clinically significant depressive symptoms, abnormalities in B12, rapid plasma regain, or thyroid function, use of psychoactive and investigational medications, significant visual or auditory impairment, or systemic illness that would impair completion of the study.

Participants with early stage AD had a collateral source to record self-measures. Collateral sources included family members or caregivers. Collateral sources were secured to ensure that the data were not affected by the cognitive status of the participant. This is standard practice because individuals with dementia are often unreliable reporters (Kiyak, Borson, Teri, & Borson, 1994; Wadley, Harrell, & Marson, 2003; Weinberger et al., 1992).

Procedures
After participants consented to be in the study, data were collected in one appointment per participant and lasted anywhere from 2 to 3 hours. Participants were tested individually. A clinical assessment, described below, was administered at the beginning of the session. Medications, past medical history, education, demographic information, and family history were collected at this time. For individuals with AD, their collateral source, typically a caregiving family member, provided this information to a nurse clinician. Healthy older adults also had a collateral source, but they did not provide health information to the clinician. Blood pressure was collected, and physical and neurologic examinations were also performed. For both the healthy older adults and AD participants, their collateral source was administered the Geriatric Depression Scale (Hamilton, 1967) and the Neuropsychiatric Inventory (Galasko et al., 1997) to assess participants’ depression and presence and severity of neuropsychiatric symptoms, respectively. Functional activity levels were estimated using the Mild Cognitive Impairment Activities of Daily Living Scale (Galasko et al., 1997) using information collected from the collateral source.

After the clinical assessment was completed, a trained psychometrician administered a psychometric battery including measures of memory (Wechsler Memory Scale-Revised Logical Memory I and II, Wechsler & Stone, 1973; Free and Cued Selective Reminding Task, Grober, Buschke, Crystal, Bang, & Dresner, 1988), language (Boston Naming Task- 5 item, Kaplan, Goodglass, & Weintraub, 1983), working memory (Wechsler Memory Scale III Digit Span Forward and Backward, Wechsler & Stone, 1973; Wechsler Adult Intelligence Scale Letter-number sequencing, Wechsler, 1955), executive function (Trail Making A and B, Armitage, 1946; Verbal Fluency animals and vegetables, Hanninen et al., 1994; and Stroop...
Color-Word Test, Stroop, 1935), and visuospatial ability (Wechsler Adult Intelligence Scale Block Design, Wechsler, 1955). As a measure of global cognition, the minimental status examination (Folstein, Folstein, & McHugh, 1975) was administered.

\( \text{VO}_{2\text{peak}} \), an objective measure of cardiorespiratory fitness examining the peak oxygen intake during a treadmill task, was then assessed using the method described below and previously (Burns, Cronk, et al., 2008). The Physical Activity Scale for the Elderly (PASE) was then administered. Additional data including the Physical Performance Test (Shah et al., 2004), a 14-sample intravenous glucose tolerance test, and dual energy x-ray absorptiometry to determine fat-free mass, fat mass, percent body fat, and total body mass, were collected at this time.

**Assessments and Measures**

**Clinical assessment.** The clinical assessment included a semistructured interview with the participant and with a collateral source knowledgeable about the participant. Diagnostic criteria for AD require the gradual onset and progression of impairment in memory and at least one other cognitive and functional domain (McKhann et al., 1984). The presence or absence of dementia, and its severity (if present), was determined using the clinical dementia rating scale (CDR; Hughes, Berg, Danziger, Coben, & Martin, 1982; Morris, 1993). These methods have a diagnostic accuracy for AD of 93% (Berg et al., 1988). On the basis of the collateral source and participant interviews, a global CDR score was derived from individual ratings in each domain such that CDR 0 indicates no dementia and CDR 0.5, 1, 2, and 3 represent very mild, mild, moderate, and severe dementia, respectively.

**Executive function.**

**Trail Making A and B.** Executive function was measured using three different tests. Trail Making A and B measure speed of processing, mental flexibility, scanning, and visual search information (Tombaugh, 2004). Trail Making A consists of “drawing lines sequentially connecting 25 encircled numbers distributed on a sheet of paper” (Tombaugh, 2004, p. 203). This portion of the test examines a person’s sequencing ability (Atkinson et al., 2009). Scores ranged from 16 to 63 for healthy older participants \((M = 33.70, SD = 9.57)\), and scores ranged from 16 to 180 for participants with AD \((M = 67.12, SD = 43.50)\). There were significant group mean differences, indicating that healthy older participants had significantly lower mean scores (i.e., better task performance). In Trail Making B, the person must draw lines alternating between numbers and letters (e.g., 1, A, 2, B, 3, C). This examines the person’s shifting ability (Atkinson et al., 2009). Scores ranged from 37 to 150 for healthy older participants \((M = 82.40, SD = 25.29)\), and scores ranged from 52 to 300 for participants with AD \((M = 147.89, SD = 42.66)\). There were significant group mean differences, indicating that the healthy older participants had significantly lower mean scores (i.e., better task performance). The scores on A and B represent the amount of time it took to complete each task; lower scores indicate better performance.

**Verbal fluency (animals and vegetables).** The second test of executive function, verbal fluency, requires respondents to name as many types of a cue word as possible (i.e., for cue “animal,” as many animals as possible). The score is the number of appropriate responses given in a 60-s time period. The sum of verbal fluency was also calculated by adding an individual’s verbal fluency-animals and verbal fluency-vegetables scores. Verbal fluency-animals scores ranged from 9 to 29 for healthy older participants \((M = 19.34, SD = 4.21)\) and 1 to 25 for participants with AD \((M = 13.62, SD = 5.28)\). Verbal fluency-vegetables scores ranged from 8 to 25 for healthy older participants \((M = 15.66, SD = 3.76)\) and 2 to 22 for participants with AD \((M = 9.86, SD = 4.56)\). Mean performance on both tasks were similar across groups. Total fluency scores ranged from 22 to 51 for healthy older participants \((M = 35.0, SD = 6.61)\) and 5 to 46 for participants with AD \((M = 23.49, SD = 8.82)\). There were significant group mean differences, indicating that healthy participants had significantly better mean performance.

**Stroop Color-Word Test.** The Stroop Color-Word Test requires the respondent to read the name of the ink color, but the actual word is a different color (e.g., the word red is typed in blue ink; the correct response is blue). The score indicates reaction time; smaller numbers indicate better performance. Because the color naming, as opposed to the word reading, is of interest, reaction times of the color naming were examined. The test has been shown to yield highly reliable measures in interference proneness, among other measures (Jensen & Rohwer, Jr., 1966). Past research has suggested that older adults with AD may experience inhibition breakdowns more quickly than healthy older adults.
Scores on color naming ranged from 50 to 97 for healthy older participants \((M = 73.10, SD = 10.71)\), and scores for participants with AD ranged from 13 to 97 \((M = 56.42, SD = 18.31)\). Scores on word reading ranged from 63 to 130 for healthy older participants \((M = 96.13, SD = 12.10)\), and scores for participants with AD ranged from 22 to 109 \((M = 76.11, SD = 18.93)\). There were significant mean differences in both Stroop scores, indicating that, for both, healthy older participants had better mean scores.

**Working memory.**

*Wechsler Memory Scale III Digit Span Forward and Backward.* The digit span tasks present a list of numbers, and participants are asked to repeat the numbers back (either forward or backward). If an individual repeats the sequence correctly, that person is given a longer list of numbers to repeat back (i.e., if an individual gets a list of three numbers and repeats it successfully, a list of four numbers is presented). The test continues until the person repeats a sequence incorrectly. The score is based on the highest number of digits a person is able to correctly recall. Digit span has been found to have high internal consistency and test-rest reliability (Iverson, 2001). Scores on digit span forward ranged from 5 to 14 for healthy older participants \((M = 8.93, SD = 2.23)\), and scores for participants with AD ranged from 4 to 14 \((M = 10.16, SD = 1.83)\). There were no significant mean group differences. Digit span backward scores ranged from 3 to 12 for healthy older participants \((M = 6.84, SD = 2.14)\) and ranged from 2 to 11 for participants with AD \((M = 5.22, SD = 1.77)\). Mean scores were similar across groups.

*Wechsler Adult Intelligence Scale Letter-number sequencing.* Letter-number sequencing is a verbal task that measures working memory. Respondents are presented with a random series of numbers and letters ranging from two to nine letter-number combinations. The person is then instructed to say the numbers first in order from lowest to highest and then the letters in alphabetical order; they can also present letters first and numbers after. Answers are considered correct if the person is able to create the correct sequence of numbers and letters. Their score is composed of the number of correct trials. Scores on letter-number sequencing ranged from 6 to 14 for healthy older participants \((M = 10.16, SD = 1.83)\), and scores ranged from 1 to 14 for participants with AD \((M = 7.45, SD = 2.70)\). There were significant mean group differences, indicating that healthy older participants had significantly better mean task performance.

**Cardiorespiratory fitness.** \(\text{VO}_{2\text{max}}\) is a standard of cardiorespiratory fitness (McAuley et al., 2011). It was measured during a symptom-limited graded treadmill test with a protocol designed for a geriatric population (described in Hollenberg, Ngo, Turner, & Tager, 1998). Participants wore a nonrebreathing facemask and were attached to a 12-lead electrocardiograph to monitor cardiac stability. \(\text{VO}_{2\text{max}}\) is the highest peak of oxygen intake during an incremental, submaximal physical test (Whipp, 2010). Because the testing to determine \(\text{VO}_{2\text{max}}\) is more time consuming and may not yield the true highest attainable oxygen intake, the \(\text{VO}_{2\text{peak}}\) is an appropriate measure to use to examine cardiorespiratory fitness. Washburn, McAuley, Katula, Mihalko, and Boileau (1999) also noted that older adults are often unable to obtain verifiable \(\text{VO}_{2\text{max}}\), making \(\text{VO}_{2\text{peak}}\) a more viable option in this population.

Scores on \(\text{VO}_{2\text{peak}}\) ranged from 12.54 to 44.60 in the healthy older sample \((M = 21.75, SD = 6.04)\), and scores ranged from 11.99 to 28.41 in the sample with AD \((M = 19.46, SD = 3.94)\). There were significant group mean differences, indicating that healthy older participants had higher \(\text{VO}_{2\text{peak}}\) than those with AD.

**PASE.** The level of habitual physical activity was measured using the PASE (Washburn, Smith, Jette, & Janney, 1993). It is a reliable and valid (Schuit, Schouten, Westerterp, & Saris, 1997; Washburn et al., 1999) measure of habitual activity. It is a self-report questionnaire that asks questions about physical activity in three domains: occupational, household, and leisure activities (i.e., walking, yardwork, caring for another person, or housework) over a 7-day period. The questions are a mix of yes or no, Likert scale, and open-ended questions. Scores on the PASE have a possible range from 0 to 361, with 0 being no engagement in physical activity. For healthy older participants, scores ranged from 27 to 276 \((M = 127.11, SD = 57.45)\), and scores for participants with AD ranged from 0 to 267 \((M = 85.20, SD = 55.33)\). There were no group mean differences in overall PASE scores. Validity and reliability for populations with dementia have not been established.

In the current study, the PASE was modified by administering it to the participant’s study partner for participants with AD. Instead of analyzing data by domains (occupational, household, leisure), the total PASE score and answers to specific items...
were analyzed. For the correlation and regression analyses, some individual PASE questions were modified (i.e., sports, housework, lawn, and garden) to combine overlapping questions. Other studies have used similar combinations to increase variability and to reduce the number of zeros (Dallosso et al., 1988). In the current study, participation in sports at light, moderate, and strenuous intensity were combined into a single indicator with a higher score indicating more strenuous sports participation. Due to the high degree of overlap between the lawn and yard work item with the gardening item, they were combined into a single item, called yardwork and gardening.

Results
Sample Characteristics
Sample descriptive statistics for demographic characteristics, cognitive test performance, and physical activity measures are presented in Table 1. The overall sample size was N = 146, and there were 74 healthy older adults and 72 older adults with early stage AD. The mean age for healthy older adults was 74 (SD = 7.20), and the mean age for older adults with AD was 74.90 (SD = 6.40). Of the healthy older adults, 56% of the sample were women, and 62% of the AD sample were men. The mean number of years of education for the healthy sample was 16.40 (SD = 2.80) years and was 15.10 (SD = 3.10) for the sample with AD.

VO\textsubscript{2}\textsuperscript{peak} and Cognition
After controlling for age, sex, and education in a simultaneous multiple regression, VO\textsubscript{2}\textsuperscript{peak} was not found to significantly predict performance on the executive function tests in adults with or without AD. In these models, education in adults without AD predicted performance on verbal fluency sum score (β = .24, p = .01) such that higher education predicted higher verbal fluency performance. Age significantly predicted performance on Trail Making A (β = .32, p = .02), Trail Making B (β = .41, p = .002), and the Stroop task (β = .32, p = .02); older adults performed significantly worse on the three tasks. Sex significantly predicted performance on verbal fluency-vegetables (β = .35, p = .03) such that men performed significantly worse.

In individuals with AD, sex predicted performance on verbal fluency-animals (β = .36, p = .02), indicating that men performed better than women on this task. In both individuals with and without AD, VO\textsubscript{2}\textsuperscript{peak} did not significantly predict performance on working memory tests.

In older adults without AD, education significantly predicted digit span backward (β = .32, p = .01), indicating that individuals with higher education performed significantly better on this task. Age significantly predicted performance on digit span substitution (β = .36, p = .03) and letter number sequencing (β = .37, p = .01) such that older adults performed significantly worse.

In older adults with AD, sex predicted performance on digit span forward (β = .39, p = .01), and education predicted performance on digit span substitution (β = .35, p = .04). Men performed significantly better on digit span forward, and individuals with higher education performed significantly better on digit span substitution.

PASE and Cognition
Results of multivariate analysis demonstrated that, after controlling for age, sex, and education, the PASE item walking predicted verbal fluency-animals (β = .30, p = .01) for individuals without AD; individuals with a higher walking composite score performed better on this task.

In multivariate analysis adjusting for age, sex, and education, the total PASE score predicted the sum of verbal fluency categories (β = .29, p = .01).
physical activity scores predicted better task performance. No covariates significantly predicted performance on the executive function tasks.

In multivariate analysis adjusting for age, sex, and education for older adults without AD, the PASE item walking was found to significantly predict digit span forward ($\beta = .31, p = .04$), and the PASE item caring for another significantly predicted letter number sequencing ($\beta = .25, p = .03$). A higher walking composite score and caring for another are associated with better performance on digit span forward and letter number sequencing, respectively.

In older adults with AD, no PASE item or overall PASE score predicted performance on working memory items. Education significantly predicted performance on the digit symbol substitution task ($\beta = .35, p = .04$) such that higher education predicted better performance.

**Discussion**

The current study found that physical activity, as measured by the PASE and VO$_2$peak, is a poor predictor of cognition in older adults with and without AD. Overall, the PASE was a better predictor of performance on executive function and working memory tasks than VO$_2$peak. There are a few possible explanations for this finding. It could be that activities that older adults report are insufficient to impact their cardiorespiratory fitness but may be enough to see modest impacts on cognition. Although enhanced cardiorespiratory fitness in older adults without AD is associated with preventing cognitive decline, it could be that cardiorespiratory fitness does not have the same cognitive benefits in AD despite the fact that individuals with AD have comparable levels of cardiorespiratory fitness (Burns, Mayo, Anderson, Smith, & Donnelly, 2008). Relatedly, it is possible that the areas of the brain needed for executive function and working memory are already so damaged that cognitive improvement by physical activity engagement is unlikely. For example, it could be that so much neurological damage has occurred that neurogenesis from physical activity cannot replace the missing neurons. It could also be that the PASE captured some activity that the VO$_2$peak was unable to capture. For example, instead of walking benefitting cognition, it could be that the social interaction with a walking partner provides cognitive stimulation. With regard to the individual PASE items, walking significantly predicted performance on cognitive tasks more often than any of the other items. This is likely because walking was one of the activities that participants most reported doing.

With regard to the covariates (age, sex, education), age was the largest predictor of cognitive tasks for participants without AD. The relationship between aging and the cognitive tasks suggested that the older that participants without AD were, the poorer their performance was on executive function and working memory tasks. These results supported previous findings that older age is associated with poorer executive function performance (de Frias, Dixon, & Strauss, 2006) and working memory (Dobbs & Rule, 1989). In participants with AD, education positively predicted performance on all of the working memory tasks, suggesting that more years of education are associated with better working memory. Previous research has found that individuals with early stage AD perform better on global cognition tests if they have higher educational attainment, but these benefits may be lost as the disease progresses in severity (Koepsell et al., 2008).

It is interesting to note that, in the executive function and working memory tests for participants without AD, the individual PASE questions were significant predictors of performance, whereas executive function tasks for individuals with AD were significantly predicted by the total PASE score. A possible reason for significant results obtained by looking at individual items is that some of the activities in the PASE were ones that participants did not do. More people engaged in walking, for example, than home repair. For many of the items, there was a floor effect. That is, many participants indicated that they did not participate in the activity. Because of the number of floor effects, this could suggest that the total PASE may not be an accurate reflection of the types of activities in which older adults engage. It is possible that fewer activities should be included in the PASE. More common activities like walking may be better correlates and predictors of cognitive function. In participants with AD, however, it could be that total PASE scores are more variable than scores on any one item, allowing for better prediction.

The data in the present study are cross-sectional, so we were unable to determine whether these results would be stable across time or how they might change with increasing dementia severity. A second limitation was that the PASE did not seem to measure much activity in this relatively physical inactive sample of older adults. Many of
the PASE items were activities in which participants did not engage. Future research should investigate better ways to measure physical activity in older adults with AD or very physically inactive older adults.

Despite the limitations, the current study suggested that there are effects of physical activity on cognition, and there was support for the relationship between the two that could be further explored in either a longitudinal study or a physical activity intervention study. Future studies should explore longitudinally whether physical activity over time influences performance on executive function and working memory tasks. It is possible that lifetime or long-term physical activity is needed to have lasting effects on the brain or has a stronger effect than short-term activity. Inactivity may be a risk factor for developing AD, caused by AD, or it may be both (Friedland et al., 2001).

The current study suggested that physical activity, especially walking, should be further explored to see whether it is able to longitudinally predict performance on cognition or if the results obtained were specific for this age and time period. Because the study sample was fairly physically inactive, the results might have underestimated the true relationship between physical activity and cognition. Future studies could explore ways to motivate older adults to engage in physical activity so its role in cognition could be more clearly defined. The present study showed that self-reported measures of physical activity did capture some variability in cognition in older adults with and without AD despite the messiness of the measure. Leading a physically active lifestyle may not just help the body but the mind as well.

References


Sprague, Watts, and Burns | Physical Activity and Cognition


Author Note: Briana N. Sprague and Amber S. Watts, Department of Psychology, University of Kansas; Jeffrey M. Burns, Department of Nutrition, University of Kansas Medical Center.

Briana N. Sprague is now at the Department of Human Development and Family Studies, The Pennsylvania State University, University Park, PA.

This research was supported in part by grants from the University of Kansas Ronald E. McNair Post-baccalaureate Achievement Program.

Correspondence concerning this article should be addressed to Briana N. Sprague, Department of Human Development and Family Studies, The Pennsylvania State University, University Park, PA 16802.

E-mail: sprague.briana@gmail.com
Gay, lesbian, bisexual, pansexual, and queer (GLBPQ) individuals have higher rates of negative psychological outcomes, but the reasons for these trends are unclear. Gay-related stress including internalized homophobia, perceived stigma, and overt discrimination may contribute to these heightened rates, but gay identity may buffer these consequences. GLBPQ individuals (N = 1,169) completed online surveys of gay-related stress, protective factors, and outcomes including depression, anxiety, stress, nonsuicidal self-injury, and suicidality. Hierarchical regressions were used to examine predictive relationships between gay-related stress, gay identity, and negative outcomes. Among gay/lesbian participants, overt discrimination predicted all outcomes; gay identity predicted Center for Epidemiologic Studies Depression Scale (CES-D) depression, Depression Anxiety Stress Scale (DASS) depression, and DASS stress; and internalized homophobia predicted Perceived Stress Scale (PSS) stress (p < .001). Among bisexual and pansexual participants, perceived stigma predicted CES-D depression, DASS depression, and PSS stress, and overt discrimination predicted CES-D depression (p < .001). Gay identity predicted reduced negative outcomes among gay and lesbian people, but did not affect bisexual or pansexual people. Gay identity did not interact significantly with overt discrimination, perceived stigma, or internalized homophobia, challenging the hypothesis that it would buffer their effects. These results indicated that the factors may function differently in different populations, highlighting the need for further research on the topic.

Predictors of Psychological Outcomes in Nonheterosexual Individuals

Emily K. Sanders and Holly McCartney Chalk
McDaniel College

ABSTRACT. Gay, lesbian, bisexual, pansexual, and queer (GLBPQ) individuals have higher rates of negative psychological outcomes, but the reasons for these trends are unclear. Gay-related stress including internalized homophobia, perceived stigma, and overt discrimination may contribute to these heightened rates, but gay identity may buffer these consequences. GLBPQ individuals (N = 1,169) completed online surveys of gay-related stress, protective factors, and outcomes including depression, anxiety, stress, nonsuicidal self-injury, and suicidality. Hierarchical regressions were used to examine predictive relationships between gay-related stress, gay identity, and negative outcomes. Among gay/lesbian participants, overt discrimination predicted all outcomes; gay identity predicted Center for Epidemiologic Studies Depression Scale (CES-D) depression, Depression Anxiety Stress Scale (DASS) depression, and DASS stress; and internalized homophobia predicted Perceived Stress Scale (PSS) stress (p < .001). Among bisexual and pansexual participants, perceived stigma predicted CES-D depression, DASS depression, and PSS stress, and overt discrimination predicted CES-D depression (p < .001). Gay identity predicted reduced negative outcomes among gay and lesbian people, but did not affect bisexual or pansexual people. Gay identity did not interact significantly with overt discrimination, perceived stigma, or internalized homophobia, challenging the hypothesis that it would buffer their effects. These results indicated that the factors may function differently in different populations, highlighting the need for further research on the topic.
stress as an outcome in GLBPQ people are rare (Lewis, Derlega, Griffin, & Krowinski, 2003). Life stress results from life events, whereas gay-related stress comes from navigating a society in which GLBPQ people are marginalized and oppressed (Lewis et al., 2003; Meyer, 1995). Both types of stress have been linked to depressive symptoms, highlighting the need for research into how they interact to predict psychological outcomes (Lewis et al., 2003).

GLBPQ individuals have a much greater risk for suicide compared to their heteronormative counterparts, with most studies citing at least a twofold increase (Gilman et al., 2001; King et al., 2008; Russell & Joyner, 2001). Some studies have placed suicide attempt rates of GLBPQ individuals at five or six times greater than their heterosexual peers, which would represent 20 to 40% of GLBPQ individuals attempting suicide (Almeida et al., 2009; Cambre, 2011; Herrell et al., 1999). One nationally representative survey suggested that experiences of victimization more than double the risk of a suicide attempt in adolescents (Russell & Joyner, 2001).

Suicide among GLBPQ youth is so prevalent that organizations and social media campaigns such as The Trevor Project and the It Gets Better Project have emerged specifically to combat it (Jorgensen, 2015; Savage, 2010).

Finally, GLBPQ people have an increased risk of nonsuicidal self-injury (Almeida et al., 2009; House, Van Horn, Coppeans, & Stepleman, 2011; King et al., 2008; Walls, Laser, Nickels, & Wisnecki, 2010). Nonsuicidal self-injury has been less extensively studied than suicidality, but studies have shown that GLBPQ youth report up to twice the rate of nonsuicidal self-injury as their heterosexual peers, and GLBPQ young adults report up to three times the rate (Walls, Laser, et al., 2010). Minority stress and experiences of discrimination significantly predict nonsuicidal self-injury (House et al., 2011; Walls, Laser, et al., 2010). Specifically, participants who reported discrimination, depression, or attempting suicide in the past year were 2.3, 3, and 10 times more likely to cut themselves, respectively (Walls, Laser, et al., 2010).

The purpose of this research was to explore the combination of variables that best predicts distress in GLBPQ persons. Identifying these factors would allow mental health professionals and researchers to design and implement intervention programs to reduce the negative psychological outcomes in this population.

**Gay-Related Stress**

Gay-related stress is a form of minority stress: the concept that members of minority populations experience chronic psychosocial stress as a result of their minority status and subsequent stigmatization (Brooks, 1981, as cited in Meyer, 1995). Meyer (1995) was the first to propose that minority stress may cause negative psychological outcomes in GLBPQ individuals, finding that it could double or triple their risk. He argued that the root cause of minority stress is the incongruence between the needs, wants, and goals of the mainstream and marginalized populations, and identified external stressors such as discrimination and hate crimes, and internal stressors such as perceived stigma and internalized prejudice (Meyer, 1995). Gay-related stress has been shown to contribute to depressive symptoms separately from general life stressors (Lewis et al., 2003).

One type of gay-related stress is perceived stigma. GLBPQ individuals “fear of being rejected or discriminated against because of their minority status” (Fingerhut, Peplau, & Gable, 2010, p. 101). Perceived stigma has been shown to be detrimental to the mental health of GLBPQ individuals, as evidenced by positive correlations with demoralization, guilt, suicidality, and depressive symptoms (Lewis et al., 2003; Meyer, 1995). Among lesbian women, perceived stigma has been linked to increased social constraints (i.e., feeling like individuals cannot discuss their sexual orientation with others), emotional distress, gay-related stress, physical symptoms, intrusive thoughts, and internalized homophobia (Lewis, Derlega, Clarke, & Kuang, 2006).

Internalized homophobia, a component of gay-related stress, refers to the negative feelings that GLBPQ individuals have toward themselves because of their GLBPQ status (Herek, Cogan, Gillis, & Glunt, 1998). It has been linked to psychological symptoms in GLBPQ individuals including increased rates of guilt, suicidality, sexual problems, demoralization, depressive symptoms, and perceived stigma, as well as decreased rates of self-esteem and outness (Herek et al., 1998; Meyer, 1995). In GLBPQ women, internalized homophobia has been linked to general psychological distress and maladaptive coping styles (Szymanski & Henrichs-Beck, 2014). Internalized homophobia may partially explain the elevated rates of distress among GLBPQ individuals.

The internal experiences of distress and perceived stigma are presumably at least partially the...
result of external forces. GLBPQ individuals are more likely to experience overt discrimination, which is associated with psychological distress (Almeida et al., 2009; Mays & Cochran, 2001). Overt discrimination covers a wide range of acts motivated by anti-GLBPQ prejudice like homophobic slurs, losing or not getting a job, or violent hate crimes. Overt discrimination predicts demoralization, guilt, suicidality, nonsuicidal self-injury, and depressive symptoms (Almeida et al., 2009; Huebner, Rebczok, & Kegeles, 2004; Meyer, 1995). GLBPQ people who experience overt discrimination are twice as likely to report suicidal ideation as those who do not report such discrimination (Huebner et al., 2004). Some researchers have suggested that experiences of overt discrimination could explain the elevated rates of psychological distress among GLBPQ people compared to heterosexual people (Birkett, Espelage, & Koenig, 2009; Huebner et al., 2004; Mays & Cochran, 2001). Therefore, overt discrimination is a crucial factor in any investigation of psychological outcomes in GLBPQ individuals.

Protective Factors

Gay identity, the extent to which GLBPQ individuals feel that they belong to the GLBPQ community, has been shown to offer some protection from the negative effects of gay-related stress (Fingerhut et al., 2010). In early research, gay identity was linked to less perceived stigma, greater outness, and more positive self-perceptions (Frable, Wortman, & Joseph, 1997). Fingerhut et al.’s (2010) work expanded upon these direct effects and revealed negative correlations between gay identity and both perceived stigma and depressive symptoms, respectively. Additionally, the positive correlation between perceived stigma and depressive symptoms disappeared in people with high gay identity (Fingerhut et al., 2010). The current work was a crucial replication of Fingerhut et al.’s (2010) analysis of gay identity as a moderator.

The purpose of the present study was to examine the relationships between gay-related stress, protective factors, and psychological distress in GLBPQ individuals, and to replicate the interaction pattern found by Fingerhut et al. (2010). Given that GLBPQ individuals experience depressive symptoms, anxiety, perceived stress, nonsuicidal self-injury, and suicidal ideation and attempts at higher rates than the general population, it seems likely that gay-related stress including perceived stigma, internalized homophobia, and overt discrimination may contribute to these negative outcomes. Furthermore, gay identity seems to be protective against these effects. Therefore, the hypotheses of the present study were: (a) that perceived stigma, internalized homophobia, and overt discrimination would positively predict negative outcomes; (b) that gay identity would predict fewer negative outcomes; (c) that gay identity would interact with perceived stigma and internalized homophobia, respectively, such that neither would predict negative outcomes in people with high gay identity; and (d) that gay identity would not moderate the relationship between overt discrimination and the outcomes, as in previous research.

Method

Participants

Participants were 1,169 people recruited through a snowball sampling technique. The survey was delivered via Qualtrics and was distributed to contacts in the GLBPQ community via e-mail, online groups, and community events relevant to GLBPQ people. McDaniel College students could complete the study to fulfill a research experience requirement. Finally, the study was advertised with a paid Facebook® advertisement. This ad was targeted toward users who liked pages pertaining to topics relevant or related to GLBPQ people. At the time, Facebook ads could only direct to Facebook pages, so users who clicked on the ad were directed to a Facebook page created for this study. This page prominently displayed the link to the survey itself, and also provided resources related to the outcomes and more information about the project. Recruitment materials requested that people share the survey with eligible others regardless of whether they themselves chose to participate.

On the survey, participants answered demographic questions regarding race/ethnicity, highest level of education, income, and religiosity, and indicated their age, biological sex, gender identity, and sexual orientation. Sexual orientation was determined by participants’ selection from a list of options including gay/lesbian/homosexual, bisexual, pansexual, queer/nonheterosexual, or straight/heterosexual; actual or fantasized sexual or romantic behavior was not considered. Participants could also select None of those accurately reflect my sexual orientation, which I describe as: [blank] and fill in their preferred identity. Participants were eliminated from the survey if they were underage, intersex, transgender, heterosexual, or did not select a sex or gender, leaving 940 valid participants (see Table 1).
Most of these (81.8%) were men. Gay and lesbian people made up 72.0% of the sample, followed by bisexual then pansexual people. The mean age was 34.02 (SD = 13.998). Ages ranged from 18 to 83 with a median of 30 and an interquartile range of 23. Participants could select as many racial identities as they felt applied to them; the majority (82.98%) selected White/European American.

Materials
Participants completed nine scales regarding gay identity, overt discrimination, perceived stigma, internalized homophobia, depressive symptoms, anxiety, perceived stress, non-suicidal self-injury, and suicidal ideation. Resources related to all outcomes were available on recruitment materials, a link on the informed consent page, and the debriefing page. See Table 2 for sample items and anchors, and Table 3 for number of items, means, standard deviations, and ranges. All scores were computed as sums; higher scores indicate higher levels of the construct.

Depression, anxiety, and perceived stress. Depression, anxiety, and perceived stress were each measured by the 21-question Depression Anxiety Stress Scale (DASS) developed by Lovibond and Lovibond (1995). The DASS includes three subscales for depression, anxiety, and stress with seven items each. Participants used a 4-point Likert-type scale to rate how much each item applied to them. This measure demonstrated reliability in the present study (α = .93 for depression, .84 for anxiety, .88 for perceived stress, and .95 for the total scale).

Depression was also measured by the Center for Epidemiologic Studies Depression Scale (CES-D; Radloff, 1977). Participants used a 4-point Likert-type scale to rate how well each statement described their mood or behavior during the past week. The measure demonstrated reliability in Radloff’s (1977) study (α = .85) and the current study (α = .94). Construct validity for the CES-D and DASS depression scales were supported by their strong correlation (r = .87).

Perceived stress was also measured by the Perceived Stress Scale (PSS) as adapted by Cohen and Williamson (1988) to include 10 items. Participants marked how often in the last month they had felt or behaved in ways indicative of stress on a 5-point Likert-type scale. This version showed reliability in the present study (α = .86). The PSS and DASS stress scales were correlated at r = .64, indicating that they may assess different components of stress.

Gay identity. Gay identity was assessed using Phinney’s (1992) Multigroup Ethnic Identity Measure as adapted by Fingerhut et al. (2010). This version includes seven items about a participant’s feeling of belonging to the GLBPQ community such as “I feel good about being gay/lesbian” (Fingerhut et al., 2010). For the current study, the term gay/lesbian was changed to nonheterosexual to be inclusive of more identities. Participants rated how much they agreed with these statements on a 5-point Likert-type scale. The measure demonstrated reliability in Fingerhut et al.’s (2010) study (α = .90) and the present study (α = .92).

Internalized homophobia. Internalized homophobia was assessed using the Internalized Homonegativity Scale, originally developed by Martin and Dean (1987) for use with men and adapted for use with women and mixed-sex samples by Herek et al. (1998). For the present study, it was adapted by including gender neutral language and replacing the terms lesbian/bisexual or gay/bisexual with nonheterosexual. Participants rated statements.

<table>
<thead>
<tr>
<th>TABLE 1</th>
<th>Participant Demographics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Race</strong></td>
<td></td>
</tr>
<tr>
<td>White/European American</td>
<td>n = 562</td>
</tr>
<tr>
<td>Black/African American</td>
<td>n = 45</td>
</tr>
<tr>
<td>Asian/Asian American</td>
<td>n = 12</td>
</tr>
<tr>
<td>Hispanic/Latino</td>
<td>n = 70</td>
</tr>
<tr>
<td>Native American</td>
<td>n = 20</td>
</tr>
<tr>
<td>Other</td>
<td>n = 16</td>
</tr>
<tr>
<td><strong>Highest Level of Education (Mode)</strong></td>
<td>Some college</td>
</tr>
<tr>
<td></td>
<td>(40.5%)</td>
</tr>
</tbody>
</table>

Note: Participants could select multiple races.
on a 5-point Likert type-scale. The scale demonstrated reliability in Herek et al.’s (1998) study ($\alpha = .83$ for men and $\alpha = .71$ for women) and in the present study ($\alpha = .88$).

**Nonsuicidal self-injury and suicidal ideation.** The authors wrote the measures for nonsuicidal self-injury and suicidal ideation for the present study. For both suicidal ideation and nonsuicidal self-injury, participants answered four questions: how many times in their life they had attempted it, when they last attempted it, how many times they had seriously considered it, and when they last seriously considered it. For questions of how many times participants had attempted or considered the behavior, they could select 0, 1, 2, or 3+ (coded for scoring as 1, 2, 3, and 4, respectively). For questions of when participants had last attempted or considered the behavior, they could select never, within the last six months, within the last year, within the last five years, or more than five years ago (coded for scoring as 0, 1, 2, 3, and 5, respectively). Nonsuicidal self-injury was defined as deliberately harming a person’s own body without the intent to complete suicide. Responses were coded into numbers as described and summed. Both scales demonstrated acceptable reliability (nonsuicidal self-injury: $\alpha = .88$; suicidality: $\alpha = .81$).

**Overt discrimination.** Overt discrimination was measured using the Schedule of Racist Events (Landrine & Klondoff, 1996) as adapted by Fingerhut et al. (2010). Participants indicated how often they had experienced discrimination on a 5-point Likert-type scale. The measure showed reliability in Fingerhut et al.’s (2010) study ($\alpha = .93$) and the present study ($\alpha = .93$).

**Perceived stigma.** Perceived stigma was assessed using the Homosexual Devaluation and Discrimination Scale, developed by Link (1987) and adapted by Ortiz (2001) and Theuninck (2000) for use with GLBPQ people. This revised version had 11 items to measure GLBPQ people’s perceptions of societal attitudes toward nonheterosexuality. The scale was adapted for use in the present study by changing words such as gay or homosexuality to nonheterosexual and nonheterosexuality and by making the statements gender neutral. Participants rated their agreement with each statement on a 5-point Likert-type scale. The measure demonstrated reliability in Fingerhut et al.’s (2010) study ($\alpha = .85$) and the present study ($\alpha = .90$).

### Procedure

All procedures and materials used in the present study were approved by the institutional review board of McDaniel College. All study materials were presented online via Qualtrics, including informed consent, debriefing, and warning alert for answers indicative of possible suicidal or self-injurious behavior. After completing an informed consent form explaining the risks and benefits of participation, participants answered demographic questions about age, race, education, income, religion, sex, gender identity, and sexual orientation. They then completed the above measures, the order of which was randomized for each participant. After completing all measures, participants viewed a debriefing page with an explanation of the study and resources for psychological distress.

### Results

All outcome variables correlated with each other. Patterns of correlations for gay identity, perceived stigma, internalized homophobia, and overt

<table>
<thead>
<tr>
<th>Construct</th>
<th>Measure</th>
<th>Sample Item</th>
<th>Anchors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depression</td>
<td>Center for Epidemiological Studies Depression Scale</td>
<td>I felt sad.</td>
<td>$0 = \text{rarely or none of the time; less than 1 day; } 3 = \text{most or all of the time; } 5–7 \text{ days}$</td>
</tr>
<tr>
<td>Anxiety</td>
<td>Depression Anxiety and Stress Scale-Depression</td>
<td>I felt that I had nothing to look forward to.</td>
<td>$0 = \text{did not apply to me at all; } 3 = \text{applied to me very much, or most of the time}$</td>
</tr>
<tr>
<td>Stress</td>
<td>Depression Anxiety and Stress Scale-Anxiety</td>
<td>I felt I was close to panic.</td>
<td>$0 = \text{did not apply to me at all; } 3 = \text{applied to me very much, or most of the time}$</td>
</tr>
<tr>
<td>Perceived Stress Scale</td>
<td>In the last month, how often have you felt nervous and “stressed?”</td>
<td></td>
<td>$0 = \text{never; } 4 = \text{very often}$</td>
</tr>
<tr>
<td>Gay Identity</td>
<td>Multigroup Ethnic Identity Measure</td>
<td>I feel good about being nonheterosexual.</td>
<td>$1 = \text{strongly disagree; } 5 = \text{strongly agree}$</td>
</tr>
<tr>
<td>Internalized Homophobia</td>
<td>Internalized Homonegativity Scale</td>
<td>I wish I weren’t nonheterosexual.</td>
<td>$1 = \text{strongly disagree; } 5 = \text{strongly agree}$</td>
</tr>
<tr>
<td>Nonsuicidal Self-Injury</td>
<td>Author Written</td>
<td>See text.</td>
<td>See text.</td>
</tr>
<tr>
<td>Overt Discrimination</td>
<td>Schedule of Racist Events</td>
<td>How many times have you been called a homophobic name like fag or “dyke”?</td>
<td>$1 = \text{never; } 6 = \text{almost all of the time}$</td>
</tr>
<tr>
<td>Perceived Stigma</td>
<td>Homosexual Devaluation and Discrimination Scale</td>
<td>Most people will think less of a person if he or she is nonheterosexual.</td>
<td>$1 = \text{strongly disagree; } 5 = \text{strongly agree}$</td>
</tr>
</tbody>
</table>
...discrimination were consistent with hypotheses (see Table 4). Gay identity, perceived stigma, internalized homophobia, and overt discrimination correlated as expected with most outcome variables (see Table 5). Suicidality and nonsuicidal self-injury were not significantly correlated with gay identity, nor was suicidality significantly correlated with internalized homophobia. Age was correlated with several variables (see Tables 4 and 5). Although some relationships that were significant in past research did not reach significance here, all significant relationships were consistent with existing literature.

Demographic variables were examined using one-way Analyses of Variance (ANOVAs). Because seven ANOVAs were run per group, we divided the customary alpha of .05 by 7, for an adjusted alpha level of .007. There were no significant differences based on race/ethnicity and only one significant difference for religiosity. Religious participants were more likely to report internalized homophobia ($M = 17.03$, $SD = 8.16$) than nonreligious participants ($M = 14.44$, $SD = 6.90$), $F(1, 687) = 19.90$, $p < .001$, $\eta^2 = .03$. Because religiosity predicted only one outcome, it was not included as a covariate.

There were significant differences found based on gender identity (see Table 6), age, and highest level of education, so these variables were included as covariates in the regression analyses. Participants who had completed some college ($M = 19.94$, $SD = 13.60$) had higher CES-D depression scores than participants who had completed a postcollege degree ($M = 14.14$, $SD = 12.29$), $F(5, 617) = 3.42$, $p = .007$, $\eta^2 = .03$. Biological sex was not assessed as a predictor because any participant whose biological sex did not match their gender identity was excluded from data analysis; assessing both factors would be redundant.

There were significant differences on the basis of sexual orientation in both predictors and outcomes, so regression analyses were performed separately for gay/lesbian people and bisexual/pansexual people. Too few people wrote in their own identity or identified as queer/nonheterosexual to conduct analyses on these groups with meaningful statistical power. There was also too much variability in what these identities might mean to the individuals espousing them. Some people identify as queer because they prefer it over traditional labels, but others use it as a political marker of solidarity with the GLBPQ community, even if they themselves only engage in heterosexual relationships. The fill-in option had varied responses that could not reasonably be grouped together. Because the queer/nonheterosexual and write-in groups were small and heterogeneous, they were excluded from further analysis.

Hierarchical regressions were conducted to predict each outcome variable separately and conducted separately for gay/lesbian people and bisexual/pansexual people. Because participants could not select gay or lesbian as separate identities (i.e., the option in the survey was gay/lesbian/homosexual), these identities could not be separated further. Bisexual and pansexual people were combined because both identities involve attraction to multiple genders, so the groups may experience similar marginalization from heterosexual people and the gay and lesbian community (Mohr & Rochlen, 1999). Covariates were entered first including gender identity, age, and highest level of education. The next step included perceived stigma, gay identity, overt discrimination, and internalized homophobia. The third step included the interaction between gay identity and perceived stigma, and the interaction between gay identity and overt discrimination. The final step included the interaction between gay identity and internalized homophobia.

Among gay/lesbian participants, overt discrimination was a statistically significant predictor for all outcomes, and gay identity and internalized homophobia...
homophobia each predicted a few. Gay identity and overt discrimination significantly predicted DASS depression, $F(7, 396) = 20.22, p < .001$, $R^2 = .26$, Cohen's $f^2 = .35$, and DASS stress, $F(7, 389) = 18.30, p < .001$, $R^2 = .25$, Cohen's $f^2 = .33$. Gay identity, overt discrimination, and education significantly predicted CES-D depression, $F(7, 372) = 21.63, p < .001$, $R^2 = .29$, Cohen's $f^2 = .41$. Internalized homophobia and overt discrimination significantly predicted PSS stress, $F(7, 243) = 7.25, p < .001$, $R^2 = .17$, Cohen's $f^2 = .20$. Overt discrimination was the only significant predictor for DASS anxiety scores, $F(7, 396) = 21.51, p < .001$, $R^2 = .28$, Cohen's $f^2 = .39$, and suicidality, $F(7, 403) = 12.02, p < .001$, $R^2 = .17$, Cohen's $f^2 = .20$. Overt discrimination, age, gender, and education significantly predicted nonsuicidal self-injury, $F(7, 407) = 14.11, p < .001$, $R^2 = .20$, Cohen's $f^2 = .25$.

The regression analyses for bisexual/pansexual participants revealed fewer and different predictors of distress. Overt discrimination and perceived stigma significantly predicted CES-D depression, $F(7, 93) = 8.16, p < .001$, $R^2 = .38$, Cohen's $f^2 = .61$. Perceived stigma significantly predicted DASS depression, $F(7, 102) = 8.84, p < .001$, $R^2 = .38$, Cohen's $f^2 = .61$, and PSS stress, $F(7, 70) = 6.58, p < .001$, $R^2 = .40$, Cohen's $f^2 = .67$. No variables significantly predicted PSS stress, DASS anxious, nonsuicidal self-injury, or suicidality. The beta weights for all predictors for gay/lesbian and bisexual/pansexual participants are presented in Table 7.

Discussion

Correlations between study variables largely supported the hypotheses. All outcome variables were significantly and positively correlated with each other, suggesting that each outcome scale was a valid measure of distress. Overt discrimination and perceived stigma were positively correlated with all measures of distress, consistent with previous research and current hypotheses (Fingerhut et al., 2010). Internalized homophobia was positively correlated with most measures, supporting the hypotheses, but was not significantly associated with suicidality, potentially due to the adapted internalized homophobia scale or self-designed suicidality scale. The pattern of correlations was largely consistent with previous research (Fingerhut et al., 2010) and supported a large body of research indicating that perceived stigma, internalized homophobia, and especially overt discrimination each contribute to negative outcomes, but that gay identity may buffer against these effects.

Predictors for Gay and Lesbian People

Among gay and lesbian participants, overt discrimination was a significant predictor of all outcome variables and the only significant predictor for anxiety and suicidality. This trend was consistent with research indicating that overt discrimination strongly and positively predicts distress (Almeida et al., 2009; Fingerhut et al., 2010; Huebner et al., 2004; Mays & Cochran, 2001; Meyer, 1995; Russell & Joyner, 2001). Fingerhut et al. (2010) found that, unlike perceived stigma, overt discrimination’s effects were not attenuated by gay identity, and this lack of moderation was replicated in the current study. The variety of negative outcomes with which overt discrimination was associated, and that gay identity did not protect against it, indicated that overt discrimination is a strong and robust predictor of psychological distress.

Gay identity predicted lower depression and stress, which replicated the findings of Fingerhut et al. (2010) and supported the hypothesis that gay identity would buffer against multiple psychological outcomes. This was consistent with previous research showing gay identity to be a protective factor (Fingerhut et al., 2010; Frable et al., 1997). However, gay identity’s lack of predictive power for other outcome variables failed to support the hypothesis, indicating that these outcomes may function differently than perceived stress or depression, or that gay identity cannot be applied so broadly to outcomes of gay-related stress. It is surprising that gay identity was not associated with PSS stress but did predict DASS stress, because these constructs should be similar. However, their modest correlation ($r = .64$) suggests that they may be assessing different components of stress. Future studies should continue to employ multiple measures of stress to clarify the potentially different outcomes among different types of stress. The

<p>| TABLE 4 |
| Correlations Between Predictor Variables |</p>
<table>
<thead>
<tr>
<th>Predictor</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Gay Identity</td>
<td>1</td>
<td>-.19</td>
<td>-.53</td>
<td>.09</td>
<td>.13</td>
</tr>
<tr>
<td>2. Perceived Stigma</td>
<td>1</td>
<td>.35</td>
<td>.42</td>
<td>-.04</td>
<td></td>
</tr>
<tr>
<td>3. Internalized Homophobia</td>
<td>1</td>
<td>.10</td>
<td>-.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Overt Discrimination</td>
<td>1</td>
<td>-.02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: $p < .05$. **$p < .01$.**
The current study replicated Fingerhut et al.'s (2010) finding that gay identity does not moderate the effects of overt discrimination, but could not replicate the finding that gay identity moderated the effects of perceived stigma. In the current study, gay identity also did not moderate the effects of internalized homophobia.

Internalized homophobia significantly predicted PSS stress, but no other outcomes. This finding supported the hypothesis that internalized homophobia would predict negative outcomes, but it was surprising that only one outcome was affected in the current study. Internalized homophobia has been linked to many outcomes including depressive symptoms and suicidality in previous research (Herik et al., 1998; Meyer, 1995; Szymanski & Henrichs-Beck, 2014).

Level of education negatively predicted CES-D depression scores and nonsuicidal self-injury, but this may be explained by third variables. People who are racial minorities or members of a lower socioeconomic status exhibit more depressive symptoms, which are linked to nonsuicidal self-injury (Bakken & Gunter, 2012; Walsh, Levine, & Levav, 2012). Because higher education may be less accessible to racial/ethnic minorities and members of lower socioeconomic status, the education measure might have assessed these variables by proxy.

Gender identity and age predicted nonsuicidal self-injury. Participating women had significantly higher levels than those who identified as men, consistent with the established trend that women experience psychological distress and nonsuicidal self-injury at higher rates (Bakken & Gunter, 2012; Norman, 2004; Russell & Joyner, 2001). Younger participants also had significantly higher levels of nonsuicidal self-injury, consistent with the previous research (Walls, Laser, et al., 2010). This trend may be because most nonsuicidal self-injury appears in adolescence (Hawton, Saunders, & O'Connor, 2012). As those who self-injure learn other coping mechanisms, the incidence likely tapers off, creating the negative correlation with age.

Predictors for Bisexual and Pansexual People
Most outcomes were not significantly predicted by any input variables among bisexual and pansexual participants. However, perceived stigma significantly and positively predicted PSS stress, DASS depression, and CES-D depression, and overt discrimination predicted CES-D depression. This is consistent with the hypotheses and previous research linking both perceived stigma and overt discrimination to negative outcomes (Lewis et al., 2003; Lewis et al., 2006; Meyer, 1995).

Comparisons Between Groups
Significant predictors for bisexual and pansexual people were different from those for gay and lesbian people. Among bisexual and pansexual people, perceived stigma predicted depression and PSS stress, and overt discrimination predicted CES-D depression. Among gay and lesbian people, overt discrimination predicted all outcomes, and gay identity predicted depression and DASS stress. Some demographic variables reached significance as predictors in the model of gay and lesbian people, but none did in the model of bisexual and pansexual people.

Most interestingly, overt discrimination predicted all types of distress among gay and lesbian people. However, overt discrimination only predicted CES-D depression among bisexual and

<p>| TABLE 5 | Correlations Between Predictors and Outcomes |</p>
<table>
<thead>
<tr>
<th>Measure</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>CES-D</td>
<td>-.25*</td>
<td>.32**</td>
<td>.30**</td>
<td>.37**</td>
<td>-.13*</td>
</tr>
<tr>
<td>DASS Depression</td>
<td>-.26*</td>
<td>.32**</td>
<td>.27**</td>
<td>.34**</td>
<td>-.12*</td>
</tr>
<tr>
<td>DASS Anxiety</td>
<td>-.10</td>
<td>.24</td>
<td>.18**</td>
<td>.42**</td>
<td>-.14**</td>
</tr>
<tr>
<td>DASS Stress</td>
<td>-.16</td>
<td>.25**</td>
<td>.23**</td>
<td>.35**</td>
<td>-.17**</td>
</tr>
<tr>
<td>PSS Stress</td>
<td>-.20**</td>
<td>.21**</td>
<td>.25**</td>
<td>.25**</td>
<td>-.17**</td>
</tr>
<tr>
<td>NSSI</td>
<td>-.07</td>
<td>.11</td>
<td>.08</td>
<td>.26</td>
<td>-.22</td>
</tr>
<tr>
<td>Suicide</td>
<td>.02</td>
<td>.16</td>
<td>.05</td>
<td>.39</td>
<td>.00</td>
</tr>
</tbody>
</table>

Note. 1 = Gay Identity; 2 = Perceived Stigma; 3 = Internalized Homophobia; 4 = Overt Discrimination; CES-D = Center for Epidemiological Studies Depression Scale; DASS = Depression Anxiety and Stress Scale; NSSI = nonsuicidal self-injury.

<p>| TABLE 6 | Mean Differences by Gender Identity |</p>
<table>
<thead>
<tr>
<th>Measure</th>
<th>F</th>
<th>Male M</th>
<th>Female M</th>
</tr>
</thead>
<tbody>
<tr>
<td>CES-D</td>
<td>8.15*</td>
<td>17.60</td>
<td>21.45</td>
</tr>
<tr>
<td>DASS Depression</td>
<td>7.19</td>
<td>24.93</td>
<td>27.94</td>
</tr>
<tr>
<td>DASS Anxiety</td>
<td>11.21*</td>
<td>21.48</td>
<td>24.37</td>
</tr>
<tr>
<td>DASS Stress</td>
<td>11.95*</td>
<td>27.18</td>
<td>30.72</td>
</tr>
<tr>
<td>PSS stress</td>
<td>11.04*</td>
<td>22.31</td>
<td>25.10</td>
</tr>
<tr>
<td>NSSI</td>
<td>61.57*</td>
<td>6.62</td>
<td>9.95</td>
</tr>
<tr>
<td>Suicide</td>
<td>5.89</td>
<td>7.98</td>
<td>9.00</td>
</tr>
</tbody>
</table>

Note. Because of the number of variables assessed, an adjusted alpha score of .007 was used to prevent type 1 error. CES-D = Center for Epidemiological Studies Depression Scale; DASS = Depression Anxiety and Stress Scale; NSSI = nonsuicidal self-injury.

*p < .05. **p < .01.
pansexual people. These results indicated that overt discrimination was a more robust predictor of distress among gay and lesbian people than among bisexual and pansexual people. One possible explanation is that gay and lesbian people are more likely to disclose their sexual orientation to others, or come out, than are bisexual people (Pew Research Center, 2013). A Pew Research poll showed that 77% of gay men and 71% of lesbians were out to “all or most of the important people in their life,” but only 28% of bisexual people (2013, p. 5). Researchers have theorized that not coming out protects GLBPQ people from overt discrimination; and therefore, may be linked to more varied experiences of discrimination for bisexual and pansexual people (Poon & Saewyc, 2009).

Whereas overt discrimination was the most effective predictor for distress among gay and lesbian people, perceived stigma was the most effective predictor for bisexual and pansexual people. This may be because bisexual people experience stigma from both heterosexual people and gay and lesbian people, some of whom criticize bisexual people because they can pass for straight or question bisexuality’s existence (Mohr & Rochlen, 1999). This form of gay-related stress is unique to bisexual/pansexual people and may explain the group differences in results.

The main effects in the current study were largely consistent with existing literature, but no interactions reached significance, inconsistent with previous research. This may be due to a difference in participants’ locations. No data were collected in the current study about location but it is likely that many participants were from Maryland because the study used a snowball sample originating there. Fingerhut and colleagues (2010) sampled highly from California and Arizona, which both have unusually high rates of same-sex couples (Baumle, 2010). Living among many GLBPQ individuals may affect participants’ experiences. Future studies should recruit evenly across regions and gather location data.

In several scales including the ones from Fingerhut et al. (2010), the phrase gay/lesbian was changed to nonheterosexual in the current study to include a wider range of identities. This adaption may have affected the results. People might have found it easier to identify as an identity (gay/lesbian) than as the opposite of an identity (nonheterosexual).

### Limitations and Future Directions

Several limitations should be noted. The study’s generalizability is limited because participants were recruited nonrandomly, resulting in a sample that was likely not representative of the population. Because the survey was online, originated at a private college, and spread by word of mouth to other colleges, people of lower socioeconomic statuses, education levels, and associated demographic characteristics (e.g., people of color) were likely underrepresented. Similarly, women were underrepresented in the sample, which limited statistical power for gender comparisons. Future researchers should recruit equally across demographic groups and assess whether findings hold across groups.

Given the small size of the recruitment pool at McDaniel College, the majority of the sample was likely recruited through the Facebook ad. Advertising targeted by Facebook likes, comments, and shares might have drawn in a disproportionate number of people who were comfortable with their sexual orientation, whereas targeting by online search terms might have drawn in people who were ashamed of their sexual orientation. Recruiting one
or both of these extremes might have caused distortions in the results. If the Facebook ad targeted different people than Fingerhut et al.’s (2010) study, this would explain some of the difference in the results.

The present study did not assess outness. Coming out can be stressful and harm relationships, or it can help people find support and comfort in their identity (Iwaski & Ristock, 2007). Outness also varies by sexual identity; significantly more gay and lesbian people are out than bisexual and pansexual people (Pew Research Center, 2013). Coming out can make someone a target for discrimination, prompting some people not to come out (Poon & Saewyc, 2009). Because of the repercussions of coming out, future research should assess participants’ level of outness, reasons for choosing that level of outness, and feelings about their outness.

Finally, given the different outcomes in gay/lesbian and bisexual/pansexual people, these groups should be analyzed separately. The current study combined bisexual and pansexual participants due to small samples, but it is possible that these groups are also distinct from each other. Research is needed including pansexual individuals because most extant literature only addresses bisexual people. Similarly, queer/nonheterosexual and write-in identities were not included in analyses due to small samples, so future research should make efforts to recruit enough people with unique identities to have statistical power. Lastly, further research is needed to establish the validity of using the DASS, CES-D, and PSS with GLBPQ people because many predictor variables predicted stress or anxiety on one scale but not the other.

Implications

Despite these limitations, these results have important implications for clinicians. Overt discrimination is significantly associated with anxiety, depression, stress, nonsuicidal self-injury, and suicidal ideation among gay and lesbian people, and the relationships are not buffered by gay identity. Clinicians should assess the effects of overt discrimination in gay and lesbian people who seek therapy. Perceived stigma predicted stress and depression among bisexual and pansexual people, and also was not buffered by gay identity, so this factor may be helpful to explore with bisexual and pansexual clients presenting with mood disorders.

Policy makers should also be aware of the risks posed by overt discrimination to GLBPQ individuals. Enforcing strong policies against overt discrimination in schools and workplaces may curtail the impact. For example, students in schools with gay-straight alliances experience less anti-GlBPQ bullying and violence, suggesting that such organizations may discourage overt discrimination (Goodenow, Szalacha, & Westheimer, 2006; Heck, Flentje, & Cochran, 2011; Walls, Kane, & Wisnecki, 2010). Although similar groups may be difficult to establish outside of schools, they could play an important role in reducing overt discrimination.

The present study contributed to a growing body of evidence indicating that perceived stigma, internalized homophobia, and especially overt discrimination predict psychological distress in persons identifying as GLBPQ. It also explored differences in the experiences of gay/lesbian people and bisexual/pansexual people. Overt discrimination emerged as a stronger predictor of distress among gay and lesbian people, and bisexual and pansexual people seemed more affected by perceived stigma. More research is needed to fully understand these relationships, but with improved efforts to combat gay-related stress, it is possible to reduce its powerful deleterious effects on GLBPQ individuals.

References


Author Note. Emily K. Sanders, Department of Psychology, McDaniel College; Holly McCartney Chalk, Department of Psychology, McDaniel College.

Emily K. Sanders is now at Department of Psychology, University of Illinois at Urbana-Champaign.

This work was supported with several McDaniel College Research and Creativity Grants.

Correspondence concerning this article should be addressed to Holly McCartney Chalk, Department of Psychology, McDaniel College, 2 College Hill, Westminster, MD 21157. E-mail: hchalk@mdcmail.edu

SUMMER 2016

PSI CHI JOURNAL OF PSYCHOLOGICAL RESEARCH

110 COPYRIGHT 2016 BY PSI CHI, THE INTERNATIONAL HONOR SOCIETY IN PSYCHOLOGY (VOL. 21, NO. 2/ISSN 2164-8204)
The Relationship Between Corumination and Health in Young Women: Evidence from Facebook Communications

CaSandra Swearingen, Jennifer Byrd-Craven*, and Shelia M. Kennison*
Oklahoma State University

ABSTRACT. Prior research has shown that face-to-face communications that involve excessive focus on problems and negative aspects of situations are related to higher stress. This type of communication has been called corumination (Rose, 2002). We hypothesized that corumination can also occur in communications carried out on Facebook®. We analyzed 100 female college students’ communications carried out through Facebook and examined their self-reported corumination behaviors from daily life. The results supported the hypothesis that individuals coruminate when communicating through Facebook conversations similar to the way that individuals communicate face to face. The results showed that corumination in daily life had a significant negative association with health in an analysis that took into account the positive effects of social support, $F(3, 89) = 3.81$, $p = .02$. Implications for interventions designed to reduce stress and to improve overall health of Facebook users are discussed.

Reliance on technology for communication has increased dramatically over the last three decades (Tapscott, 2009). Social networking websites such as Facebook® provide users with new ways to communicate. Recent statistics estimated that more than 500 million people use Facebook daily (Facebook, 2012). As new ways of communicating develop, there is a need to understand the effects that these new forms of communication may have on the health and well-being of individuals. Prior research has shown that some forms of interpersonal communication may be adversely associated with health and well-being (Byrd-Craven, Geary, Rose, & Ponzi, 2008; Byrd-Craven, Granger, & Auer, 2011; Calmes & Roberts, 2008; Rose, 2002; Rose, Carlson, & Waller, 2007). Rose (2002) identified a type of communication that focuses on negative topics, in which individuals excessively rehash problems, as corumination. Higher levels of corumination have been positively related to symptoms of depression and anxiety (Rose, 2002; Rose et al., 2007; see also Calmes & Roberts, 2008). Recent work has also shown that higher levels of corumination are related to higher levels of cortisol and alpha amylase, which are two biological indicators of stress (Byrd-Craven et al., 2008; Byrd-Craven et al., 2011). The purpose of the present research was to investigate the novelty of the extent to which individuals using Facebook engage in corumination and whether corumination carried out on Facebook is associated with health.

Researchers have found that corumination occurs more often in the friendships of young girls than the friendships of young boys (Rose, 2002; Rose et al., 2007). This research has shown that corumination is positively related to beneficial aspects including relationship satisfaction and to detrimental aspects including depression. Young women and girls who report higher levels of corumination in their friendships feel closer to their friends than do those who report lower levels of corumination (Rose, 2002). Friends who engage in higher levels of corumination also rate their
friendships as higher quality than those who engage in lower levels of corumination (Rose et al., 2007). However, number of friends has been negatively associated with corumination in which individuals who coruminate more often have fewer friends. (Tompkins, Hockett, Abraibesh, & Witt, 2011). Research on corumination has also shown that the amount of corumination occurring in friendships was positively related to depression and anxiety (Rose et al., 2007).

Recent work (Byrd-Craven et al., 2008; Byrd-Craven et al., 2010) has shown that corumination has biological consequences, which are likely related to negative health outcomes. In a study of young college women, Byrd-Craven et al. (2008) found that higher levels of corumination occurring between friends were related to higher levels of cortisol, a stress hormone. In a more recent study, Byrd-Craven et al. (2011) found that higher levels of corumination were related to higher levels of alpha-amylose, an enzyme associated with the stress response, as well as higher levels of cortisol. The authors pointed out that the biological effects of corumination are complex, involving both the hypothalamic-pituitary-adrenal axis (cortisol) and the sympathetic nervous system (alpha-amylose). Both cortisol and alpha-amylose have been associated with various times throughout a stressful situation (Engert et al., 2011). Stress has been shown to have negative effects on the immune system (for a review, see Denson, Spanovic, & Miller, 2009). Stress is also believed to be a major factor in many physical illnesses such as cancer, asthma, and gastrointestinal conditions or stomach problems (Goldstein & Kopin, 2007).

Social networks have been associated with beneficial aspects to relationships including increasing social capital through maintenance of close and distant relationships (Steinfield, Ellison, & Lampe, 2008; Subrahmanyam, Reich, Waechter, & Espinoza, 2008). Raacke and Bonds-Raacke (2008) found that social networks such as Facebook are used to locate old friends and form new friendships. Users of Facebook are able to communicate with one another in several ways such as communicating instantly or asynchronously. Users may also play games (e.g., Farmville®, Mafia Wars®), which frequently involve other users. Many of these games often allow users to exchange messages in real time (Pempek, Yermolayeva, & Calvert, 2009). Sheldon (2008) investigated individual differences in willingness to communicate and found that users of Facebook communicate via Facebook the way they do in face-to-face communications.

In the present study, we aimed to determine the extent to which college students engage in corumination when using Facebook and the extent that corumination through Facebook relates to health. Adult women are more likely than adult men to endorse Facebook relationship status as containing more meaning such as exclusive relationships and long-term stability (Fox & Warber, 2013) and report having more online relationships (Nice & Katzev, 1988). Women are also more likely than men to exhibit higher levels of internalizing problems (Tompkins et al., 2011). Although Facebook is used more often for social interaction, Facebook has been associated with identity expression in emerging adulthood (Pempek et al., 2009). Older adolescents have also been shown to use the Internet to communicate with existing personal networks more than younger adolescents who are more likely to establish relationships with strangers (Valkenburg, Schouten, & Peter, 2005). Because social networks such as Facebook have been shown to relate to more young adult women than men and young adolescents, the focus of the research was on a college female population. By analyzing Facebook conversations between same-sex close friendship dyads for corumination and assessing health outcomes, the relationship between corumination, health, and social networks can be studied. Additionally, much of the research on corumination has found that women coruminate more than men, particularly with same-sex friends, in both adolescent (Rose, 2002; Rose et al., 2007) and emerging adult populations (Byrd-Craven et al., 2011). Because this was our primary construct of interest, we decided to limit the sample to young women. We hypothesized that individuals would coruminate in their Facebook communications in a similar way that individuals communicate face-to-face. Furthermore, we hypothesized that corumination both on Facebook and self-reported would be negatively related to overall health. Specifically, we hypothesized that higher levels of corumination would be associated with lower health scores as has been documented in the literature for face-to-face relationships. In addition, we hypothesized that social support would be positively related to health.

Method

Participants
Participants were recruited using an online system (i.e., SONA) that allows researchers to post descriptions of research studies and allows
participants to sign up for appointments. Only women were included in the study. Participants were given either course credit or extra credit in a course as compensation for participating in the study. Participants were 100 female undergraduates enrolled in psychology and speech communication courses at a large public university. All participants were 18 years of age or older. All 100 participants submitted copies of their Facebook interactions with a close friend of the same sex. The 100 participants included 11 dyads (i.e., 22 participants in which both friends participated in the study and contributed 11 Facebook conversations). No further demographic information was gathered from participants.

Materials
All participants completed the corumination questionnaire (Rose, 2002), the Research and Development (RAND) Health Survey (Version 1.0, Ware & Sherbourne, 1992), the Positive and Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988), and the Medical Outcomes Study (MOS) Social Support Survey (Sherbourne & Stewart, 1991). A summary of the descriptives of the measures that pertain to the study are displayed in Table 1.

Corumination. The corumination questionnaire (Rose, 2002) was administered to assess participants’ levels of corumination in their friendships. The questionnaire contains 27 statements that relate to the type of conversation that participants have when they talk to another person about their problems and contains statements such as “After my friend tells me about a problem, I always try to get my friend to talk more about it later.” Participants were asked how well the statement described them on a 5-point Likert-type scale of 1 (not at all true) to 5 (really true). The scale ranges from 27 to 135 in which higher scores on the corumination questionnaire indicate that participants engage in higher amounts of corumination. The corumination questionnaire has been found to have high internal reliability (α = .96, Rose, 2002). In the current study, we observed its reliability to be α = .98.

Health. The RAND 36-Item Health Survey (Version 1.0, Ware & Sherbourne, 1992) was used to establish a general overview of participants’ health conditions. The RAND asked questions that covered participants’ personal physical and emotional health. Some questions were reversed scored to account for both positive and negative stated health questions. The RAND includes two questions that measure global self-rated health and how a person’s health has changed over the past 12 months. The RAND also includes eight other subscales: Energy Level, Pain, Physical Functioning, Role Limitations resulting from Physical Problems (physical limitations), Role Limitations resulting from Emotional Problems (emotional limitations), Social Functioning, General Health Perceptions, and General Mental Health. A total RAND score can be calculated by combining average responses to each of the 10 components (i.e., 2 general questions and questions for the 8 subscales). Each subscale on the RAND ranges from 0 to 100 percent with higher scores on the RAND indicative of a healthier response to questions. The reliability of the RAND has been found to be α ≥ .70 (Busija et al., 2011). In the current study, we observed its reliability to be α = .98.

### Mood.
The PANAS questionnaire (Watson et al., 1988) was used to assess the average mood of participants. The PANAS contains 20 different positive and negative adjectives. Participants were asked “in the past few weeks” how well the adjectives described them on a 5-point Likert-type scale from 1 (not at all true) to 5 (really true). The scale ranges from 20 to 100 in which higher scores on the PANAS indicate a more positive mood. The PANAS has been found to have high internal reliability (α = .96, Rose, 2002).

### TABLE 1

<table>
<thead>
<tr>
<th>Measure</th>
<th>n</th>
<th>M</th>
<th>SD</th>
<th>Variance</th>
<th>Min.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical Functioning</td>
<td>99</td>
<td>90.66</td>
<td>18.96</td>
<td>359.51</td>
<td>5.00</td>
<td>100.00</td>
</tr>
<tr>
<td>Limitations (Physical)</td>
<td>99</td>
<td>91.92</td>
<td>18.84</td>
<td>354.95</td>
<td>0.00</td>
<td>100.00</td>
</tr>
<tr>
<td>Limitations (Emotional)</td>
<td>99</td>
<td>68.35</td>
<td>37.31</td>
<td>1391.70</td>
<td>0.00</td>
<td>100.00</td>
</tr>
<tr>
<td>Energy</td>
<td>98</td>
<td>54.80</td>
<td>16.40</td>
<td>269.03</td>
<td>0.00</td>
<td>100.00</td>
</tr>
<tr>
<td>Emotional Well-Being</td>
<td>99</td>
<td>74.63</td>
<td>13.74</td>
<td>188.79</td>
<td>24.00</td>
<td>100.00</td>
</tr>
<tr>
<td>Social Functioning</td>
<td>98</td>
<td>79.85</td>
<td>22.46</td>
<td>594.62</td>
<td>25.00</td>
<td>100.00</td>
</tr>
<tr>
<td>Pain</td>
<td>99</td>
<td>83.28</td>
<td>18.16</td>
<td>329.80</td>
<td>0.00</td>
<td>100.00</td>
</tr>
<tr>
<td>General Health</td>
<td>99</td>
<td>70.61</td>
<td>21.77</td>
<td>474.12</td>
<td>15.00</td>
<td>100.00</td>
</tr>
<tr>
<td>Physical Health</td>
<td>99</td>
<td>84.12</td>
<td>13.17</td>
<td>173.37</td>
<td>20.00</td>
<td>100.00</td>
</tr>
<tr>
<td>Emotional Health</td>
<td>97</td>
<td>69.21</td>
<td>17.57</td>
<td>308.58</td>
<td>20.38</td>
<td>97.50</td>
</tr>
<tr>
<td>Total RAND</td>
<td>100</td>
<td>120.5</td>
<td>11.63</td>
<td>135.32</td>
<td>81.00</td>
<td>146.00</td>
</tr>
<tr>
<td>MOS</td>
<td>100</td>
<td>81.00</td>
<td>8.93</td>
<td>79.78</td>
<td>51.00</td>
<td>90.00</td>
</tr>
<tr>
<td>PANAS (Positive)</td>
<td>95</td>
<td>14.07</td>
<td>1.94</td>
<td>3.76</td>
<td>8.00</td>
<td>26.00</td>
</tr>
<tr>
<td>PANAS (Negative)</td>
<td>97</td>
<td>4.69</td>
<td>5.29</td>
<td>28.01</td>
<td>2.67</td>
<td>8.67</td>
</tr>
<tr>
<td>Self-reported Corumination</td>
<td>100</td>
<td>77.42</td>
<td>23.39</td>
<td>547.24</td>
<td>27.00</td>
<td>135.00</td>
</tr>
<tr>
<td>Facebook Corumination</td>
<td>90</td>
<td>30.63</td>
<td>8.86</td>
<td>78.47</td>
<td>12.00</td>
<td>59.00</td>
</tr>
</tbody>
</table>

Note. RAND = Research and Development Health Survey; MOS = Medical Outcomes Study; PANAS = Positive and Negative Affect Schedule.
Corumination and Health in Young Women | Swearingen, Byrd-Craven, and Kennison

1 (very slightly or not at all) to 5 (extremely). Questions were also divided into positive and negative adjectives to describe the mood. Each subscale of the PANAS ranges from 10 to 50 with higher scores on the PANAS indicative that participants considered themselves to have more positive attitudes or negative attitudes. The reliability of the PANAS has been found to be .85 for negative affect and .89 for positive affect (Crawford & Henry, 2004). In the current study, we observed its reliability to be $\alpha = .90$ for negative affect and $\alpha = .90$ for positive affect.

Social support. The MOS of Social Support (Sherbourne & Stewart, 1991) measures the overall functional social support that participants receive and contains four subscales of social support. Scores range from 0 to 100 percent with higher scores on the overall social support and on the four subscales indicative of a higher social support. Participants were instructed to complete the survey when considering the social support of their friend. The reliability of the MOS has been found to be between .92 and .97 (Sherbourne & Stewart, 1991). In the current study, we observed its reliability to be $\alpha = .93$ for overall social support.

Procedure

After Oklahoma State University institutional review board approval (AS-10-113) was given, participants were able to sign up for the study using the online system. Participants who signed up for the present study received an e-mail asking them to come to the laboratory with a copy of Facebook conversations that occurred between the participant and a close female friend. They were sent a consent form for their friend, which gave permission for the conversations to be used in the research. The conversations were to have occurred over a 3-day period for at least 15 min a day. Participants were asked to save the formatted conversations into a Microsoft® Word® document and to replace the names involved in the conversation with the letters A for themselves and B for the other party in the conversation. When participants arrived to the lab session, they were asked to sign a consent form and provide the consent form from their friend in the conversation. They were asked for the formatted conversations and asked to fill out the four additional questionnaires. All participants received instructions in the e-mail to (a) pick a close female friend with whom they regularly use Facebook Chat; (b) chat with the friend at least 15 min each day through Facebook Chat; (c) save each file with their participant identification number; (d) copy and paste all conversation from their text window on Facebook Chat into the Word document; (e) remove their name from the transcript; and (f) print the transcript to submit to the laboratory.

The Facebook conversations were coded for the content and amount of corumination within the conversation. The coding for each conversation was conducted by three trained undergraduate research assistants. The system of coding corumination developed by Rose, Schwartz, and Carlson (2005) was adapted for the present study. The conversations were divided into four areas of corumination. Examples of each of the areas were provided for clarification for the coders. Rehashing was characterized by “talking about every detail of the problem, talking about parts of the problem over and over.” Examples of speculating were “talking about why the problem might have happened, talking about bad things that might happen because of the problem, talking about parts of the problem that are not understood.” Dwelling on negative affect pertained to “talking about how bad the person with the problem feels, talking about how upset, sad, or mad the person with the problem feels (or should feel).” Encouraging problem talk was exemplified by “trying to keep one another talking about the problems, trying to get each other to tell every detail about the problems.” Each coder rated the conversations on a 1 to 5 scale for the four different areas. Coders rated a core set of conversations until their reliability was $\alpha > .90$. After the initial core set of conversations, every 10th participant was double coded to ensure consistent reliability. The interrater reliability for each area was as follows: rehashing ($\alpha = .82$); speculation ($\alpha = .88$); negative affect focus ($\alpha = .92$); mutual encouragement ($\alpha = .89$). When the total scores for all areas were calculated, the average of the three raters was taken whenever the raters score fell within 2 points of each other. If differences between the raters were more than 2 points apart, the outlier was dropped and the average of the two similar scores was taken. The averages from the coders provided an indication of the amount of corumination that was found within the conversations.

Results

We analyzed participants’ self-reported daily corumination, their corumination carried out through Facebook, as well as their self-reported health, mood, and social support. In an effort
to address the data dependency between dyads that participated together, participants filled out questionnaires separate from their dyadic partner. Conversations were only scored based on the individual responses without consideration to what the other partner in the conversation stated. Table 1 displays the descriptive statistics for these variables. The results supported that individuals coruminate in daily life when communicating through Facebook. Similar amounts of corumination were observed in each of the three conversations carried out through Facebook. Total observed Facebook corumination was used as a predictor in lieu of average observed corumination to reduce the influence of outliers, as was done in previous studies (Byrd-Craven et al., 2008; Byrd-Craven et al., 2011). The total observed Facebook corumination in the first conversation significantly predicted the total observed Facebook corumination in the second conversation, $F(1, 88) = 34.53$, $p < .001$, $\beta = .53$, $r^2 = .22$, and the total observed Facebook corumination in the second conversation significantly predicted the total observed Facebook corumination in the third conversation, $F(1, 88) = 24.42$, $p < .001$, $\beta = .47$, $r^2 = .22$.

Regression analyses were also conducted to analyze the subcomponents of corumination (i.e., Rehashing, Speculation, Encouraging Problem Talk, and Negative Affect). The results showed that Rehashing observed in Facebook communications predicted the amount of rehashing occurring in reported communications, $F(1, 88) = 16.18$, $p < .001$, $\beta = .39$, $r^2 = .16$. The amount of Rehashing in the second conversation on Facebook predicted Rehashing in the third conversation, $F(1, 88) = 25.18$, $p < .001$, $\beta = .54$, $r^2 = .22$. Speculation in the first Facebook conversation predicted Speculation in the second Facebook conversation, $F(1, 88) = 5.40$, $p = .02$, $\beta = .23$, $r^2 = .06$. Negative Affect Focus in the first Facebook conversation predicted Negative Affect Focus in the second Facebook conversation, $F(1, 88) = 10.75$, $p = .001$, $\beta = .32$, $r^2 = .11$, and Negative Affect Focus in the second Facebook conversation predicted Negative Affect focus in the third Facebook conversation, $F(1, 88) = 14.64$, $p < .001$, $\beta = .45$, $r^2 = .14$. Encouraging Problem Talk in the first Facebook conversation predicted Encouraging Problem Talk in the second Facebook conversation, $F(1, 88) = 16.83$, $p < .001$, $\beta = .35$, $r^2 = .16$, and Encouraging Problem Talk in the second Facebook conversation predicted Encouraging Problem Talk in the third Facebook conversation, $F(1, 88) = 16.55$, $p < .001$, $\beta = .49$, $r^2 = .16$.

Correlational analyses were conducted to investigate the interrelatedness of the subscales of the RAND, the PANAS, the MOS, corumination on Facebook, and self-reported daily corumination. A summary of these results are displayed in Table 2. As in prior research (Hays, Sherbourne & Mazel, 1995), the subscales of the RAND health measure were positively related to one another. Total social support measured through the MOS scale was positively correlated with general health scores ($r = .23$, $p = .02$). Further analyses suggested that self-reported daily corumination was positively associated with emotional limitations ($r = .20$, $p = .05$), and observed Facebook corumination was positively associated with physical health ($r = .21$, $p = .05$) and physical functioning ($r = .23$, $p = .03$). Physical functioning was negatively related to the positive subscale of the PANAS ($r = -.21$, $p = .04$), as higher levels of physical function were related to lower levels of positive affect.

To investigate how self-reported daily corumination and corumination on Facebook predicted health, we conducted a multiple regression using a hierarchical approach. We used the total score for the RAND for each participant as the dependent variable in the hierarchical regression. We investigated whether self-reported daily corumination and corumination on Facebook could significantly improve model fit, explaining variance in addition to the variance accounted for by the total social support as measured by MOS. Table 3 displays a summary of the variables, standard errors, beta values, standardized betas, and $R^2$ from this analysis. Both models had acceptable Variance Inflation Factor (VIF) and Tolerance values. In Block 1, VIF ranged from 1.00 to 1.03, and Tolerance ranged from .70 to .87. In Block 2, VIF ranged from 1.05 to 1.44, and Tolerance ranged from .97 to .99. Additionally, with means at or around zero, the following diagnostic statistics indicated that both models were a good fit: Residual, Cook’s Distance, Studentized Residual, and Centered Leverage Value. The predictor entered into the first block was the MOS, which accounted for 4% of the variance, Block 1, $F(1, 89) = 3.35$, $p = .07$. There was a nonsignificant trend for those with higher levels of social support having higher levels of overall health. The predictors added into Block 2 along with MOS were self-reported daily corumination and corumination on Facebook. The predictors included in Block 2, $F(3, 89) = 3.81$, $p = .02$, $r^2 = .12$, accounted for 12% of the variance in overall health. As expected, the addition of corumination significantly improved
Corumination and Health in Young Women | Swearingen, Byrd-Craven, and Kennison

the fit of the model. Controlling for social support, self-reported daily corumination was a significant predictor of overall health, indicating that a higher level of daily corumination was associated with lower levels of health. Social support was also a significant positive predictor of health. See Table 3 for the standardized beta weights, standard errors, beta values, and $R^2$ change.

**Discussion**

The results of the present research supported that corumination does occur in young women’s communications through Facebook. The elements of corumination (i.e., Rehashing, Speculation, and Negative Affect) were related in the three conversations carried out on Facebook, which suggested that corumination through Facebook continues throughout the interaction. Correlation results also showed that the observed Facebook corumination was significantly positively related to health factors, and self-reported corumination was also positively, but not significantly, related to health factors. Although this result may seem contradictory, beneficial aspects of corumination such as higher quality relationships (Rose, 2002) have to be considered. The results confirmed that, when the positive benefits of social support were controlled in a hierarchical regression analysis, self-reported daily corumination negatively predicted overall health as measured by participants’ total RAND scores. It is important to note that reported self-reported daily corumination and observed Facebook corumination are not correlated but are both predictors of the health factors. Given that both having observed Facebook conversations and daily self-perspective of corumination add unique view points to the situation, the lack of association was not surprising. The lack of association may also be influenced by observed corumination in Facebook conversations only accounting for a small portion of corumination that an individual is partaking on daily basis.

The results suggested that some users of social networking sites such as Facebook unknowingly engage in communications that may adversely affect their health. Because of the increasing availability of smart cellular phones and other devices such as tablets that allow users to communicate through social networking throughout the day, we expect that individuals who regularly engage in corumination with close friends will have more opportunities to do so. Those individuals engaging in high levels of corumination are likely to not only experience higher levels of stress at the time...
that the communications occur, but may also put
themselves at risk in the future because overall
health can decline as the result of chronic stress.

In the present study, we allowed participants
to select any conversation that occurred over the
3-day period to submit to the study for analysis.
We recognize that this was an important limita-
tion of the study. It is possible that participants’
choices might have led to an underestimation of
the corumination occurring in their relationships
because participants might have avoided sharing
conversations that focused on especially private life
experiences. In future research, it may be possible
to enroll participants in the study and to collect
conversations (with participants’ permission) as
they occur in real time. This method would create
a sample of conversations that are more represen-
tative of the entire range of conversation topics
that young women experience. Additional charac-
teristics of the individuals’ social network such
as number of friends and time spent using social
networks and in depth demographics were also
not assessed. Number of friends has been shown
to be negatively related to personality factors such
as shyness (Orr et al., 2009) and to corumination
(Tompkins et al., 2011). Future research should
expand upon details of individuals’ social networks
that could influence online behaviors. In addition,
small effect sizes and correlations limited the inter-
ruption of the data and warrant future replication
and more research.

Future research should also explore the extent
to which corumination carried out on social net-
working sites such as Facebook lead to changes in
the biological markers for stress (i.e., cortisol and
alpha amylase). Prior research has found that coru-
mination occurring in face-to-face communications
was related to higher levels of cortisol and alpha
amylase (Byrd-Craven et al., 2008). We expect that
similar results would be obtained during commun-
ication regardless of the mode of communication
(i.e., face-to-face vs. on a cellular phone vs. via a
social-networking site). Because individuals may
communicate more often via cellular phones and
social networking sites than they communicate
face-to-face and because there may be differences
in characteristics of the communications (e.g.,
duration, intensity), differences in the biological
stress responses may be found for the different
methods of communicating.

In sum, the present research provided evi-
dence that young women coruminate when
communicating through Facebook. Their levels of
corumination in daily life predicted overall
health with those engaging in the highest levels of
corumination in daily life having the lowest health
scores when the positive benefits of social support
were statistically controlled. We anticipate that the
use of social networking sites such as Facebook will
increase in the future. Consequently, there is the
danger that the prevalence of corumination will
increase. We hope that this initial study of corumi-
nation on Facebook provided an impetus for others
to investigate corumination, health, and/or stress
response in studies of human communication.

References

Busija, L., Pausenberger, E., Haines, T. P., Haymes, S., Buchbinder, R. &
Osborne, R. H. (2011). Adult measures of general health and health-
related quality of life: Medical Outcomes Study Short Form 36-Item
(SF-36) and Short Form 12-Item (SF-12) Health Surveys, Nottingham
Health Profile (NHP), Sickness Impact Profile (SIP), Medical Outcomes
Study Short Form 6D (SF-6D), Health Utilities Index Mark 3 (HUI3), Quality
of Well-Being Scale (QWB), and Assessment of Quality of Life (AQOL).
Arthritis Care and Research, 63, S383–S412. doi:10.1002acr.20541

increases stress hormone levels in women. Hormones and Behavior, 53,

Byrd-Craven, J., Granger, D. A., & Auer, B. J. (2011). Stress reactivity to co-
rumination in young women’s friendships: The relationship between
cortisol and alpha-amylase. Journal of Social and Personal Relationships,
28, 469–487. doi:10.1177/0265407510382319

Does co-rumination explain gender differences in emotional distress and
relationship satisfaction among college students? Cognitive Therapy and
Research, 32, 577–590. doi:10.1007/s10608-008-9200-3

Crawford, J. R., & Henry, J. D. (2004). The Positive and Negative Affect
Schedule (PANAS): Construct validity, measurement properties, and
normative data in a large nonclinical sample. British Journal of Clinical
Psychology, 43, 245–265. doi:10.1348/0144665031752594

Denson, T. F., Spanovic, M., & Miller, N. (2009). Cognitive appraisals and
emotions predict cortisol and immune responses: A meta-analysis of
crural laboratory social stressors and emotion inductions. Psychological

Engert, V., Vogel, S., Efanov, S. I., Duchesne, A., Corbo, V., Ali, N., &

Swearingen, Byrd-Craven, and Kennison | Corumination and Health in Young Women

TABLE 3
Hierarchical Regression Results With Social Support, Self-Reported Corumination, and Corumination on
Facebook Predicting Health

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Block 1</th>
<th></th>
<th></th>
<th>Block 2</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>101.11</td>
<td>10.65</td>
<td>101.11</td>
<td>11.53</td>
<td>101.22</td>
<td>25.04</td>
</tr>
<tr>
<td>Social Support</td>
<td>0.13</td>
<td>0.07</td>
<td>0.13</td>
<td>0.24</td>
<td>0.24</td>
<td>0.02</td>
</tr>
<tr>
<td>SR Corumination</td>
<td>0.04</td>
<td>0.12</td>
<td>0.03</td>
<td>0.04</td>
<td>0.12</td>
<td>0.03</td>
</tr>
<tr>
<td>FB Corumination</td>
<td>-0.13</td>
<td>-0.27</td>
<td>0.01</td>
<td>0.05</td>
<td>0.26</td>
<td>0.01</td>
</tr>
<tr>
<td>R²</td>
<td>0.04</td>
<td>0.12</td>
<td>0.03</td>
<td>0.04</td>
<td>0.12</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Note. Dependent variable = Total RAND; SR = Self-Reported; FB = Facebook.
Corumination and Health in Young Women


Author note. CaSandria Swearingen, Jennifer Byrd-Craven, and Shelia M. Kennison, Department of Psychology, Oklahoma State University.

Correspondence concerning this article should be addressed to CaSandria Swearingen, 116 North Murray Hall, Department of Psychology, Oklahoma State University, Stillwater, OK 74078. E-mail: casanls@okstate.edu
Expressions of humor can be found across all cultures (Martin, 2010). Humor has a behavioral component that includes physical responses such as laughing and a physiological component that involves the heart muscle, circulatory system, endorphins, and muscles (Martin, 2001). Additionally, the stereotyped characteristics of laughter such as internote interval and decrescendo serve as a social component used to initiate and maintain social discourse (Provine & Yong, 1991). These stereotyped features along with the reinforcing nature of laughter establish humor as a universal phenomenon ideal for human communication (Martin, 2010). Because of the physical, hormonal, and endocrinological changes (Berk, Felten, Tan, Bittman, & Westengard, 2001; Hubert, Möller, & de Jong-Meyer, 1993), humor benefits mental, emotional, and physical health including the immune system (Bennett & Lengacher, 2007; Dillon, Minchoff, & Baker, 1986).

Research has shown multiple sources of humor such as oddities, incongruences, playfulness, and unexpected comicality (Goldstein & McGhee, 2013). However, the underlying sources of humor may not always be benevolent. It is not uncommon for people to laugh at someone else’s misfortune or defilement of character and reputation. This type of humor is called disparagement humor (Ferguson & Ford, 2008), also known as schadenfreude (Ben-Ze’ev, 1992). In the present study and the related literature review, we focused on the perception of disparagement humor or finding a joke about someone’s misfortune amusing.

**ABSTRACT.** Explanations of why individuals find other people’s misfortune amusing range from unconscious urges to elevated self-esteem. The present investigation was about the influence of induced empathy on reducing ratings of disparagement humor. Sixty-four undergraduate students participated in this experimental study. In the pretest, they rated video clips eliciting disparagement humor in terms of funniness and pleasure. Afterward, the experimental group read an empathy-inducing story whereas the control group read a neutral story. Both groups then watched and rated 2 new video clips in the posttest. Participants also completed questionnaires for dispositional empathy and humor. As expected, the experimental group showed a significant decrease in ratings of funniness, $F(1, 61) = 3.76$, $p = .05$, $\eta^2_p = .05$, and pleasure, $F(1, 61) = 5.43$, $p = .023$, $\eta^2_p = .08$, whereas the control group did not. Dispositional empathy was positively correlated with coping humor, $r = .34$, $p = .006$, and negatively correlated with negative attitudes toward humor, $r = -.42$, $p = .001$. Future research could expand the understanding of disparagement humor by investigating the cognitive appraisal that takes place and the associated feelings created before the experience of disparagement humor. Empathy appears to have the potential to alter the feelings resulting from such appraisal, thereby, tempering the perception of humor.

**The Effects of Empathy on Disparagement Humor**

Peter Bui, Maria Kalpidou*, Lauren DeVito, and Todd Greene
Assumption College
Effects of Empathy on Disparagement Humor | Bui, Kalpidou, DeVito, and Greene

Why Is Disparagement Humor Funny?
Disparagement humor has puzzled philosophers and psychologists alike. From the psychoanalytic perspective, engagement in disparagement humor serves as a defense mechanism to protect a person’s self-worth (Ferguson & Ford, 2008). Freud (1960) speculated that disparagement or hostile humor allows people to relieve themselves of unacceptable and impure unconscious desires, a process similar to catharsis. To test this theory, Singer (1968) induced participants with aggressive motivation and then successfully reduced this unacceptable emotion by exposing participants to hostile humor. However, Berkowtiz (1970) failed to demonstrate the cathartic role of hostile humor. Women participating in the study evaluated a female job applicant less favorably when they were exposed to offensive hostile humor than when they were exposed to nonhostile humor. Therefore, the evidence for the psychoanalytic perspective has been rather conflicting.

An alternative explanation to disparagement humor is the social identity theory (Tajfel & Turner, 1986). People may find happiness in another’s group misfortune in an area where two groups may be competing. Leach, Spears, Branscombe, and Doosje (2003) showed that Dutch soccer fans, who were induced with the belief that the German soccer team was a threat, expressed greater pleasure when hearing that the Germans lost the World Cup. Similarly, other studies have supported that perceived funniness of disparagement humor is greater when made against an outgroup (Abrams & Bippus, 2011). This suggests that group biases influence the experiences of malicious pleasure. However, social identity theory does not explain why people laugh at the misfortune of strangers whose affiliations with groups are unknown (Ferguson & Ford, 2008).

Superiority theory somewhat addresses the limitation of social identity theory by suggesting that pleasure at another’s misfortune derives from negative feelings for that person (Gruner, 1997). In support of superiority theory, feelings of dislike have been found to regulate experiences of disparagement humor (Hareli & Weiner, 2002; Wicker, Barron, & Willis, 1980). Hareli and Weiner (2002) found that induced feelings of dislike, but not envy, increased the amount of pleasure in another’s misfortune presented in hypothetical scenarios, with a mild level of misfortune generating greater pleasure. Hareli and Weiner (2002) speculated that feelings of dislike might have heightened the humorist’s self-esteem. Building on this idea, Van Dijk, van Koningsbruggen, Ouwerkerk, and Wesseling (2011) discovered that participants with lower self-esteem experienced greater pleasure at a fellow student’s misfortune when they also perceived this person as threatening.

Unlike the other explanations, superiority theory does not rely on the humorist’s identification of a group or unconsciously denied impulses. However, the underlying commonality in all perspectives is that the humorists harbor some sort of negative or unpleasant feelings. If the superiority explanation is accurate, a person should expect that inducing positive feelings such as empathy would decrease the level of amusement in disparagement humor. The current study explored this possibility.

The Relationship Between Humor Perception and Empathy
Empathy, defined as the ability to adapt a person’s way of thinking and behaving in order to understand another’s perspective and emotions (Rogers, 1980) has been associated with the ability to comprehend humor. Furthermore, inability to take someone else’s perspective has been linked to poor understanding of humor as illustrated in studies with adults (Samson, 2013) and children with autism (Baron-Cohen, 1997; Emerich, Creaghead, Grether, Murray & Grasha, 2003).

The direction of the relationship between empathy and humor is unclear. Earlier evidence suggested a positive correlation (Hampes, 2001). Using questionnaire data, Hampes (2010) found that humor aiming to create affiliations with others was positively correlated with empathic concern, but aggressive humor (i.e., laughing at others) was negatively correlated with empathic ability. In an experimental study, Greitemeyer, Osswald, and Brauer (2010) showed that playing prosocial video games increased empathy and decreased schadenfreude toward a male celebrity. The above studies supported a negative relationship between empathy and aggressive or disparagement humor, but there is limited causal evidence for this relationship.

The Current Study
Based on previous findings that inducing dislike increased the experience of disparagement humor (Hareli & Weiner, 2002; Van Dijk et al., 2011) whereas increasing empathy decreased schadenfreude (Greitemeyer et al., 2010), we sought to provide stronger experimental evidence for the

The Current Study
role of empathy. Unlike Greitemeyer et al. (2010), we took pre- and postratings of humor and used targets that were not known to participants. We expected that participants in the induced-empathy condition would report lower ratings of funniness and pleasure after watching the video clips than participants in the control group. A significant interaction between the pre- and postmeasurement of humor and the control versus experimental group would support this hypothesis.

Additionally, we included questionnaires that assessed humor and empathy as dispositional traits. Dispositional humor involves the tendency to produce humor, use humor as a coping mechanism, and hold a positive attitude toward humor and humorists (Thorson & Powell, 1993). Dispositional empathy includes the ability to share other people’s emotions and respond sensitively to them (Spreng, McKinnon, Mar, & Levine, 2009). Based on the relationship between empathy and humor in individuals with autism (Baron-Cohen, 1997; Emerich et al., 2003) and the work of Hampes (2001, 2010) on aggressive and affiliative humor as discussed above, we expected significant correlations between empathy and humor, the direction of which would depend on the type of humor (i.e., a positive correlation between using humor as coping and empathy, a negative correlation between negative attitudes toward humor and empathy).

**Method**

**Participants**

Participants were 64 undergraduate students from a small, liberal arts Catholic college in the northeast. The average age of participants was 20.20 years ($SD = 1.48$). Most participants were women (92.2%), European American (81.3%; 9.4% Asian American; 9.3% other), and 15.6% were seniors, 39.1% juniors, 37.5% sophomores, and 7.8% first-year students. Participants’ majors included 40.7% in the social sciences, 35.9% in humanities, 18.8% in sciences, and 4.8% in business studies. Most participants who attended the college came from middle-class families. Participants were recruited from undergraduate psychology courses and may have received extra course credit for their participation. The student researchers e-mailed psychology professors for permission to recruit from their classes. At the beginning of class, researchers briefly described the study and passed out sign-up sheets for interested students to provide contact information. There were no exclusion criteria, and the sample size was determined by the restricted period available for data collection (two weeks in spring 2013).

**Measures**

Measures included a demographic scale and questionnaires about dispositional humor and empathy.

**Dispositional humor.** The Multidimensional Sense of Humor Scale (MSHS; Thorson & Powell, 1993) was designed to measure four dimensions of dispositional humor as well as an overall score. The initial scale has 24 items and yields the following dimensions: (a) Humor Production and Social Uses of Humor (e.g., “I can often crack people up with the things I say”), (b) Coping Humor (e.g., “Humor helps me cope”), (c) Negative Attitudes Toward Humor (e.g., “People who tell jokes are a pain in the neck”), and (d) Positive Attitudes Toward Humor (e.g., “I appreciate those who generate humor”). Participants rated the items using a 5-point Likert scale, which ranged from 1 (strongly disagree) to 5 (strongly agree). To calculate the overall score for dispositional humor, we reversed seven negatively worded items. To improve reliability of MSHS, we removed one item (“I can actually have some control over a group by my uses of humor”) and achieved a Cronbach alpha of .85. The alphas for the MSHS subscales ranged from .62 to .87.

**Dispositional empathy.** The Toronto Empathy Questionnaire (TEQ; Spreng et al., 2009) was used to measure participants’ level of dispositional empathy. The TEQ consists of 16 items, eight of which are positively worded (e.g., “When someone else is feeling excited, I tend to get excited too”) and eight of which are negatively worded (e.g., “Other people’s misfortunes do not disturb me a great deal”). Items assess comprehension of others’ emotions, experience of sympathetic physiological arousal, altruism, prosocial behavior, understanding of emotional states in others based on sensitivity responses, and the extent to which a person experiences the same emotional state of another. Participants rated the items using a 5-point Likert-type scale (1 = never, 2 = rarely, 3 = sometimes, 4 = often, 5 = always). The Cronbach’s alpha for TEQ was .72 with the removal of one item (“I find it silly for people to cry out of happiness”). The average score of the remaining items in each scale was used in the analyses.

**Procedure**

The study was approved by the Assumption College institutional review board (#2013-15) and took place in the computer lab of the psychology
department. The lab had eight computer stations divided by walls. Six to eight participants were tested at the same time. To increase privacy, participants used headphones when watching video clips. After signing the consent forms, participants received a package of the questionnaires and instructions about the order of the tasks. The video clips were available on the desktop, and participants worked independently to complete the procedure, which included the following tasks in this order: completing the MSHS, experiencing the experimental or the control condition, and completing the TEQ. Two undergraduate student researchers were present during the experiment.

Experimental condition. After completing the MSHS, participants watched two short clips (Clip 1: child is being hit by a swinging stick and Clip 2: child falls off a bike into a mud pit), each of which lasted less than 30 s. Following each clip, participants rated on a 5-point Likert scale (1 = strongly disagree to 5 = strongly agree) the extent to which they agreed with three statements: “I found this video to be funny” and “Viewing the clip was pleasurable.” The third statement “I knew the people in this video” was added to ensure that participants did not recognize the protagonists in the video clips. Next, participants read a nonfictional short narrative entitled “NPR Social Work” (NPR, 2008). The story is about a social worker who was robbed at knife-point on his way home from work but, instead of turning the young thief to the police, he treated the thief to a meal. In return, the thief turned over the knife and the social worker’s wallet.

Thereafter, participants rated the extent to which they felt sympathetic, warm, compassionate, soft-hearted, tender, and moved using a 7-point Likert-type scale from 1 (not at all) to 7 (extremely; Batson, Klein, Highberger, & Shaw, 1995). Next, participants watched two new video clips (Clip 3: a child rolls off a bike and Clip 4: a child experiences a failed trust fall) and rated the extent to which they found the videos to be funny, pleasurable, and familiar. Lastly, participants completed the TEQ.

Control condition. Participants in the control group experienced the same procedure as participants in the experimental group, but they read a neutral story instead of the empathic story. We used an excerpt from “Nature’s Destiny” (Denton, 2002), which explained the alternatives to light energy as a means for organic life.

Design. Participants were assigned to either the control or the experimental conditions in a counterbalanced order, resulting in equal sample size (n = 32). The order of the paired clips before and after the story (empathic or neutral) was also counterbalanced, thus creating four combinations: (a) pretest Clips 1 and -2, experimental condition, posttest Clips 3 and -4; (b) pretest Clips 1 and -2, control condition, posttest Clips 3 and -4; (c) pretest Clips 3 and -4, experimental condition, posttest Clips 1 and -2; (d) pretest Clips 3 and -4 control condition, posttest Clips 1–2. Completion of the procedure took approximately 30 min. At that time, participants were debriefed about the experiment and were asked not to discuss the study with others. All participants who showed up were able to participate and complete the procedure. The study was concluded during the course of one academic semester.

Results

We first conducted analyses to ensure that the empathic story was effective, that the two groups were not different in dispositional humor and empathy, and that the participants were not familiar with the protagonists in the video clips. We tested the effectiveness of the empathic story by comparing the control and the experimental groups using a multivariate analysis of variance with group (control, experimental) as the independent variable and each of the empathy adjectives as the dependent variables. The analysis showed a significant main effect of group on ratings of sympathy, $F(1, 62) = 120.76, p < .001, \eta^2 = .66$, warmth, $F(1, 62) = 207.65, p < .001, \eta^2 = .77$, compassion, $F(1, 62) = 242.80, p < .001, \eta^2 = .80$, softheartedness, $F(1, 62) = 104.59, p < .001, \eta^2 = .63$, tender, $F(1, 62) = 178.89, p < .001, \eta^2 = .63$, and moved, $F(1, 62) = 27.78, p < .001, \eta^2 = .31$. As shown in Table 1, the scores of these perceived empathic states were significantly higher in the experimental than the control group. We also compared the two groups

<table>
<thead>
<tr>
<th>TABLE 1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Descriptive Statistics for Induced Empathy</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Experimental</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>M (SD)</td>
<td>95% CI</td>
<td>M (SD)</td>
</tr>
<tr>
<td>n = 32</td>
<td>n = 32</td>
<td></td>
</tr>
<tr>
<td>Sympathetic</td>
<td>5.43 (1.52)</td>
<td>(4.88, 5.98)</td>
</tr>
<tr>
<td>Warm</td>
<td>5.84 (0.88)</td>
<td>(5.52, 6.16)</td>
</tr>
<tr>
<td>Compassionate</td>
<td>5.96 (0.82)</td>
<td>(5.67, 6.26)</td>
</tr>
<tr>
<td>Soft-hearted</td>
<td>5.37 (1.43)</td>
<td>(4.85, 5.89)</td>
</tr>
<tr>
<td>Tender</td>
<td>5.09 (1.22)</td>
<td>(4.65, 5.53)</td>
</tr>
<tr>
<td>Moved</td>
<td>6.40 (0.94)</td>
<td>(6.06, 6.74)</td>
</tr>
</tbody>
</table>
in terms of dispositional humor and dispositional empathy to ensure that there were no preexisting differences that could potentially confound the effects of the experimental manipulation. These analyses did not yield any significant results (see Table 1). Lastly, we analyzed participants’ responses in the third statement (“I knew the people in this video”) after viewing the video clips and found that both groups were not familiar with the targets in the pairs of clips.

To test the hypothesis that invoking empathy would hinder the perception of how funny or pleasurable disparagement clips were, we conducted two repeated measures Analyses of Variance with dependent variables being the average scores across the clips for funniness and pleasure. The between-group variable was the control or the experimental group and the within-group factor was time (pre and post). We included the order of video clips as a covariate because, although we counterbalanced the order of the video clips across the control and experimental groups, we did not counterbalance the order of the clips within each pair (Clips 1 and 2 for the pretest and Clips 3 and 4 for the posttest).

There was a significant interaction effect between order of clips and time, $F(1, 61) = 7.58$, $p = .008$, $\eta^2 = .11$. Watching Clips 3 and 4 in the pretest and Clips 1 and 2 in the posttest resulted in greater reduction in ratings of funniness in both the experiment and the control group. Through this analysis, it was evident that the interaction between order and time meant that the disparagement humor clips were not all equally funny. There was no significant main effect of the group, $F(1, 61) = .008$, $p = .928$, or the order in which the pairs of clips were viewed, $F(1, 61) = .001$, $p = .982$. As shown in Figure 1, there was a significant interaction effect between time and group that supported our hypothesis, $F(1, 61) = 3.76$, $p = .05$, $\eta^2 = .05$. The experimental group experienced a decrease in how funny the video clips were ($M_{\text{pre-exp}} = 3.49$, $SD = 0.66$; $M_{\text{post-exp}} = 3.22$, $SD = 0.78$), but the control group did not ($M_{\text{pre-control}} = 3.33$, $SD = 0.87$; $M_{\text{post-control}} = 3.42$, $SD = 0.76$). Lastly, there was a significant main effect of the within-factor time, $F(1, 61) = 4.65$, $p = .035$, $\eta^2 = .07$, indicating that the overall rating of funniness decreased, ($M_{\text{pre}} = 3.41$, $SD = 0.77$; $M_{\text{post}} = 3.52$, $SD = 0.77$).

Similar analyses were performed with the ratings of experienced pleasure from watching the clips as the dependent variable. Again, in support of the hypothesis, we found a significant interaction between the within and between factor (Time x Group), $F(1, 61) = 5.43$, $p = .023$, $\eta^2 = .08$. As shown in Figure 2, the experimental group experienced a decrease in the amount of pleasure ($M_{\text{pre-exp}} = 3.22$, $SD = 0.79$; $M_{\text{post-exp}} = 2.78$, $SD = 0.88$). The control group remained about the same from $M_{\text{pre}} = 3.23$ ($SD = 0.84$) to $M_{\text{post}} = 3.24$ ($SD = 0.74$). The main effect of the within factor of time and the interaction effect between the order of the clips and time were not significant.
Lastly, we conducted correlational analyses between humor and empathy as measured by the questionnaires. As shown in Table 2, there was a significant positive correlation between empathy and humor, $r = .36$, $p = .005$, indicating that a positive approach to humor was related to higher empathy. This finding was in line with our prediction. We further explored the relationship of empathy with specific types of humor to investigate the negative association between negative humor and empathy found in the literature. As shown in Table 2, although empathy positively correlated with coping humor, $r = .34$, $p = .006$, there was a negative correlation between empathy and negative attitudes toward humor, $r = -.42$, $p = .001$. Lastly, there was a marginally significant correlation between positive attitudes toward humor and empathy, $r = .20$, $p = .097$.

**Discussion**

Inducing empathy significantly reduced the ratings of how pleasurable and funny participants perceived disparagement humor to be, thus supporting the main hypothesis. In contrast, participants who read the neutral story experienced similar levels of funniness and pleasure. Moreover, participants who experienced higher dispositional empathy tended to like humor whereas participants with lower dispositional empathy tended to hold negative attitudes toward humor. The results implied that empathy is a powerful emotion that changes the way people think about others in situations that are seemingly funny but potentially harmful.

It appears that lack of empathy might have brought feelings of amusement at someone’s misfortune. This significantly expanded previous research by Greitemeyer et al. (2010) who showed that empathy induced by playing prosocial video games decreased the experience of schadenfreude with regard to a male celebrity’s calamity. The current findings were different in that the target was a stranger, participants in the control and experimental groups did not differ in dispositional empathy or humor, and empathic states before and after the reading of the empathic or neutral story were taken into consideration. Therefore, we demonstrated the effect of empathy on disparagement humor in a much tighter experimental design.

In support of superiority theory, Hareli and Weiner (2002) showed that induced dislike increased feelings of pleasure at another’s misfortune. Similarly, we showed that induced empathy decreased disparagement humor as additional evidence for the superiority theory. It is possible that empathy reduced the need to feel superior by elevating self-esteem (Van Dijk et al., 2011), although including self-esteem as another measure would have allowed us to directly elucidate such effects.

The positive association between dispositional humor and dispositional empathy, also found in previous research (Hampes, 2001), highlighted the possibility of a common underlying mechanism in perceiving humor and experiencing empathy. This is likely the ability to understand the other person’s mental states such as intentions and feelings (Ramachandran, 2011). Moreover, our findings indicated that it was important to consider the relationships between different types of humor and empathy. Similar to previous evidence of Hampes (2010) that empathy was negatively associated with aggressive forms of humor (i.e., teasing, sarcasm, ridicule), we found that participants who reported lower empathy also held more negative attitudes toward humor. In contrast, positive forms of humor such as using humor as a coping strategy and holding positive attitudes toward humor have been associated with increased empathy (Hampes, 2010). These results were in line with the experimental findings, thus further supporting the role of empathy in understanding the perception of disparagement humor.

The findings were important because, in addition to strengthening existing literature, they made significant implications for future theory development and research. Although our findings supported the superiority theory, they did not necessarily dispute the social identity or the psychoanalytic perspective because both perspectives assume that the humorist harbors some unpleasant feeling that is potentially altered by induced empathy. However, by focusing on empathy as part

<table>
<thead>
<tr>
<th>TABLE 2</th>
<th>Correlations Between Dispositional Humor and Dispositional Empathy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Empathy</td>
</tr>
<tr>
<td>Overall Dispositional Humor</td>
<td>.36*</td>
</tr>
<tr>
<td>Humor Production</td>
<td>.18</td>
</tr>
<tr>
<td>Coping Humor</td>
<td>.34*</td>
</tr>
<tr>
<td>Negative Humor</td>
<td>-.42*</td>
</tr>
<tr>
<td>Positive Humor</td>
<td>.20</td>
</tr>
</tbody>
</table>

*Note. N = 64, *p < .05, **p < .01.
of the processes involved in disparagement humor, researchers can propose a model that integrates all perspectives and current research. The process of disparagement humor may begin with the humorist engaging in appraisal of all known information about the target who experiences misfortune including social identity (Tajfel & Turner, 1986), sex (Greitemeyer et al., 2010), and various biases. Such appraisal may also include an estimate of the gravity of the misfortune (e.g., slipping on a banana peel is seen as harmless; Hareli & Weiner, 2002), often in light of the information about the target (i.e., the humorist hypothetically decides that rich celebrities deserve a misfortune because there are overindulged). This cognitive appraisal likely results in either negative or uncaring feelings (e.g., dislike) that allow the humorist to find humor in the target’s mishap. The relationship between appraisal and the associated emotions is probably bidirectional and happens quickly and largely out of awareness (Freud, 1960). Lastly, it is possible that the resulting humor makes people feel better about themselves, thus supporting the superiority theory (Van Dijk et al., 2011). Empathy likely interferes with both the cognitive appraisal and emotional component of the process. Inducing empathy reminds people of their social connection with each other, the moral value of caring for others, and wanes the tendency to respond with laughter. This integrative model merits further investigation.

A future replication of the current project could be improved by expanding the ratings of disparagement humor to include statements such as “I found this video clip humorous,” “This video clip was amusing to me,” “I really enjoyed this video clip,” and “This video was entertaining.” Such an approach would better capture the multiple dimensions of experiencing humor. People may find something very entertaining but not equally funny. This distinction that was not made here and in previous studies may be an important limitation to consider. The generalizability of the present findings was limited to the population of undergraduate students. Therefore, future research should include older adults who might have a different sense of humor and different levels of empathy than college students do. Research has shown an age-related declining ability to understand humor (Mak & Carpenter, 2007) and a linear age-related increase in emotional empathy and prosocial behavior (Sze, Gyurak, Goodkind, & Levenson, 2012). These changes in empathy and humor would possibly lead to a decline in disparagement humor with age.

Expanding the targets in age, sex, and familiarity would be another way to extend future research. Ratings of funniness and pleasure might have decreased more if the targets had been adults similar to the participants because people empathize more with those who are similar to them (Batson, Lishner, Cook, & Sawyer, 2005). Future research should consider sex effects on both empathy and humor. In the literature, women tended to have higher dispositional empathy than men did (Jonason, Lyons, Bethell, & Ross, 2013), and men tended to prefer more affective, self-enhancing, aggressive, and self-defeating kinds of humor than women did (Martin, Puhlik-Doris, Larsen, Gray, & Weir, 2003). A person would expect that induced empathy might be more effective in reducing disparagement humor in women than in men. This brings up a limitation in the present study because most participants were women due to recruiting from psychology courses, which were attended mostly by women. The overrepresentation of women raised the possibility that the effect we observed was stronger than it would have been with more men in the sample. Lastly, a person could consider personality effects. High scores on narcissism, psychopathy, and Machiavellianism tended to associate with low levels of dispositional empathy (Jonason et al., 2013). Individuals with such characteristics would potentially experience higher disparagement humor and not respond to induced empathy.

The current results demonstrated a causal relationship between empathy and disparagement humor. Moreover, empathetic people were more likely to enjoy positive uses of humor and less likely to enjoy aggressive forms of humor. Future research could expand the understanding of disparagement humor by investigating the cognitive appraisal that takes place and the associated feelings created before the experience of disparagement humor. Empathy appears to have the potential to alter the feelings resulting from such appraisal, thereby tempering the perception of humor.

References

Effects of Empathy on Disparagement Humor | Bui, Kalpidou, DeVito, and Greene

doi:10.1037/0022-3514.68.6.1042


Author Note. Peter Bui, Maria Kalpidou, Lauren DeVito, and Todd Greene, Department of Psychology, Assumption College. P. Correspondence concerning this article should be addressed to Maria Kalpidou at Psychology Department, Assumption College, 500 Salisbury Street, Worcester, MA 01609. E-mail: mklapido@assumption.edu; phone: 508-767-7253, fax: 508-767-7263.

Steven V. Rouse
Pepperdine University

ABSTRACT. Set in the context of the history of null hypothesis significance testing, a debate among contemporary researchers and statisticians focuses on whether inferential statistical methods should be revised, reformed, or completely rejected. Although significance tests have been misused and misunderstood in the past, these statistical methods do provide information that, when interpreted accurately, may be valuable. Researchers who use these methods should follow best practices such as calculating and reporting effect sizes, constructing and reporting confidence intervals, and presenting replicated or repeated analyses of the same research questions. When presenting nonsignificant research results such as when the findings contradict previous literature or widely held common assumptions, researchers should evaluate findings in the light of statistical power and score reliability. Several best practices for the use of significance tests help researchers ensure that they are most likely to use inferential statistics in an appropriate manner.

Tartling changes have been afoot in the last two years regarding statistics and the publication process. Two years ago, Trafimow (2014), the incoming editor for the journal Basic and Applied Social Psychology (BASP), announced that inferential statistics were no longer required for journal submissions. Because of concerns about the validity of Null Hypothesis Significance Testing (NHST), authors were no longer required to calculate or provide \( p \) values. Then, last year, Trafimow and Marks (2015) announced an even more extreme policy: NHST procedures were completely banned from publication in BASP. They clarified that any manuscripts including \( p \), \( t \), or \( F \) values, or any discussion of statistical significance would have to be cleaned of all reference to NHST before publication. They concluded,

(w) e hope and anticipate that banning the NHSTP [null hypothesis significance testing procedure] will have the effect of increasing the quality of submitted manuscripts by liberating authors from the stultified structure of NHSTP thinking thereby eliminating an important obstacle to creative thinking. The NHSTP has dominated psychology for decades; we hope that by instituting the first NHSTP ban, we demonstrate that psychology does not need the crutch of the NHSTP, and that other journals follow suit. (p. 2)

Shortly after this landmark decision, editors of several other psychology journals were asked whether they would follow BASP’s lead in banning NHST. Uniformly, the editors of Personality and Social Psychology Bulletin, Journal of Research in Personality, and Journal of Personality and Social Psychology: Interpersonal Relations and Group Processes (Vazire, Wegener, Lucas, & Kawakami, 2015) indicated that they were not considering a similar ban. Neither is Psi Chi Journal. Furthermore,
Teacups and \( t \) Tests | Rouse

the Executive Director of the American Statistical Association released a statement noting that the BASP policy “may have its own negative consequences, and thus the proper use of inferential methods needs to be analyzed and debated in the larger research community” (R. Wasserstein, personal communication, March 3, 2015). However, the issues raised by Trafimow (2014) and Trafimow and Marks (2015) deserve attention. Their decision has sparked valuable conversations about the role of NHST in contemporary psychological research, and provided the editorial board of *Psi Chi Journal* an opportunity to discuss and communicate best practices in the use of NHST.

**Living With the Rules of a Food Dare**

In many ways, contemporary empirical psychological research is still being affected by a single cup of tea that was prepared in the 1920s (Senn, 2012). The beginning of the story differs from one account to another. Salsburg (2001) claimed that the beverage was prepared at a formal tea party for Cambridge University administrators and their wives, while Box (1972) reported that it was a simple tea break among three scientists. From there, however, the stories converge. A cup of tea was poured for a woman, identified as Blanche “Muriel” Bristol in Box’s account, who politely declined, saying that she witnessed the tea being made incorrectly. She explained that she saw the milk being poured into the cup first, followed by the brewed tea, and she explained that the taste of tea was superior when the tea was poured into the cup first, followed by the milk. A lively debate ensued—was it possible to differentiate milk-then-tea cups from tea-then-milk cups by taste alone? A statistician named Ronald A. Fisher proposed an empirical test: eight cups of tea would be prepared for the woman making this claim, but four would be prepared in a tea-then-milk manner and four would be prepared in a milk-then-tea fashion. Then, the order of the eight cups would be randomized. If she could correctly identify all eight cups of tea, her claim would be supported. After all, Fisher reasoned, there are 70 possible sequences of four tea-then-milk and four milk-then-tea cups. Therefore, the probability of guessing the correct sequence by chance alone is merely 1.4% (Fisher, 1956). According to Box (1972) and Salsburg (2001), both of whom heard the story from someone who claimed to be present at the time (in Box’s case, her own father—Fisher), the eight cups were prepared and randomized, and the woman correctly identified each tea-then-milk and milk-then-tea cup.

Certainly, the past must be filled with countless other friendly debates about the ability to differentiate between Pepsi and Coke or between fresh-squeezed and made-from-concentrate orange juice. But this food challenge earned a place in the history of psychology; years later, in his highly influential book on experimental design (arguably the single most influential book in the foundation of research methodology), Fisher (1935) used this particular event as a demonstration of the rigor with which we can examine the veracity of a claim. Indeed, this one event exemplifies the epistemological assumption of NHST: rather than accepting a claim, start with the assumption that the claim is not correct (that is, the null hypothesis), and only reject the null hypothesis if data can be collected that shows a sufficiently low probability of obtaining the results by chance alone. In this case, Fisher was willing to accept a 1.4% probability that he would be wrong—that he would accept Muriel’s claim when she was merely guessing—but given the way the test was constructed, it was highly improbable that she would be right by chance alone. In other words, he was willing to accept a p value of .014 as sufficiently rigorous for rejecting the hypothesis that she was merely guessing. But the question remained of whether that level of rigor was high enough or too high.

In the early years of NHST, as methods were first being created to statistically calculate p values, no universally accepted standard had been set. Cowles and Davis (1982) noted that Fisher himself wrote that different p value cut-off levels could be used, and as late as 1950, statisticians acknowledged that different p value cut-offs could be used for different purposes. However, in 1955, a line was clearly drawn; Cramer (as cited by Cowles and Davis, 1982) set .05 and .01 as cut-off levels that differentiated significant and nonsignificant results. After that, the \( p < .05 \) rule moved from a common standard to a strict rule to a dogmatic divide between publishable and nonpublishable research.

Although a formulaic and dogmatic adherence to a rigid cut-off level (i.e., \( p < .05 \)) developed during the second half of the 20th century, a growing number of researchers and statisticians have highlighted numerous flaws of NHST in the past several years (for in-depth explorations of these flaws, see Cumming, 2012, and Kline, 2013). First, NHST addresses whether an effect could have been observed on the basis of chance alone, but it ultimately does not speak to whether that effect has
practical significance. For example, if we compare SAT Math scores for one million West Coast high school students and one million East Coast high school students, even negligible differences in the mean scores might be statistically significant simply because of the impressive sample size. More information would be needed to determine whether this difference was meaningful. Second, the dichotomous language of NHST promotes unrealistic and simplistic dichotomous thinking styles. Many aspects of human behavior are too complex to be boiled down to “True or Not True” conclusions, and yet the language of NHST subtly communicates fallacious and absolutist approaches to evaluating claims. Third, NHST has promoted an environment in which “p hacking” is a concern. Although p hacking can take many different forms, consider one fictional example: imagine that researchers have collected data to examine the relationship between playing a violent videogame and aggressive behavior in a chatroom. After calculating the average number of insults written by those who played an aggressive and nonaggressive video game, they saw that the mean scores did differ in the direction that they predicted, but the p value fell just slightly above .05, and therefore into a range that many journals would consider unpublishable. They discovered, however, that if they eliminated the outliers who simply did not write much at all in the chatroom, the p value dropped down below the cut-off level, so they published the article without indicating that this was a post-hoc decision. This and other forms of p hacking, in which statistics are manipulated to lower the p value, become a greater concern when rigid and dogmatic cut-off levels are used in simplistic ways.

For these and other reasons, many have called for reformations in the use of statistics. For example, Kline (2013) concluded that the “role of significance testing will continue to get smaller and smaller to the point where researchers must defend its use” (p. 25). It was in this context that Trafomow and Marks (2015) banned NHST from their journal.

**Other Forms of Information**

The concerns raised about NHST are legitimate, and it is clear that this approach to statistical analyses has been misused for many years. We believe the kind of information provided by NHST still has an important place in contemporary empirical research. In short, NHST asks the question “If the proposed effect is, in fact, nonexistent, how probable is it that we would have obtained these results by chance alone?” If Fisher’s tea-sipping acquaintance supported her claim simply by showing that she could correctly identify two cups of tea (at a p level of .50), her claim would not be as compelling as it was for eight cups of tea (at a p level of .014), and the rationale behind NHST gives us a means to quantify (or at least estimate) the likelihood of obtaining these results by chance alone. However, this information is incomplete, and other statistical approaches provide answers to different questions.

**Effect Sizes**

According to the Journal Article Reporting Standards (JARS; Cooper, 2010) endorsed by the American Psychological Association, effect sizes should always be presented to accompany any statistical significance tests. Psi Chi Journal follows this recommendation by requiring effect sizes be reported with significance tests. Consider the SAT example provided above. Imagine a fictional situation in which mean scores on the SAT Math test varied between East Coast and West Coast students; one coast yielded a mean of 500.5 (n = 1,000,000; SD = 100) while the other yielded a mean of 499.5 (n = 1,000,000; SD = 100). Although a t value calculated for this data surpasses highly stringent criteria for statistical significance (t = 7.07; df = 1,999,998; p < .0001), a difference of 1.00 on an SAT scale is negligible and would have little practical importance. Therefore, if a researcher simply reported a highly significant difference between the two groups, it would likely lead to misinformation, misunderstanding, and possibly even nurture unrealistic stereotypes. However, effect sizes could be used to communicate that the difference (though highly unlikely to be due to chance factors alone) represents a very small effect.¹

Ellis (2010) explained that there are two general families of effect sizes. The first, the d family, assesses the difference between groups. In the example above, the difference between SAT scores for East and West Coast test takers only represented 1/100th of a standard deviation (d = .01).

¹Although some statisticians have provided benchmarks to represent “small,” “medium,” and “large” effect sizes, many others have urged against treating effect sizes like T-shirt sizes; see Ellis (2010) for a full discussion of this debate. Setting cut-off levels raises the risk that researchers will treat effect sizes in a mindlessly formulaic way, as happened when statistical significance levels of .05 changed from being a common standard to a strict, dogmatic criterion. The focus should be on the practical impact of an effect, rather than how it compares against arbitrarily selected cut-off levels.
Teacups and t Tests | Rouse

a very weak effect despite its statistical significance. Examples of effect sizes in this family are Cohen’s $d$ (which quantifies the effect of a t test), risk ratios, and odds ratios (both of which can be used for 2 x 2 tables to provide effect sizes for $\chi^2$ analyses). The second, the $r$ family, assesses the strength of the relationship between variables. In the example above, the relationship between SAT Math scores and region can be represented as $r = .005$; this is simply an alternative way to present the very weak but statistically significant effect. Examples of effect sizes in this family are $r$ squared ($R^2$), which is used for multiple regression analysis, and several measures used to quantify the effects on ANOVAs: eta squared ($\eta^2$), partial eta squared ($\eta^2_p$), omega squared ($\omega^2$), and Cohen’s f squared ($f^2$).

Among recent articles published in Psi Chi Journal, Stirling, Greskovich, and Johnson (2014) provided an example of best practices in the use of effect sizes. They presented participants with a series of pictures that showed either a snake (which was used as a fear-inducing stimulus) or a salamander (a fear-irrelevant stimulus), and the pictures were either visually clear or blurry. They were instructed to push one key on a keyboard when they saw a snake and a different key when they saw a salamander. Errors occurred more frequently when participants were shown salamanders than when they were shown snakes ($\eta^2 = .57$), and more frequently for blurry pictures than for clear ones ($\eta^2 = .59$). An interaction was observed ($\eta^2 = .39$), such that participants were most likely to make an error when shown blurry pictures of salamanders and least likely to make an error when shown clear pictures of snakes. These findings were interpreted within an evolutionary context; because the cost of failing to respond to a dangerous stimulus is greater than the cost of responding to an innocuous stimulus, humans might be primed to err in the direction of perceiving threat, especially in ambiguous situations. The use of effect sizes in this study was valuable in showing relatively strong effects that can be compared across the types of effects and the interaction of these effects.

Confidence Intervals

Generally, inferential statistics are based on a sample from a population, but they are used to reach generalized conclusions about the population itself. Therefore, these statistics are best thought of as estimates of the values we would obtain if we studied the entire population. Nevertheless, statistics are often reported as though they are definite, obscuring the reality that estimates generally have some level of error. However, confidence intervals (CIs) more accurately communicate the uncertainty inherent in statistical analyses by showing a range of values within which the population value is most likely to fall. For example, if researchers obtained the mean SAT Math score for a sample of West Coast students and for a sample of East Coast students, a 95% CI could be calculated for each of those sample means. If the mean score for one sample falls within the CI for the other sample, it would be unwise to conclude that the two samples represent different populations—any observed difference in the means might simply be due to sampling error.

Among recent Psi Chi Journal articles, Baumgartner, Bauer, and Bui (2012) published an exemplar of the best practices in the use of CIs. Participants answered a questionnaire regarding discriminatory attitudes about homelessness, along with a battery of attitude, belief, and personality measures, each of which was selected to measure characteristics that were hypothesized as a basis of antihomless discrimination. A multiple regression analysis was used to predict discriminatory attitudes on the basis of these test scores. For the five predictor variables, 95% CIs were calculated around the beta weights. Only Locus of Control [-.14, -.03] had a CI that did not include the value 0, while the CIs for Collectivism [-.01, .26], Individualism [-.19, .05], Belief in a Just World [-.24, .02], and Controllability [-.09, .01] all included the value 0. These results provided the greatest level of support for a Locus Hypothesis, that people who make internal attributions for situations are most likely to hold discriminatory attitudes toward homeless people. This study is notable for its effective use of CIs. Although the information gained from the CIs is consistent with the NHST results (in that Locus of Control was the only variable which had a statistically significant beta weight, i.e., $p = .004$), the use of CIs communicates that these beta weights are merely estimates, but that Locus of Control was the variable that was most likely to have a nonzero weight.

Replicated and Reproduced Results

Whether a strict replication (in which the exact procedures from one study are followed for a new sample) or a new study developed to address a previous study’s research question using new methods, the systematic gathering of converging and diverging evidence is an important part of
the scientific process. Even when NHST is used effectively, every study poses the risk of error. For this reason, confidence in a claim is bolstered when several studies all provide support; even more exciting for the development of knowledge is the situation in which some studies support a claim and others do not, opening an opportunity to begin exploring the complexity of the topic. Nevertheless, for many years, replications were rare, in part because some journals refused to publish articles that did not make a unique contribution to the literature, and some journals refused to publish articles that did not attain statistical significance. In recent years, however, renewed attention has been directed toward this important part of the scientific process. This is demonstrated by recent decisions of the American Psychological Association and the Association for Psychological Science to create initiatives that support replication science (Drew, 2013; Novotny, 2014). *Psi Chi Journal* has recently engaged a replication initiative (Edlund, 2016).

As part of this initiative, we highlight replication articles with a footnote and have added a replication subject area in our online submission system to help us identify replication studies and reviewers who are replication experts.

Among recent *Psi Chi Journal* articles, Naylor, Kim, and Pettijohn (2013) provided an exemplar of research that benefitted from using more than one independent study to explore a research question. Examining the relationship between personality, creativity, and mood, they first conducted a small pilot study in which they measured extraversion levels prior to inducing a happy or sad mood state. Then, they asked participants to generate creative solutions to four different problems. The results suggested an interaction effect, in which extroverted people were more likely to generate more solutions when a positive mood was induced than when a negative mood was induced; although the opposite pattern was observed for introverted people, the effect was not as strong. The effect size for the interaction of mood and extraversion levels ($\eta^2 = .37$) was substantially higher than the effect size for extraversion levels ($\eta^2 = .04$) or mood ($\eta^2 = .15$) alone. This smaller pilot study was followed by a larger second study; although they changed the mood induction procedure and added a mood manipulation check (which appeared to be validated by the data), in other ways, the procedure repeated the prior study. The second study generally reproduced the results of the first. Once again, extroverted participants were more likely to produce a greater number of responses following a positive mood induction than a negative mood induction. Again, the opposite pattern was observed for introverted participants, but with a much weaker effect. Once again, the effect size for the interaction of mood and extraversion levels ($\eta^2 = .06$) was higher than the effect size for extraversion levels ($\eta^2 = .04$) or mood ($\eta^2 = .01$) alone. Although both studies obtained statistically significant results on their own, the reporting of two sets of results from two independent data sets leads to greater confidence in the general conclusion reached—that extroverted people may be able to generate more solutions when in a positive mood state than in a negative one, and that the same pattern is not observed among introverted people (and may even be reversed).

**When Nothing Is as Important as Something**

As mentioned before, many journals have had previous policies to refuse any manuscript that did not have statistically significant results. There are many negative consequences of a policy such as this, one of which is that such a policy might prevent publication of a manuscript for which the lack of statistical significance is in itself empirically significant. Imagine a fictional example in which a researcher conducted a study that replicated an influential study. However, in the replication, the main effect was negligible, despite being very strong in the initial research. If published, this manuscript might have the potential to encourage other researchers to begin exploring when the effect is likely to be observed and when it is not. However, if journals maintain simplistic policies that only accept statistically significant results, this inconsistency in findings might not come to light. Or, imagine a different fictional situation in which a researcher conducted a study that would support common-sense assumptions, and yet found that the expected effects were weak or negligible. Again, policies that only allow statistically significant results to be published would prevent the researcher from bringing to light the error in the common-sense assumptions. In some cases, then, a lack of statistical significance might be highly informative. *Psi Chi Journal* accepts manuscripts whose findings are nonsignificant.

However, when one of the main points of a manuscript is to highlight the lack of expected statistical significance, it is important to rule out other possible explanations for the unexpectedly weak effect. It might be that the expected effect simply
was not present, but it also might be the case that flaws in the empirical process or the data prevented the expected effect from being observed. At a minimum, two considerations should be addressed.

Reliability Information
First, when expected effects are not observed, researchers should examine the reliability of the scores included in the analysis. A low level of reliability sets an artificially low ceiling for that variable’s relationship with any other variable. After all, low reliability is a reflection of a high level of measurement error, so if a high level of random error variance is present in a set of scores, those scores are unlikely to systematically relate to any other variable or condition. For this reason, it is important for score reliability to be reported for all studies. Because reliability is a property of a set of scores, not a property of the test itself, it is not sufficient to simply report that the test developer found high levels of reliability. Rather, reliability coefficients should be calculated for the data at hand and reported in the manuscript. If the reliability estimate calculated for the data is sufficiently high, measurement error can be ruled out as a probable explanation for the lack of statistical significance.

*Psi Chi Journal* requires that authors report reliability estimates. Extremely low reliability estimates (below .60) are considered “fatal.” Estimates at .70 or higher are considered adequate. Reliability estimates lower than .70 are sometimes allowed, when authors provide a strong explanation for the inclusion of the scale and address the potential impact of the low reliability estimate on the findings. Authors may also choose to alter scales or use single items or two items (with a strong correlation) as predictors where a scale did not prove adequate.

Power Consideration
Second, when expected effects are not observed, researchers should present information about the statistical power of the analyses. According to JARS, every journal article should provide an explanation for the sample size selected, and this should generally be based on a priori power analysis (Cooper, 2010), but information about a priori power analysis is especially important in the case of nonsignificant statistical effects. Ellis (2010) correctly argued that post-hoc power analyses are inappropriate and misleading. In other words, some researchers are occasionally misadvised to conduct a post-hoc power analysis in order to consider whether the obtained results would have been statistically significant if the sample had been larger. Ellis explained several reasons why a post-hoc analysis such as this reflects a misunderstanding of the concept of power.

*Psi Chi Journal* encourages authors to provide information about power analyses conducted during the planning stages of a study. If a conscientious researcher conducted a power analysis prior to collecting data (a step that should be completed for almost all empirical studies), and if the actual sample size matched the sample size recommended from the power analysis, then the researcher has a solid foundation for ruling out an insufficient sample size as a probable explanation for the lack of statistical significance. If, however, the researcher failed to conduct a power analysis during the planning phase or failed to secure the prescribed sample size, insufficient power cannot be ruled out as a possible explanation. We encourage researchers who are new to power analyses to refer to Murphy, Myors, and Wolach (2014) who have written a very accessible introduction to the topic and have developed a user-friendly online power analysis program to accompany their book.

Conclusion: Best Practices During Times of Change
Empirical psychology is at a point in its history when change is happening in the use of NHST. At this point, it is unclear whether that change will amount to a radical revolution away from NHST or a more measured movement away from formulaic and simplistic use of NHST to a more well-reasoned use of significance tests partnered with other forms of statistics. As the research community is examining and debating the most appropriate use of inferential statistics, we recommend that researchers continue to use the tools of NHST, but to use these tools carefully and effectively, following the best practices provided in Table 1.

In general, *Psi Chi Journal* favors empirical manuscript submissions that approach statistical reporting in a thorough and transparent manner to reflect the highest standards of academic integrity of our profession². Thorough reporting of *p*, effect size, reliability estimates, and power analyses provides evidence of scientific rigor in the approach to answering a question. These multiple indices also provide authors with the structure to clearly

---

²As noted by Brannan (2015), *Psi Chi Journal* recognizes the value of qualitative and mixed-methods research as well. When NHST is used, however, the reporting of statistics must be thorough and transparent.
TABLE 1
Best Practices in the Contemporary Use of Null Hypothesis Significance Tests

Prior to gathering data
- Conduct a power analysis to determine the necessary sample size, based on obtained effect sizes from previous research. See Murphy et al. (2014) for introduction and resources.
- Determine which effect size measures are used most frequently in the literature on the topic you are studying. If you use the same form of effect size, it will make comparisons with prior research clearer.
- Read the published research on the scales you are planning to use select one that is likely to produce highly reliable scores in your study.
- Consider whether you conduct more than one study to examine the research question with multiple independent samples.

During statistical analyses
- Perform a reliability analysis for all scales in your study based on your own data set.
- Calculate effect sizes for every significance test. Even when conducting post-hoc contrasts for an ANOVA, effect sizes should be calculated.

When writing the report
- In the Methods section of the report, explain the rationale for the sample size within the context of the power analysis conducted prior to the study.
- Report reliability estimates obtained with your sample for each scale or test.
- Report actual p values, even when they are not statistically significant.
- Report effect sizes for every p value.
- When different statistical significance tests are being compared with each other, make the comparisons on the basis of higher and lower effect sizes, not higher or lower p values.
- When presenting group means (such as in the context of a t test or ANOVA), provide 95% confidence intervals for each group mean or provide a 95% confidence interval for the difference between means.
- When presenting regression analyses, provide R² values, especially when contrasting results from different models.
- Provide 95% confidence intervals for each beta weight in a regression analysis.

articulate the evidence that led to their nuanced and conceptually driven interpretation of data. It is up to all readers of Psi Chi Journal to determine whether they agree or disagree with the authors’ conclusions. Careful and complete reporting of statistical results allows for a critical and independent analysis of the published research.

References

Author Note. Steven V. Rouse, Pepperdine University. Steven V. Rouse is also Associate Editor of Psi Chi Journal of Psychological Research.
Correspondence concerning this article should be addressed to Steven V. Rouse, Social Sciences Division, Pepperdine University, 24255 Pacific Coast Highway, Malibu, CA 90263. E-mail: Steve.Rouse@pepperdine.edu

SUMMER 2016
PSI CHI JOURNAL OF PSYCHOLOGICAL RESEARCH
Journal Submissions

The *Psi Chi Journal of Psychological Research* encourages all Psi Chi members—undergraduate students, graduate students, and faculty—to submit manuscripts for publication. Submissions are accepted for review throughout the year. Although manuscripts are limited to empirical research, they may cover any topical area in the psychological sciences. Replication studies, which are vital for reinforcing the science of psychology, are also welcome.

Submission basics:

- Please send ALL submissions and inquiries through our portal. Manuscripts are peer reviewed, which takes approximately 10–12 weeks.
- Authors’ work is judged in comparison to others at their developmental level. The *Journal’s* Editorial Board are especially dedicated to providing undergraduates with special direction when needed so that the quality of their manuscripts will ultimately match graduate and faculty submissions.
- Once the work is peer reviewed, the primary author will receive e-mail notification (approximately three months after initial submission) for:
  - acceptance,
  - acceptance with minor revisions,
  - the encouragement for major revisions and resubmission, or
  - rejection.
- Accepted manuscripts are generally published within a year after initial submission according to submission dates, revision turnaround time, and at the discretion of the Editor.
- If you have any questions about the submission process, please e-mail the Managing Editor at psichijournal@psychi.org.

What to Submit:

1. A **cover letter** that includes:
   - identifying information for the primary author, including an e-mail address;
   - the primary author’s Psi Chi membership ID number;
   - a description of the primary author’s educational status (e.g., an estimated or actual date of graduation, or description of faculty appointment);
   - a statement that the manuscript is original (not published or accepted for publication elsewhere); and
   - a statement that the research was carried out with approval of an institutional review board and following proper procedures for the protection of human participants or animal subjects.

2. For research papers with an undergraduate listed as the first author, submit a **sponsoring statement** that specifies:
   - the research adhered to APA ethical standards;
   - the mentor has read and critiqued the manuscript on
   - content,
   - method,
   - APA style,
   - grammar, and
   - overall presentation; and
   - the planning, execution, and writing of the manuscript represent primarily the work of the undergraduate student.

3. A **cover page** in APA style (with manuscript title, authors’ names, institutional affiliations, and possibly an author note).

4. A **masked manuscript** following these guidelines:
   - All authors’ identifying information (e.g., name and school) is removed from all sections of the entire manuscript.
   - The manuscript is in Microsoft Word®.
   - The manuscript includes figures, tables, and charts generated in either Microsoft Word or Excel® (in black and white only).
   - Scanned images or illustrations (also in black and white or grayscale only) must have a resolution of at least 600 dpi resolution, and authors must be able to show that they have permission to use all images/illustrations.
   - The manuscript must adhere to APA style.
Journal Submissions

The *Psi Chi Journal of Psychological Research* encourages all Psi Chi members—undergraduate students, graduate students, and faculty—to submit manuscripts for publication. Submissions are accepted for review throughout the year. Although manuscripts are limited to empirical research, they may cover any topical area in the psychological sciences. Replication studies, which are vital for reinforcing the science of psychology, are also welcome.

**Author Checklist:**

Prior to submission, check:

1. Is the primary author a Psi Chi member?
   - The first author of a submitted manuscript must be a member of Psi Chi at the time of submission.
   - Additional authors may include non-Psi Chi members.

2. Is the manuscript empirical?
   - *Empirical* articles include original data collection, secondary data analysis, or meta-analysis.
   - Manuscripts must follow a solid research design, use appropriate research analyses, and report all relevant statistical data including internal consistency and effect sizes for *p* values.
   - Specific values (think =, not > or <) should always be provided regardless of statistical significance, unless the number is < .001.

3. Is the contribution original?
   - Only *original manuscripts* (not published or accepted for publication elsewhere) will be accepted.

4. Is this paper ready to be judged by reviewers?
   - Reviewers will be prompted to evaluate your paper based on the *level of work* expected of undergraduate students, graduate students, or faculty.
   - Authors who completed their research as an undergraduate student may submit it after graduation; however, if those students are enrolled in a graduate program, the work will be compared to that of graduate students rather than other undergraduates.

5. Is the research timely?
   - The Editor reserves the right not to review projects completed more than one year before submission.

6. For faculty authors, is a student coauthor included?
   - Faculty must have *at least one student coauthor*. All authors who submit work they completed as an undergraduate must have a faculty mentor who has reviewed the manuscript and affirmed the work was primarily that of the student(s).

7. Is the manuscript *fewer than 35 pages* including all references, tables, figures, and appendices?

8. Has the manuscript been written according to *APA Style*?
   - All manuscripts must be prepared according to the *Publication Manual of the American Psychological Association* (6th ed.).
   - Refer to the “Checklist for Manuscript Submission” found on APA’s website to check the accuracy of your paper for APA style prior to submission.

9. Does the manuscript reference other *Psi Chi Journal* articles? (We highly encourage this!) Our Publication Search tool at psichi.org/?Publications_Search will help you find interesting and relevant articles.
Online Journal Submission Process

All Psi Chi undergraduates, graduates, and faculty* are invited to submit their research to the *Psi Chi Journal of Psychological Research* through the new web based manuscript submission, tracking, and peer review software solution. Better than email submissions used in the past, this software allows users to create personal accounts to make the submission process more efficient.

Four items are required for all submissions:

1) **Cover Letter**  
Include primary author’s education status, manuscript originality statement, IRB approval

2) **Sponsoring Statement**  
Undergraduate first authors only

3) **Cover Page**  
Author names, school affiliation, and any author note

4) **Masked Manuscript**  
MS Word with all personal information removed

Simply register an account, then click Submit Manuscript:

1) **Upload Files**  
Files can be removed, replaced, or reorganized

2) **Enter Manuscript Information**  
E.g., title, abstract, authors, keywords, etc.

3) **Review Manuscript Material**  
Summary of all information/files submitted

4) **Submit Manuscript**  
Receive email confirmation

New software benefits:

- Allows users to track their manuscripts’ progress
- Inserts multiple files including cover letters, manuscripts, and figures
- Permits users to prioritize files and coauthors
- Checks for mistakes in the submission process and points out any errors
- Streamlines the process for authors and reviewers

* Psi Chi member ID number required

Register an account: http://pcj.msubmit.net/cgi-bin/main.plex
Tutorial videos: http://www.ejpress.com/demos

NOW INDEXED IN psycINFO!