Guidance for Researchers When Using Inclusive Demographic Questions for Surveys: Improved and Updated Questions
Jennifer L. Hughes¹, Abigail A. Camden¹, Tenzin Yangchen², Gabrielle P. A. Smith³, Melanie M. Domenech Rodríguez², Steven V. Rouse⁶, C. Peeper McDonald¹, and Stella Lopez⁷
¹ Department of Psychology, Agnes Scott College
² Department of Psychological Sciences, Auburn University
³ Department of Epidemiology, Brown University
⁴ Department of Psychology, Texas Woman’s University
⁵ Department of Psychology, Utah State University
⁶ Social Sciences Division, Pepperdine University
⁷ Department of Psychology, University of Texas at San Antonio

Bicultural Identity and Social Support Seeking Processes: The Effects of Cultural Priming Among East Asian Americans
Maggie Yao, Delancey C. Wu, and Nancy L. Collins
Department of Psychological & Brain Sciences, University of California, Santa Barbara

Responses to Shame: Influences of Adherence to Masculinity Norms
Zachary Stine, Taylor Broughton¹, Glena Andrews¹, and Chris Spromberg¹
George Fox University, School of Clinical Psychology

Examining Relationships Between Transgender Prejudice, Gender Essentialism, and Defining and Categorizing Transgender People
Bec Stargel and Angela C. Bell¹
Department of Psychology, Lafayette College

The Patterns of Children’s and Caregivers’ Gender-Typed Exhibit Choices in a Pop-Up Children’s Museum
Brandon Garcia, Natassia Aleman-Teweles, and Jennifer Dyer-Seymour¹
Department of Psychology, California State University, Monterey Bay

Mental Health Symptoms Predicting American College Students’ Academic Performance: The Moderating Role of Peer Support
Gabrielle M. Goselin and Nicolette P. Rickert¹
Department of Psychology, Georgia Southern University

The Predictive Ability of Early Maladaptive Schemas for Aggression
Derek A. Lavoie and Elizabeth A. Harwood¹
Department of Psychology, Rivier University
Guidance for Researchers When Using Inclusive Demographic Questions for Surveys: Improved and Updated Questions

Jennifer L. Hughes¹, Abigail A. Camden², Tenzin Yangchen³, Gabrielle P. A. Smith⁴, Melanie M. Domenech Rodríguez⁵, Steven V. Rouse⁶, C. Peeper McDonald¹, and Stella Lopez⁷

¹Department of Psychology, Agnes Scott College
²Department of Psychological Sciences, Auburn University
³Department of Epidemiology, Brown University
⁴Department of Psychology, Texas Woman’s University
⁵Department of Psychology, Utah State University
⁶Social Sciences Division, Pepperdine University
⁷Department of Psychology, University of Texas at San Antonio

ABSTRACT. This article is an updated version of Hughes et al.’s (2016) article, which encouraged authors to think about and update the demographic questions they use in their research surveys. Hughes et al. (2016) wrote the original article because they could not find a comprehensive resource that gave researchers examples of well-written and representative demographic questions based on the research literature. Since that original article, new and revised terminology related to demographics has emerged and scholarship on equity, diversity, and inclusion has flourished, so the need to present a set of updated demographic questions arose. Based on the recommendation from the APA Journal Article Reporting Standards, Appelbaum et al. (2018) recommended that researchers report the following major demographic characteristics for their samples, which are presented in this article (i.e., age, gender identity and sex assigned at birth, ethnicity and race, and socioeconomic status and social class). They also suggested that researchers assess other demographic characteristics that are important to their specific research, so the most common additional options from the psychological literature (i.e., children, citizenship and immigration status, disability, education, employment, income, language, location, relationship status, religion, and sexual orientation) are included as well. For each of these demographic domains, both questions and more inclusive answer choices are presented. This work is important because it can help researchers to gather and present more accurate information about survey participants’ identities and demonstrates that researchers value inclusion and diversity when conducting their research (Hughes et al., 2016).

Keywords: demographics, demographic questions, survey, questionnaire, inclusion
we relied on our own expertise for the different demographic categories, and we also extensively reviewed the research literature and typical practices for each of the categories. In addition, we pilot tested earlier versions of the questions listed in this article and received valuable feedback from researchers who study these areas, as well as those who are not involved in academia. That feedback led to revising the questions and response options.

As mentioned in the original article, we have only recently started to see a shift in how some governmental agencies and researchers are presenting demographic questions (Hughes et al., 2016). Historically, they asked questions with what now seems to be simplistic categories that only represented majority groups of individuals (Hughes et al., 2016; Pew Research Center, 2022b). However, there is much improvement that still needs to happen. By using consistent and inclusive options for demographic questions, both research participants and those reading about research will be able to see themselves represented (Betz, 2020). Asking participants to select options that do not represent them can lead to frustration, marginalization, and also not knowing how to respond (Hughes et al., 2016). This frustration can decrease the number of participants willing to answer certain questions or even take surveys (Hughes et al., 2016; Tourangeau & Yan, 2007).

The American Psychological Association Journal Article Reporting Standards (JARS; APA, 2021b; Applebaum et al., 2018) are designed to increase transparency and the scientific rigor of journal articles. These standards suggested that researchers should report the major demographic characteristics (i.e., age, sex, ethnicity, and socioeconomic status, SES) of their samples (Applebaum et al., 2018). In addition, they suggested that researchers should include additional characteristics that are important to their specific research.

We believe at a minimum that all research articles should include information about age, gender identity and sex assigned at birth, ethnicity and race, and SES and social class. Depending on what the researchers are examining, while taking into consideration the length of their survey and respondent burden, we suggest using some of these other common categories when applicable: children, citizenship and immigration status, disability, education, employment, income, language, location, relationship status, religion, and sexual orientation. We also acknowledge that there are other additional demographic characteristics (e.g., political party affiliation, military veterans or active service) that researchers may choose to ask depending on the research topic, but in this article, we have focused on those most widely utilized in psychological research.

Researchers should be intentional when selecting demographic questions (Alchemer, 2021) to ensure that each question is mapped to their larger survey goals and that there is a plan for how to analyze the data. This means being thoughtful about the selection of the demographic questions and only using ones that are relevant to the research being conducted.

In addition, researchers will want to be mindful of the following issues when using demographic questions in surveys. First, in our original article, we discussed the reasons why researchers collect demographic information (Hughes et al., 2016). One reason is to answer research questions about identity, and another reason is to accurately describe the sample of participants. By describing the sample in detail, researchers can determine if the sample they recruited represented the population they wanted to study. In addition, describing the sample’s demographic characteristics helps readers to understand the sample better, to tell if the findings are generalizable, and to compare the sample to other studies. This also can help when it comes to replicating the research.

Second, we discussed in our original article the factors that can influence where demographic questions are placed in surveys (cf. Hughes et al., 2016). Our conclusion was that one set rule at the beginning or end of the survey should not be used and researchers should consider the types of questions they will be using and how their participants might respond to those questions.

Third, question wording is another important issue in that the choice of words and phrases indicates the meaning and intent of the question to the survey participants and wording choices can affect how participants answer the questions (Pew Research Center, 2022b). If researchers do not use established questions like the ones we present in this article, we suggest they pilot test their questions with a diverse group of individuals to receive feedback on the wording and phrases used.

Fourth, researchers should decide when they will use open-ended (i.e., participants provide a response in their own words) vs. closed-ended questions (i.e., participants are asked to choose from a list of possible answers; Pew Research Center, 2022b) for each demographic question. Open-ended questions can be useful as far as keeping the survey length shorter without compromising accuracy or inclusivity (Hughes et al., 2016). In addition, open-ended questions can be better suited for questions where all the possible response options might not be known, or participants might be more comfortable describing their answers in their own language (Cooks-Campbell, 2020). However, closed-ended questions often help with ease of scoring and coding responses. It is important to consider that, for closed-ended questions, participant responding
may be influenced by the response options given, the order in which those response options are presented, and the number of response options (Pew Research Center, 2022b).

Finally, researchers should be sensitive about whether and how they ask for personal data (Sharma & Cowley, 2019). If possible, researchers should consider making their surveys anonymous or at least confidential if they are asking about sensitive demographic information.

**Standard Demographic Questions**

In the following section, we describe the demographic categories of age, gender identity and sex assigned at birth, ethnicity and race, and SES and social class. Based on the JARS guidelines (APA, 2021b), we believe these are essential demographics to report.

**Age**

As noted in the original article by Hughes et al. (2016), age is fairly straightforward to assess, and this is true for most countries except for some East Asian countries where people believe that life begins outside of the womb at one year old (Meinlschmidt & Tegethoff, 2015). If researchers are collecting data using participants from East Asian countries, they should be aware of this.

In the original article, we suggested that researchers use an open-ended question to evaluate age, which allows the researchers to know specific ages of participants and it is easy to calculate a mean age (Hughes et al., 2016; see Figure 1). However, we received feedback from researchers who used this question that more respondents were leaving the question blank as compared to when they used a closed-ended question with response categories. We think this is the case because some respondents thought they could be identified by their exact age, especially if they were older or younger than the typical participant. Researchers should take this into account, and because of this we offer a second option for assessing age. Toor (2020) echoed this and noted that, due to the sensitive nature of age, forcing participants to give a specific number for age is generally discouraged.

A second option includes using a closed-response question (see Figure 2). Researchers who have used closed-response questions have typically divided age into only a few categories (i.e., often five categories), which have ended up with large developmental ranges (cf. U.S. Department of Education, 2009). Large categories such as 60 and older do not feel inclusive in that survey respondents could be in their 70s, 80s, 90s, or even over 100 and might not feel they are developmentally the same as those in their 60s. Instead, we offer nine response categories that represent smaller developmental ranges and another option for those who prefer not to answer.

It should also be noted that APA (2021a) suggested that authors use inclusive age-related language to replace dated terminology. They suggested avoiding the terms: “the elderly,” “elderly people,” “the aged,” “aging dependents,” “seniors,” and “senior citizens.” The terms they suggested using include: “older adults,” “older people,” and “the older population.”

**Gender Identity and Sex Assigned at Birth**

Despite significant strides in recognition of gender and sex as distinct and separate constructs, questions regarding these terms are still often conflated in research (Hughes et al., 2016; Westbrook & Saperstein, 2015). We offer two different demographic questions for asking participants about gender and sex. In contrast to an individual’s sex assigned at birth (APA, 2021; Schusterman Family Philanthropies, 2021), gender refers to an individual’s deeply felt sense of being a woman, man, and/or nonbinary individual, which may or may not align with biological sex or secondary sex characteristics (APA, Divisions 16 and 44, 2015; Hughes et al., 2016). Additionally, gender identity is fluid and can change over time (Hughes et al., 2016; Westbrook & Saperstein, 2015). The question on sex assigned at birth should only be asked if it is essential to address the research questions (e.g., research on ovarian cancer), as asking this question without reason may be offensive to individuals who do not identify with their assigned sex.
Gender Identity
Our question about gender identity was initially proposed by Moody et al. (2013) and endorsed by Hughes et al. (2016). Open-ended demographic questions are inclusive because they convey that the researcher is receptive to using the labels and identities that the participant uses to describe themselves (APA, 2020, 2021) without the researcher biasing the response through set options or the order of options, which have the potential to convey cisgenderism bias. This open-ended question is in line with APAs (2020) bias-free language guidelines, which note that language related to gender has evolved rapidly and could continue to do so. Having an open-ended response allows for future changes in language.

Conveying the fluidity in gender is important to addressing and understanding inclusivity. When writing our question in 2016 we added “currently” to convey that gender can change over time (Hughes et al., 2016; Westbrook & Saperstein, 2015). Some have commented that this qualifier may confuse those who have not questioned their gender. However, because it does not change the integrity of the question, we opted to retain this qualifier, as it conveys that gender is fluid and can change over time. Additionally, we offered additional structure to the original question proposed by Moody et al. (2013) by creating a “please specify” option where participants write their responses. Moreover, the “I prefer not to answer” option is in line with best practices that honor participants’ autonomy over whether they wish to disclose aspects of their identity to the researcher. Similarly, researchers should not force a response (e.g., if using an online survey software such as Qualtrics, researchers should not require a response to move on to the next page of the survey; ORARC, 2020).

In the years since we recommended this question, we have received feedback that the term “gender identity” versus “gender” is confusing to some participants, which may in turn lead to responses that do not reflect the spirit of this question. Some researchers have found more success with this open-ended question when they add examples in parentheses (e.g., woman, man, nonbinary). However, we have also found that some respondents may confuse “e.g.” with “i.e.,” and in turn do not perceive these parentheticals as examples, but rather as forced options, which could make the question much less inclusive than intended. Additionally, by listing only a few options, the researcher may seem to convey a lack of understanding of the gender spectrum. For example, when we piloted this question for the revised edition of this article with a couple examples of gender (e.g., woman, agender), we received feedback that more examples should be added to acknowledge the full spectrum of gender, or conversely that examples should be removed. To create a true open-ended question, we opted to remove examples to avoid bias responding or force set responses and to avoid conveying noninclusivity by listing only a few options. See Figure 3.

Some may be concerned that an open-ended question may create more work for the researcher to code participants’ responses. In our use of this demographic question, we have found that running frequencies on the variable allows for a relatively streamlined process, with some outliers if a participant has misunderstood the question or misspelled their response. Please refer to Hughes et al. (2016) for a coding schema.

Should researchers wish to ask respondents to select a categorical response, they may opt to use a question with response options consistent with Schusterman Family Philanthropies’ (2021) recommendations, with the addition of “gender questioning” and some terms used by Indigenous, Native American, and Native Hawaiian cultures to describe nonbinary genders, including “māhū” (Vanderbilt, 2022), “muxe” (Vanderbilt, 2022), and “two spirit” (Hughes et al., 2016). If it is relevant to the research questions, researchers may also wish to ask participants if they are transgender as an additional question, particularly if choosing not to qualify men and women with cisgender (Schusterman Family Philanthropies, 2021). See Figure 4.

If researchers list response options such as these, they should make several considerations in wording.
and ordering. For example, transgender should not be listed as a gender label by itself, as this conveys that transgender is an adjective and ignores the potential that transgender individuals may also identify as a gender, a man, a woman, gender-fluid, etc. (APA, 2020). Rather, researchers may ask a follow-up question regarding identification as transgender (Schusterman, 2021). Additionally, although some may wish to use “cisgender man” or “cisgender woman” as response options, some respondents are not familiar with these terms, and we would therefore recommend against it to avoid confusion in responding.

Sex Assigned at Birth
For sex assigned at birth, we endorse the demographic question provided by Badgett et al. (2014), to which we have added an intersex response option. Although intersex has only recently been utilized on corrected birth certificates (Segal, 2017), we believe that adding this response option is important to be inclusive of those with a disorder of sexual development. Additionally, as previously noted, researchers should not ask about biological sex in addition to gender unless necessary for their research question. Asking about biological sex, particularly after asking about gender identity, may be offensive to those who do not identify with their biological sex. Moreover, it should be noted the American Medical Association (2021) recently recommended that sex be removed from the public portion of birth certificates. If sex assigned at birth is necessary to the research question (e.g., research on menstruation), researchers may wish to clarify why they are asking about sex assigned at birth in addition to gender identity. The explanation about the necessity of the question regarding sex assigned at birth should be as thorough given the predetermined research parameters. See Figure 5.

Ethnicity and Race
Collecting race and ethnic information continues to be of critical importance for tracking health (Flanagan et al., 2021), academic (Fox et al., 2021), and other systematic disparities in outcomes across groups. Although race is a social construct, the impacts of racism are real and devastating, making the continued measurement of these constructs a necessity. In addition to measures of ethnicity and race possibly being essential to understanding results, post-hoc subgroup analyses in any dataset may also provide valuable insights into questions of interest. In addition to representing good scientific practice, the measurement of ethnicity and race is required in many instances. For example, in clinical trials, the National Institutes of Health requires tracking gender and ethnicity/race to ensure representation in federally funded research (National Institutes of Health, 2017).

Over the years, definitions and measurements of ethnicity and race have varied. Perhaps most notably, this is easily and strikingly captured in a historical review of the U.S. Census racial/ethnic categories (Pew Research Center, 2020). A common critical decision is whether to measure ethnicity and race separately or in combination. In the earlier article, Hughes et al. (2016) measured ethnicity and race by listing ethnic and racial categories in the same item. In this updated article, we recommend separating the categories for increased precision. Another factor that varies in measurement over time is the use of specific ethnic labels. These are in constant evolution, and there is significant variability within groups on their preference for self-identification. Researchers might choose to (a) use federally determined labels, (b) group selected labels (e.g., the national ethnic psychological associations use the panethnic terms: Black, Indian, Latinx, Asian American, and American Arab, Middle Eastern, and North African, in their association names) or (c) labels otherwise known to be preferred in the communities within which they work (U.S. Census Bureau, 2022a).

Disaggregation of Ethnicity and Race
Disaggregating questions on ethnicity and race provide a richer description of samples and can be optimally inclusive. Researchers should be as specific as appropriate while being sensitive to the needs of the target population and acquiring the necessary data for their research (APA, 2021a). If feasible, researchers should consider using open-ended questions to allow for self-identification instead of forcing a response in a predetermined category that may not address the identities of participants (Woolverton & Marks, 2021). For example, there are 574 federally recognized Native American tribes (Indian Affairs, 2022). That number increases when considering state-recognized tribes. Native Americans may or may not have tribal affiliations, and tribal affiliations and self-reports vary (U.S. Census Bureau, 2019). It would be overwhelming to attempt to collect disaggregated data. However, predetermined categories may be necessary for some research. It is vital to balance the preferences of participants (e.g., most
Latinos prefer to identify with national origins rather than the panethnic “Hispanic or Latino” label; Taylor et al., 2012) with research needs (e.g., panethnic labels can help simplify some complexities). In most cases, we provide alternatives (e.g., Black or African American), but in others, the options could be overwhelming (e.g., Hispanic or Latino/Latina/Latinx/Latine) and we selected the more used terms (Noe-Bustamante et al., 2020; Taylor et al., 2012). We also attempted to address this issue by providing the larger group categories and allowing space for respondents to provide more detail if they wish to do so.

We also want to acknowledge that panethnic labels can and will vary based on the population of interest, research question, and many other factors (Noe-Bustamante et al., 2020; U.S. Census Bureau, 2021). Further, although we often use panethnic labels, the individuals within these groups are not monolithic groups and thus have different experiences and label preferences. Therefore, when engaging in this work, we strongly suggest being cognizant of the potential implications of panethnic labels while considering the research’s context, purpose, target audience, and intended population. Also, although we use panethnic labels here, and there will be occasions that it is necessary for research, we want to reiterate that many individuals tend to prefer labels that are specific to their tribe, nationality, or ethnic group and those should be used when feasible (Lê Espiritu, 2019). See Figure 6.

Race
In addition, disaggregating ethnicity and race can help provide more clarity in data. For example, the U.S. Census does not provide an ethnic category for Middle Eastern and North African (MENA) people and instead categorizes them as White (Wang, 2022). Yet, for many, this categorization is inaccurate as their physical appearance and lived experiences may be in sharp contrast with the label White. Indeed, many MENA do not perceive themselves, nor are perceived by others, to be White (Maghbouleh et al., 2022). Also, the conflation of race and ethnicity do not allow researchers to capture the layers of identity participants may have, and “all that apply” options do not provide additional context regarding race and ethnic identification. To address these complexities, we created an ethnicity question and a race question, providing some guidance to respondents regarding the definitions of each construct. We have also added information to point to how participants see themselves and how they are perceived by others. This is part of the complexity of felt, as compared to observed, identity. Research has documented that self-report is more accurate than observer report (Moscou et al., 2003; Polubriaginof et al., 2019) and also that perceived phenotype can impact a person’s experiences in a variety of contexts (see colorism section below). To that end, questions about self-report and how the person is perceived by others could give some rich contextual information.

Some racial groups have significant social ties to a collective racial identity (e.g., Black), that may not always mirror or capture nuances that exist in ethnic or national identities (e.g., Nigerian or African American). These identities also sometimes overlap and intersect with national identity, a person’s sense of belonging to a state or nation, which often include numerous overlapping social identities (Ashmore et al., 2001). Therefore, if researchers are interested in specific aspects of identity, collection of specific ethnic or national identity may be warranted. The collection of national or specific ethnic identities may be especially significant in research spanning multiple nations, or nations consisting of a population originally from a multitude of countries. Researchers engaging participants in countries outside of the United States should ensure that categories align with categories representative of their participants while accounting for cultural norms and the legality of questions regarding ethnicity and race (White, 2015, as cited by Hughes et al., 2012).
Improved and Updated Questions | Hughes, Camden, Yangchen, Smith, Rodríguez, Rouse, McDonald, and Lopez

PSYCHOLOGICAL JOURNAL OF RESEARCH

COPYRIGHT 2022 BY PSI CHI, THE INTERNATIONAL HONOR SOCIETY IN PSYCHOLOGY (VOL. 27, NO. 4/ISSN 2325-7342)

Race is generally tied to physical characteristics such as skin tone, facial features, and height, among other characteristics, and the process of ascribing social meaning to those groups.

How do you describe yourself racially?

How do others describe you racially?

FIGURE 7

Colorism

In addition to ethnicity and race, there has been increasing visibility and scholarship on issues related to colorism. Colorism is discrimination based on skin tone, and typically includes preferences for lighter skin (Dixon & Telles, 2017). Skin tone stratification, or undue privileges and opportunities granted due to lighter skin, is evident in numerous social inequities within society (Monk, 2021). Recent scholarship has revealed significant differences in health and social outcomes according to skin tone within ethnic groups (e.g., Capielo Rosario et al., 2021). Although colorism is evident within the experiences of many racial groups, experiences with colorism vary widely between groups and nations (Monk, 2021). Although we do not provide a measure of skin tone, we do encourage researchers for whom race and ethnicity demographics are central to consider also adding a skin tone measure in their demographics if relevant to the research (e.g., research on skin lightening). There are several measures of colorism that researchers might explore including those reviewed by Dixon and Telles (2017), Capielo Rosario et al.'s (2021) revised version of the Felix vonLuschan Skin Color Chart, and Massey and Martin's (2003) popular NIS Skin Color Scale.

Socioeconomic Status and Social Class

According to the APA (2021a), “socioeconomic status (SES) encompasses quality-of-life attributes and opportunities afforded to people within society and is a consistent predictor of a vast array of psychological outcomes” (p. 17). It has been assessed by evaluating income, levels of educational attainment, and occupational prestige (Deutsch, 2017; Diemer et al., 2013). APA (2021a) also includes perceptions of social class when measuring SES. All these components of SES are presented as demographic questions in this article, and we present ways to assess them. See APA (2015) and Diemer et al. (2013) for more information about how to use these variables to measure SES. Even though it is important for researchers to evaluate SES, it has often not been assessed in survey research because of the lack of conceptual clarity and measurement issues (Diemer et al., 2013; Fernandez et al., 2016; Hughes et al., 2016). Measurement issues involve the fact that SES consists of multiple factors, which often can be labor intensive to evaluate (National Committee on Vital and Health Statistics, 2012; Reeves et al., 2018). In addition, Diemer et al. (2013) advised to carefully select the components to assess, because even though they are related, they each measure separate parts of SES and should not be viewed as interchangeable (Hughes et al., 2016).

Researchers often assess social class instead of SES. Deutsch (2017) defined social class as an individual’s position within society that goes beyond SES and can be thought of as a culture that involves group membership, norms, and socialization patterns. Diemer et al. (2013) noted that “the scales generally include a person’s judgment—based on his or her personal/human capital (occupational prestige, income), social capital (access to socially desirable information), and cultural capital (what he or she knows)—of where they stand relative to others in society” (p. 104). In addition, the effects of social class intersect with other social categories, such as gender, race, and ethnicity (Diemer et al., 2013).

To assess social class, we again present a similar question from our original article, but we increased the number of response options using Stevens’ (2018) options and repeated the question for participants so they could evaluate social class from the perspective of their childhood and their current situation (Hughes et al., 2016). The question in our original article was based on the Social Class Worldwide Model developed by Liu et al. (2004). See Figures 9 and 10.

FIGURE 8

When thinking about physical attributes usually ascribed to race, which of the following general labels describe how you would describe yourself racially: (mark ALL that apply)

☐ Asian
☐ Black
☐ Indigenous, Aboriginal, or First Nations
☐ Latino or Hispanic
☐ Middle Eastern
☐ White
☐ Other, please specify: ___________________
☐ I prefer not to answer

When thinking about physical attributes usually ascribed to race, which of the following general labels describe how others would describe you racially: (mark ALL that apply)

☐ Asian
☐ Black
☐ Indigenous, Aboriginal, or First Nations
☐ Latino or Hispanic
☐ Middle Eastern
☐ White
☐ Other, please specify: ___________________
☐ I prefer not to answer

al., 2016). We provide additional questions about race and give more options for researchers to select from. See Figures 7 and 8.
Reeves et al. (2018) cautioned that measuring social class can produce inconsistent results. Using the wording for three large national surveys (i.e., Pew, GSS/NORC, and Gallup) Reeves et al. (2018) demonstrated that the names and number of social class categories influenced responses and that most respondents put themselves in the middle class. We also believe that using labels (i.e., working class, lower middle class) can be harmful for some people. There are other ways to evaluate social class without labels (i.e., The MacArthur Scale of Subjective Social Status by Adler & Stewart, 2007), which we presented in the original article (Hughes et al., 2016). However, in pilot testing of our new possible demographic questions, we found that respondents asked for anchors to respond to when selecting their social class.

Furthermore, APA (2021a) identified terms regarding social class to avoid, including “the poor,” “low-class people,” and “poor people.” Alternative phrases were suggested, such as “people whose incomes are below the federal poverty threshold” and “people whose self-reported incomes were in the lowest income bracket” (p. 17). Additionally, the Diversity/Inclusivity Style Guide of The California State University (2020) recommended considering the term “underresourced” when referring to groups who have been underserved and underrepresented by the system at large to convey the responsibility of this system.

**Additional Demographic Questions**

In the following section, we describe the additional demographic categories. Researchers will want to use these if they apply to the research they are conducting, while again considering participant response burden. These include children, citizenship and immigration status, disability, education, employment, income, language, location, relationship status, religion, and sexual orientation. Example questions and background information about those questions are given.

**Children**

Participants’ responses to surveys can be influenced by whether the participants have children (Toor, 2020). However, there has been little consistency with the way questions about children have been written (Hughes, 2013; Hughes et al., 2016), and many of the surveys do not have response options that fit all respondents.

In the original article, Hughes et al. (2016) based their recommendations for writing questions about children on the work of Hughes (2013) and Lee and Duxbury (1998) who recommended asking about the number of children, their ages, and if they live in the household. Two questions were written, and those questions have been revised. The first question recognizes more types of parenting roles and includes additional options for types of children. See Figure 11.

The second question about children uses a grid and should only be asked of those who select options other than “no children,” “unborn children,” or “I prefer not to answer” (Hughes et al., 2016). It asks about the ages of the children using age ranges, which were expanded for this revision and include newborn, infant, and toddler. We also kept the option for adult children based on the recommendation of Lee and Duxbury (1998; Hughes, 2013; Hughes et al., 2016). We continue to think that this response option is important to include because 52% of young adult children (i.e., ages 19–29) currently live with one or both of their parents (Fry et
al., 2020). In addition, the question asks if the children live at home full-time or part-time, not at all, or are no longer living (U.S. Census Bureau, 2021). The addition of “they are no longer living” allows those who have lost a child or children to also be represented in the response options. See Figure 12.

If researchers wanted to know more specific information about the respondents’ children, they could have the participants write the gender identity of the children in the boxes instead of numbers. This was not included in our question.

Citizenship and Immigration Status

For U.S.-based surveys, researchers may want to gather information about participants’ citizenship status. According to the Pew Research Center (Budiman, 2020), 13.7% of the U.S. population is composed of immigrants. Immigrants may be authorized or unauthorized, temporarily or permanently. More permanent statuses (e.g., naturalized citizenship) provide immigrants with greater legal protections and civil rights.

This question could be sensitive and therefore it is not necessarily recommended as a standard demographic question. For example, research related to the citizenship question on the census found that those holding temporary or unauthorized statuses and Latinx tend to underreport on this question (Baum et al., 2019; Brown et al., 2019). Hence, we recommend that this question is asked in a trauma-informed way, reminding participants of confidentiality, anonymity, and the voluntary nature of the survey. For example: “Remember that your answers are not connected to you, survey responses will only be shared in aggregate form, and you may skip any questions you are uncomfortable answering.”

The wording of this question is critically important. We recommend using the terms “documented or undocumented” or “authorized or unauthorized.” The term “illegal” tends to describe an act, not a person (Colford, 2013) and its use is considered dehumanizing. Indeed, the use of humanizing language was implicated in the recent federal change to languages used in official documentation referring to undocumented or unauthorized immigrants (Rose, 2021). In addition, immigration status is fluid, and the term illegal might inaccurately capture an infraction that is relative to policy changes. An immigrant who entered a country lawfully might have overstayed their visa and have documentation but be currently unauthorized to reside there. Some recommend the use of the term “unauthorized” as more precise. For example, a refugee or asylum seeker may not have documentation, but may have authorization to be in a country. An immigrant may be unable to obtain documentation because it is prohibited by law in which case the label undocumented is somewhat incomplete (Toobin, 2015).

In recent years, researchers have worked to make visible the differential impacts to health and well-being related to immigration status (Cadenas et al., 2022; Garcini et al., 2022; How et al., 2021; Moreno et al., 2021; Venta et al., 2022), especially as public policies change (e.g., the Supreme Court DACA rescission case in 2020; Department of Homeland Security et al. v Regents of the University of California et al., 2020). Part of inclusion in demographics is to capture the diversity in participants that might be otherwise “invisible.” In a sample of U.S. participants, not reporting on citizenship status would invite an assumption that all participants are U.S. citizens. This simple question might help provide more nuance to the descriptive information provided about the sample. See Figure 13.

Disability

Disability is a broad category representing heterogeneous lived experiences, and inclusive demographic questions must recognize this. In asking about disability, person-first language is imperative (APA, 2021). Our recommended question and response options are based on those from the Schusterman Foundation (2021) and Fernandez et al. (2016), the latter based on the National Survey of Student Engagement. We have altered these questions to include important categories (e.g., mobility impairment or physical disability) and disaggregate some categories. We recommend these
updated disability questions over the ones presented in our previous article (BrckaLorenz et al., 2014; Hughes et al., 2017; Moody et al., 2013), as they better recognize the scope of disability identity and aim to address underreporting. See Figure 14.

Historically, disability has been underreported (Fernandez et al., 2016). Several factors can be seen as contributing to this, including the phrasing of demographic questions. Namely, some respondents may be unsure what the researcher is asking if it is not clearly defined, given that disability is broad, and some individuals may not consider themselves to have a disability despite falling into a relevant category. Additionally, disability can be a sensitive topic (Schusterman Foundation, 2021). Phrasing should recognize this and clarify why the researcher is asking this information. Given this, we recommend explaining what is meant by disability and why the researcher is asking, by using the definition from the American with Disabilities Act (i.e., conditions that substantially limit one or more life activities; ADA, 1990). This language is preferable over demographics questions that ask about specific areas of life that are affected by the disability (e.g., those presented on the census; U.S. Census Bureau, 2018), as it allows the respondent to dictate what important areas of life are impacted for them. Moreover, our recommended questions clarify that the researcher is interested in understanding disability status regardless of accommodations or diagnosis, in an aim to combat underreporting of disability. As a follow-up question, if it is important to the research questions, researchers may wish to ask whether the disabilities have been formally diagnosed or require accommodations. See Figure 15.

**Education**

Education is often assessed in surveys, but researchers are not consistent in the ways they ask their questions (Hughes et al., 2016). The original educational attainment question from Hughes et al. (2016) was retained and updated. New questions about college enrollment and online and in-person education were added. Those additional questions could be useful for researchers wanting to know more about their student samples.

**Educational Attainment**

In the original article by Hughes et al. (2016), one question was used to assess education attainment, and it was based on the U.S. Census’ (2010) education question with some additions. For this revision, the new question was based on the American Community Survey (U.S. Census, 2022b), but again some revisions were made. Those include the following.

First, the American Community Survey’s educational attainment question (U.S. Census, 2022b) included specific grades. We did not include those, but instead, we asked if participants had completed elementary school, or middle or junior high. We knew that we would be adding other response options, and this helped to keep the length more manageable.

Second, Mahmutovic (2021) encouraged those writing demographic questions about education to also include options for apprenticeships in order to include those who continue their education but not at colleges...
or universities. Some technical, trade, and vocational schools give students certificates when students complete the programs, so “certificate” was added to that option as well. The response option for “vocational training” was expanded to include the wording “technical, trade, or vocational school certificate or apprenticeship.” For a review of these types of programs see the Center for Employment Training (2022).

Third, when people think of postsecondary education, they often default to bachelor’s degrees, but many people are earning subbaccalaureate credentials (i.e., associate degrees or completing certificates; Hudson, 2018). Those respondents should also feel represented with survey options. The current U.S. Census (2021b) uses a new distinction for associate degrees, including occupational associate degrees (i.e., a degree given for a specific occupation) and academic associate degrees (i.e., a degree typically in the arts and sciences with the work being able to be transferred to a bachelor’s degree). These options were added to the question.

Fourth, we added an option for those who just started college but had not obtained any college credit yet. The American Community Survey’s educational attainment question (U.S. Census, 2022b) did not represent this group.

Fifth, the specialist degree (EdS) was added to the list of possible degrees. Indeed (2021) noted that this degree is an alternative to a doctorate in education, and it typically is a more specialized degree in the areas of school psychology, education leadership and policy, and special education, which does not take as long to complete as a doctorate.

Two additional changes included, an “other, please specify” option and an option for “I prefer not to answer.” This way every possible participant is represented or if respondents did not feel comfortable responding to the question, they would not have to do so. See Figure 16.

College Enrollment
Two additional questions about college enrollment were added that ask respondents if they are part-time or full-time undergraduate or graduate students. Response options were added for those who used to attend college and those who never attended. Although, if researchers think they might have postbaccalaureate or certificate students, they might want to add additional questions for those students. These questions are useful for researchers and reviewers who want to know the percent of college students in samples. See Figures 17 and 18.

Online and In-Person Education
The final question about education asks about the mode of delivery for courses. This is a specialized question, and only some researchers might want to know this information. See Figure 19.
Employment

Survey questions about employment vary considerably (Toor, 2020). For example, some researchers ask about employment status, number of work hours, type of employee, industry, size of organization, years of work experience, etc. (Toor, 2020). In our original article, three questions were given to assess different aspects of employment (Hughes et al., 2016). Those included a question about employment status, including hours worked, type of employer, and type of industry. Those questions were retained but revised. In addition, the option to answer those employment questions about more than one job was added. A more specific separate question about the number of hours worked overall was added and the number of hours worked per job was added as well. Finally, a question asking about occupational titles could be used for those evaluating SES.

Employment Status

The first question asks respondents whether they are employed. As in the original article, the phrase “not employed” was used as an option instead of “unemployed” because the word unemployed has connotations of not being able to find work and in addition some people may not be looking for work (Hughes et al., 2016). In this revised question, “not employed” was broken into “not employed and not looking for work” and “not employed but looking for work.” An option of “retired” was also added, and a new option “I prefer not to answer” was added to every employment question. See Figure 20.

Total Hours Worked Per Week

This question asks about the number of hours employees work per week including multiple jobs, if they have more than one job. The original question included a statement about hours worked at an office, in the field, or at home, and this was retained (Hughes, 2013; Hughes et al., 2016). This is especially important with 25% of all workers in the United States working from home (Robinson, 2022). In addition, the U.S. Census Bureau (2021) noted that those who were on a leave of absence (e.g., family leave, medical leave) might be employed but not have work hours to record. For this reason and to be more inclusive of those in different work situations, a response option representing this was added to the question.

The question asks respondents to use the past week when answering the question. Researchers should be aware that they might want to adjust the time of year for when they collect data (i.e., around holidays). By adding this question that asks about the number of work hours, researchers can differentiate between part-time and full-time employees (Hughes, 2013; Hughes et al., 2016). As noted in Hughes et al. (2016), the U.S. Bureau of Labor Statistics (2022) uses 35 or more hours to define full-time work, so for the question the hours were listed in 5-hour increments in case researchers wanted to divide employees into part-time and full-time employees. However, it should be noted that there is some variation in what is considered to be part-time and full-time (Indeed, 2022). For example, the Affordable Care Act uses 30 hours as the cutoff for full-time work, whereas some employers use 40 hours as the cutoff for full-time work (Indeed, 2022). Finally, the number of work hours given to select from range from under 20 to over 71 to allow for those who might work more than one job or work long hours. See Figure 21.

Hours Worked Per Week

This question is like the question above about total hours worked per week, but respondents would give their hours individually for each job that they worked in the prior week. See Figure 22.

Type of Employee

We based our original question on the American Community Survey’s type of employment question, and we used their recently revised question for our updated question (Hughes et al., 2016; U.S. Census Bureau, 2022b). However, an option for “other type of employee” was included with a place to specify the type. See Figure 23.
If you are NOT on a leave of absence, how many TOTAL hours did you work last week including time at an office, in the field, or working at home? If you have multiple jobs, please add up the hours for all your jobs.

- 71 or more hours
- 66 to 70 hours
- 61 to 65 hours
- 56 to 60 hours
- 51 to 55 hours
- 46 to 50 hours
- 41 to 45 hours
- 36 to 40 hours
- 31 to 35 hours
- 26 to 30 hours
- 21 to 25 hours
- 20 or fewer hours
- On a leave of absence (for example: family leave, medical leave, etc.)
- I prefer not to answer

If you are NOT on a leave of absence, how many hours did you work last week at this job, for all your jobs.

- 71 or more hours
- 66 to 70 hours
- 61 to 65 hours
- 56 to 60 hours
- 51 to 55 hours
- 46 to 50 hours
- 41 to 45 hours
- 36 to 40 hours
- 31 to 35 hours
- 26 to 30 hours
- 21 to 25 hours
- 20 or fewer hours
- On a leave of absence (for example: family leave, medical leave, etc.)
- I prefer not to answer

What type of employee are you? (mark ALL that apply)

- Private Sector Employee
  - For-profit company or organization
  - Non-profit organization including tax-exempt and charitable organizations
- Government Employee
  - Active-duty U.S. Armed Forces or Commissioned Corps
  - Federal government civilian employee
  - Local government (for example: city or county school district)
  - State government (including state colleges/universities)
- Self-Employed or For-Profit Family Business Employee
  - Owner of incorporated business, professional practice, or farm
  - Owner of non-incorporated business, professional practice, or farm
  - Work without pay in a for-profit family business or farm for 15 hours or more per week
- Other
  - Other type of employee, please specify: _______________________
  - I prefer not to answer

Industry

Some researchers recruit participants using people from a specific occupation or industry, which makes the occupational information easier to report but others use a wide variety of occupations (Hughes et al., 2016). When researchers recruit respondents from a wide variety of occupations, it is important for the researchers to consider that respondents may feel uncomfortable giving their specific occupation because they might question whether they will be able to be identified by their response. Using industry instead of having participants list their occupation makes this less likely to happen (Hughes et al., 2016).

In the original article, the U.S. Census Bureau’s (2016) list of industries was used, but some participants reported that their occupation fell into more than one of the industries listed. For this revision, the National Center for O*NET Development’s (2022) list of job industries was selected to use instead because they overlap less. Another advantage for using their list is that it has fewer options making the length of the question more manageable. Again, for this question, an option was added for writing in the industry if the participant did not see their industry represented (Hughes et al., 2016). See Figure 24.

Occupation Title

The National Committee on Vital and Health Statistics (2012) outlined possible steps to evaluate SES. One of those included using survey respondents’ occupational titles to assess prestige. They noted that standard occupation codes can be used to code occupations using existing classification systems. If researchers want to use this to evaluate SES, they should include a free-response question asking participants to list their occupational titles. However, as noted before, some participants might be uncomfortable listing their specific occupational title or titles.

Finally, we did not include temporary or contract workers in our employment questions. If researchers want to know that type of information, they could add an additional question that asks about that.

Income

We did not include income questions in the original article (Hughes et al., 2016). In this revision, we added demographic questions about individual income and family income. We included income because the APA (2015) recommended that it was important to include in surveys to be able to evaluate SES. They also suggested that income should be measured using multiple sources, in addition to wages and salary. They gave examples such as dividends and interest, Social Security, unemployment insurance, and disability income. The American Community Survey’s income questions use eight categories of income to come up with total income (U.S. Census, 2022b). For our questions, we used their eight sources of income, but instead of individually assessing each one of those as separate questions, we condensed their options into one question for both individual and family income.

Besides individual income, some researchers assess either family income or household income. However, the definitions for those differ. The U.S. Census (2021) defines a family as two or more individuals living together, and
How would you describe the industry your job would be in? (mark ALL that apply)

- Accommodation and food services
- Administrative and support services
- Agriculture, forestry, fishing, and hunting
- Arts, entertainment, and recreation
- Construction
- Educational services
- Finance and insurance
- Government
- Health care and social assistance
- Information
- Management of companies and enterprises
- Manufacturing
- Mining, quarrying, and oil and gas extraction
- Professional, scientific, and technical services
- Real estate and rental and leasing
- Retail trade
- Transportation and warehousing
- Utilities
- Wholesale trade
- Other industry, please specify:

I prefer not to answer

What is your total individual income in U.S. dollars for the past 12 months from all the following sources? Please use all the examples below when thinking about your overall income. If you do not know your exact income, please estimate.

- Wages, salary, commissions, bonuses, or tips for all jobs (Report amount before deductions for taxes, bonds, dues, or other items)
- Any public assistance or welfare payments from the state or local welfare office
- Interest, dividends, net rental income, royalty income, or income from estates and trusts (Include even small amounts credited to an account)
- Retirement income, pensions, or survivor or disability income (Include income from a previous employer or union, or any regular withdrawals or distributions from IRA, Roth IRA, 401(k), 403(b), or other accounts specifically designed for retirement. Do NOT include Social Security)
- Self-employed income from own nonfarm businesses or farm businesses including proprietorships and partnerships (Include NET income before business expenses)
- Social security or railroad retirement
- Supplemental security income (SSI)
- Any other sources of income regularly received such as Veterans' (VA) payments, unemployment compensation, child support, or alimony (Do NOT include lump sum payments such as money from an inheritance or the sale of a home)

- No income
- Less than $25,000
- $25,000–$50,000
- $50,000–$100,000
- $100,000–$200,000
- $200,000–$500,000
- More than $500,000
- I prefer not to answer
with changes made to increase inclusivity through adherence to inclusive language guidelines and person-first language (e.g., APA, 2021a; OHSU, 2020). Namely, OHSU (2021) inclusive language guidelines suggest that “first language” is preferable to “English as a second language” to acknowledge that individuals may speak multiple languages. Our recommended question honors this through the plural “language(s),” to acknowledge that individuals may currently use multiple languages interchangeably or have grown up learning multiple languages simultaneously and therefore not be able to identify one first or primary language. Additionally, inclusive language guidelines dictate that language that paternalizes (e.g., “English language learner”) should be avoided (OHSU, 2021).

Therefore, questions about proficiency must be balanced with sensitivity to avoid paternalizing, appearing skeptical, or assuming nonproficiency. The U.S. Census English proficiency question has demonstrated good convergent validity with objective measures of proficiency (Vickstrom et al., 2015), and we rephrased this question to be more consistent with person-first language (APA, 2021a). See Figures 27 and 28.

We recommend presenting the proficiency question for all participants and presenting this prior to the language use question to avoid appearing reactive or skeptical of a possible language deficiency. As a follow-up to understanding which languages participants use, researchers may also wish to understand which, if any, of these the participant considers to be their

### FIGURE 26

Family income is defined as the income of two or more individuals residing together most of the year and who are related to each other by birth, marriage, partnership, or adoption.

What is your total combined family income in U.S. dollars for the past 12 months from all the following sources? Income can come from many sources, please use the examples below when thinking about your overall household income. If you do not know your exact income, please estimate.

- Wages, salary, commissions, bonuses, or tips for all jobs (Report amount before deductions for taxes, bonds, dues, or other items)

**Other Sources**

- Any public assistance or welfare payments from the state or local welfare office
- Interest, dividends, net rental income, royalty income, or income from estates and trusts (Include even small amounts credited to an account)
- Retirement income, pensions, or survivor or disability income (Include income from a previous employer or union, or any regular withdrawals or distributions from IRA, Roth IRA, 401(k), 403(b), or other accounts specifically designed for retirement. Do NOT include Social Security)
- Self-employed income from own nonfarm businesses or farm businesses including proprietorships and partnerships (Include NET Income before business expenses)
- Social security or railroad retirement
- Supplemental security income (SSI)
- Any other sources of income regularly received such as Veterans’ (VA) payments, unemployment compensation, child support, or alimony (Do NOT include lump sum payments such as money from an inheritance or the sale of a home)

- My individual income is the same as my family income
- No income
  - Less than $25,000
  - $25,000–$50,000
  - $50,000–$100,000
  - $100,000–$200,000
  - $200,000–$500,000
  - More than $500,000
- I prefer not to answer

### FIGURE 27

How well do you use English?

- Very well
- Well
- Fair
- Poorly
- Very poorly
- I prefer not to answer

### FIGURE 28

What language(s) do you use fluently or with near fluency? (Mark ALL that apply)

- Arabic
- Bengali
- Cantonese
- English
- French
- German
- Haitian Creole
- Hindi/Hindustani
- Japanese
- Javanese
- Korean
- Malay/Indonesian
- Mandarin
- Polish
- Portuguese
- Punjabi
- Russian
- Signed Language
- Spanish
- Tagalog
- Telugu
- Vietnamese
- Other, please specify: ____________________

### FIGURE 29

Which of these would you consider to be your primary language(s)? (Mark ALL that apply)

- Arabic
- Bengali
- Cantonese
- English
- French
- German
- Haitian Creole
- Hindi/Hindustani
- Japanese
- Javanese
- Korean
- Malay/Indonesian
- Mandarin
- Polish
- Portuguese
- Punjabi
- Russian
- Signed Language
- Spanish
- Tagalog
- Telugu
- Vietnamese
- Other, please specify: ____________________
- I prefer not to answer
primary language. The language options we provide in these demographic questions are those that the U.S. Census Bureau (2015a) reported to be spoken most frequently in the United States and those languages the 2020 Census was offered in (i.e., based on most frequently reported languages from 2012–2016), in addition to those recommended by the ACPA Standards for Demographics Questions (Moody et al., 2013). However, researchers should adjust these language options according to the populations they are sampling. See Figure 29.

Location

Geography is an important context for understanding a person. Indeed, a whole field of behavioral geography (Montello, 2013) examines human behavior in geographical locations. For researchers whose need for examining demographics is less central, a briefer question (see Figure 30) that groups states and territories may make more sense. The groupings are based on the U.S. Census determination of geographic regions. In addition to those, we added an option for territories. We named all territories. When considering inclusion, details such as this make a big difference for the people living in these territories. In addition to possibly feeling invisible, participants may think they are not eligible to participate in a study, or they may leave a geographical location question blank because it is unclear where they belong. In addition to better inclusion, this list provides clarity that would, hopefully, increase probability of responding.

For people seeking a more detailed examination of geography in their variables, we recommend a more disaggregated approach (see Figure 31). Although, we acknowledge that some participants may worry that giving a specific location could help to identify them, especially if the location is different from many of the other participants.

The emergence of COVID-19 caused many to move to remote work, and for some individuals, this may mean that they are residing in a state different from their primary employment. Given this, researchers might want to ask another question about the location of employment. However, this level of detail may be unnecessary, cumbersome, or even increase the probability of making participants more identifiable. Researchers should take these issues into account when choosing location questions. For this additional question, the question stem could be: “I work in the same state or territory that I primarily reside,” with the first response option being “yes,” and the second option being “no, please select the state or territory where you work (mark all that apply)” with the same states and territories listed in Figure 31. Options for “Outside of the United States or United States Territories” and “I prefer
not to answer” would be retained as well.

Depending on the needs of research, geographic designation rather than location may be more relevant. A specific state may be less meaningful than a distinction of rural or urban context. Unfortunately, the definitions of rural and urban have varied over the years (Ratcliffe, n.d.), and the 2020 Census saw more changes (U.S. Department of Commerce, 2022), making precise measures difficult and messy. Rather than concentrate on minute distinctions, a simple item that asks about the general description of the area as urban, suburban, or rural can provide rich context (Parker et al., 2018; see Figure 32).

Finally, researchers will want to be mindful of the risks inherent in survey research that might be distributed well beyond the boundaries under which it was approved. From a regulatory standpoint, Institutional Review Boards’ (IRB) approval resides within the boundaries of the United States. For researchers wanting to collect data on international samples, it is important to know that international research requires local review (Domenech Rodriguez et al., in press). A researcher hoping to collect data in Argentina would have to find a local review board to review and approve the research prior to data collection there. Research approval procedures vary tremendously across countries. A notable example was the passing of the General Data Protection Regulation (GDPR) in 2018. The GDPR has strict standards for data protections for all countries in the European Union (gdpr-info.eu) that easily surpass IRB standards in the United States. We would recommend that, as a minimum, researchers ask whether participants currently live in the United States or one of its territories (see options in Figure 30 and 31). This question would likely be included in a screening questionnaire to ensure compliance with protection standards for human participants in research.

**Relationship Status**

As relationships and living arrangements have changed considerably over the last few decades, the heterosexual intimate/sexual dyad should no longer be assumed for Americans’ relationship structure and composition of families (Seltzer, 2019). Hughes et al. (2016) stated that, up until recently, researchers have only asked about marital status, which was often treated in earlier research as a dichotomous variable between married and not married, the latter being an aggregate category for divorced, separated, widowed, and single. According to Pew Research Center, 38% of U.S. adults ages 25 to 54 were unpartnered in 2019, a steep increase from 29% in 1990 (Fry & Parker, 2022). The substantial rise in the proportion of unmarried people should not be construed as an increase in singlehood, as traditionally perceived. This perspective may be offensive to people who are in partnerships but cannot or do not want to marry (Makadon & Tillery, 2013), and may also lead to inconclusive results and inaccurate views of relationship formations and trajectories. More people in the United States are delaying or eschewing marriage, which may partially explain the general upward trend in cohabitation and divorce rates, shifting traditional meanings of intimate relationships, which are now more individualized (Horowit, et al., 2020).

The most common categories suggested and used for marital status classification are as follows: married, widowed, divorced, separated, and never married (Alchemer, 2021; SurveyMonkey, 2022; U.S. Census Bureau, 2022). DePaulo (2011) noted that it is offensive to list the option of “single” by just using the phrase “never married” because it implies marriage is the end goal for each person. The prompt provided by Alchemer (2021) used “single” because asking what a person’s marital status defaults to that as being the favored option. We included “single, not looking” and “single, looking or casually dating” based on the Pew Research Center's (2022a) classifications. We also have expanded the options beyond those listed above to include “civil union/domestic partnership” and “cohabiting or in a relationship.” A total of 32 countries, including the United States, have accorded legal recognition to same-sex marriages, and an increasing number of Western democracies without marriage equality have recognized civil unions (World Population Review, 2022).

We also included response options for individuals practicing polyamory, a type of consensually nonmonogamous relationship in which people engage in intimate romantic relationships with multiple concurrent partners (Haupert et al., 2017; Moors et al., 2021). The widely held assumption of romantic or sexual exclusivity with one partner as the natural, optimal model for healthy relating that appears in leading theoretical frameworks, such as attachment theory and the investment model of relationships (Conley et al., 2017; Moors et al., 2017), has historically been conventionalized. Although monogamy remains the most prevalent relational paradigm, there is a burgeoning public interest regarding polyamory in mainstream society and popular press (Moors et al., 2017), with one in nine people in the United States having been in a polyamorous relationship (Moors et al., 2021).
For researchers specifically interested in studying intimate relationships, we recommend adding a text box question asking about how people in nonmonogamous relationships identify themselves. Although polyamorous relationships may also take the form of triads, quads, V-structures, etc., they all tend to adhere to the same core ideals of open communication and consent by all parties involved (Balzarini et al., 2019). As a further step toward acquiring a more nuanced understanding of intimate relationships, we added additional options that allow people to indicate whether they are currently in a monogamous relationship, polyamorous relationship, multiple relationships, or not in a relationship, in this updated version of the original article (Hughes et al., 2016). We recommend that researchers ask two sets of questions about current relationship status as a more inclusive approach for capturing the diversity of intimate relationships people engage in, as shown in Figures 33 and 34.

**Religion**

Because the commonality of different religious affiliations varies across geographic regions, it is essential for researchers to be sensitive to the common religious faiths in the cultures from which data are collected. For that reason, the options provided in Figure 35 might not be relevant outside of North America; within North America, however, most respondents would find one or more of these options to be appropriately self-identifying based on recent population-wide religion surveys (Pew Research Center, 2015; Public Religion Research Institute, 2021).

This survey question was created to reflect a few important principles and best practices for inclusive religious self-identification. First, rather than directly asking participants to specify their religion, we suggest a broader question: “With which of the following do you currently identify?” This allows the question to be inclusive of those with theistic faiths, spiritual nontheistic beliefs, and nonspiritual worldviews within the same demographic question.

Second, instead of requiring respondents to only select one self-identification, we encourage allowing respondents to select as many descriptors as they would like. Although religious beliefs are often viewed as mutually exclusive categories, many people have fluid belief systems that incorporate multiple worldviews. Bidwell (2018) noted that spiritually fluid people reported feeling frustrated when asked to constrain their beliefs to a single tradition, and the requirement to select one single worldview was described by some as forcing them to exclude part of their identities. In a recent unpublished study, 5.72% of a large sample of respondents selected multiple belief systems when given the opportunity to do so. By allowing participants to select as many descriptors as are self-identifying, it is more likely that spiritually fluid participants will be able to fully communicate their identities.

Third, note that the options are included alphabetically. Alphabetic presentation prevents the perception of status, importance, or hierarchies among religions.

Fourth, we recommend the inclusion of “spiritual but not religious” (SBNR) as an identity option. As noted by Mercadante (2014), a substantial proportion of
people—especially in the generations born in the 1990s and later—consider their spirituality to be a highly salient part of their identity, and yet many do not ascribe to any formal religious tradition or shared faith. Despite common stereotypes that these individuals simply lack commitment to a particular faith, Mercandante provided evidence that SBNR individuals hold strong convictions that arise from their spirituality, providing their lives with core value systems that are as central to their lives as formal religious doctrines are to the lives of many religiously affiliated people. By providing survey respondents with an SBNR option, researchers allow these participants to communicate the importance of spirituality in their lives without simply being grouped together in a general religiously unaffiliated group.

Fifth, the inclusion of “nothing in particular” is a recommendation that is aligned with the observation that one of the fastest growing religious groups in North America is the subset who do not feel a strong resonance with any religious descriptor: the “nones.” Between 2007 and 2014, the proportion of Americans selecting this option increased from 12% to 16%, making it the fourth largest religious subgroup (i.e., behind Evangelical Protestants, Catholics, and Mainline Protestants; Pew Research Center, 2015). Burge (2021) acknowledged that this increase may, in part, be due to changes in American society that make it more socially acceptable to acknowledge that one does not feel a strong commitment to any faith tradition, but his research suggests that this cannot fully account for the growth. Many nones, he argued, are in a process of moving away from the faith tradition of their childhood but are not driven to replace it with a new belief system. Rouse (2018) found distinct differences among religiously unaffiliated groups, with nones reporting attending church, praying, and reading sacred texts more often than atheists or agnostics. Rouse conducted a survey with “nothing in particular” as an option; two weeks later, he sent those who selected this descriptor a follow-up survey, but this time without providing that option. Although 44% of the nones selected “agnostic” and 19% selected “atheist” when “nothing in particular” was not provided, roughly a third of the nones selected a specific formal religious affiliation in its absence. This suggests that many nones may not identify with other religiously unaffiliated descriptors, so the provision of a “nothing in particular” option enhances the inclusion of this rapidly growing American subset.

Sixth, we encourage the opportunity to allow participants to self-describe their faith or worldview if not listed.

**Sexual Orientation**

Research surveys have generally followed one of three different approaches when asking respondents to self-identify their sexual orientation: (a) dimensional self-ratings, (b) categorical descriptors, and (c) open-ended free response. We recommend that respondents select either a categorical approach (see Figure 36) or free response (see Figure 37).

**Limitations of Dimensional Approaches**

Within a historical context, dimensional approaches had an important place at one time for the promotion of inclusive data collection. The most well-known dimensional scale, the Kinsey Scale, represents a historical move toward greater inclusivity. At a time when sexual orientation was conceptualized by researchers as a simplistic dichotomy between “homosexual” and “heterosexual,” Kinsey et al. (1948) proposed a 7-point dimensional scale with options ranging from 0 (heterosexual) to 6 (homosexual); however, these anchors were often rephrased as “exclusively attracted to the opposite sex” and “exclusively attracted to the same sex”). Although this represented a positive first step in raising researchers’ awareness of the complexity of sexual orientations, it is insufficient for several reasons. Notably, it condenses sexual orientation into one single dimension, typically defined based on attraction.

The Klein Sexual Orientation Grid (KSOG; Klein et al., 1985) was developed specifically to rectify this limitation of the Kinsey Scale, and it is still widely used today. Using a 7-point scale, the KSOG and its variants ask respondents to rate themselves on several independent dimensions, such as rating one’s sexual attraction, sexual behaviors, sexual fantasies, self-identification, social preference, romantic preference, and “lifestyle.” Moreover, the KSOG asks respondents to evaluate each of those dimensions in three reference domains: past, present, and ideal.
There is much to admire about the KSOG’s contributions toward greater inclusivity: for example, it reflects the lived experiences of many sexual minorities in allowing for representation of fluidity across time and complexity across aspects of identity.

However, dimensional approaches such as the Kinsey Scale, KSOG, and variants of these methods have two weaknesses that are detrimental to research inclusivity. First, these approaches tend to conflate the experiences of plurisexual people (e.g., bisexual, pansexual, and omnisexual) with those of asexual people. By conceptualizing sexual orientation as one or more dimensions anchored by (in Kinsey’s original terminology) “heterosexual” and “homosexual,” the middle range might include a very heterogeneous group of people. Galupo et al. (2018) asked a diverse sample of participants to rate whether the Kinsey Scale and KSOG accurately reflected their sexuality; plurisexual people rated the scales less positively than cisgender monosexual people (i.e., including gay men, lesbian women, and straight men and women). Second, these dimensional approaches tend to often reinforce gender binaries. If not anchored by terms such as “heterosexual” or “homosexual,” they tend to be anchored by phrases such as “same sex” and “opposite sex,” which reflects an incorrect conceptualization of sex or gender as a binary characteristic.

If one were to try to be fully inclusive within the dimensional approach, a researcher would have to ask about multiple domains of identity (e.g., attraction, behavior, fantasies, and self-identity), multiple focuses (e.g., toward nonbinary people, toward gender-conforming men, and toward gender-conforming women), and multiple time frames (e.g., past and present). The resulting system would be very time consuming for respondents to complete, and although it might provide insights into a specific respondent’s lived experiences, it would defy any type of summarization for descriptive statistics of a research sample. Therefore, we discourage the use of dimensional self-rating systems of sexual orientation.

**Categorical Descriptors**

Suen et al. (2020) asked focus groups of gender and sexual minorities to reflect on the wording of categorical systems of self-identification, and their responses raise several important best practices in the use of this approach. First, they noted the importance of allowing respondents to select multiple identifiers, such as the phrase “mark all that apply” in Figure 36. Lunn et al. (2016) reported that 16.8% of the respondents in their large sample of sexual and gender minorities selected multiple identifiers when given the option to do so, so requiring respondents to only select one term is unlikely to allow all people to fully represent their identities. Second, Suen et al. (2020) noted the importance of using terms that best resonate with individuals’ identities. For example, many of their respondents found the use of the term “homosexual” to be offensive because of the tie between that term and a history of medicalization and stigmatization. Although the sexual and gender minority respondents in their sample also reported that the phrase “heterosexual” was offputting and problematic, we included the terms “straight/heterosexual” in Figure 36. After all, many straight people think of themselves as “heterosexual,” and our goal was to create a fully inclusive survey in which all people felt as though they could express their own identity in a manner that resonated with their lived experiences. Third, Suen et al. (2020) emphasized the importance of allowing an “other” option. Although most respondents in Lunn et al.’s (2016) sample of 11,476 sexual minority respondents were able to self-identify with one or more of the descriptors listed in Figure 36, a subset of respondents may have a less common term (e.g., androsexual, gynesexual, or skoliosexual) that they consider to be a better representation of their identity. Although not identified by Suen et al. (2020), we recommend a fourth best practice for those using the categorical approach—listing the options in alphabetical order to avoid communicating a hierarchy or supremacy of some identities over others.

**Free Response**

In the focus group conversations reported by Suen et al. (2020), many participants recommended a free response format such as the one in Figure 37. Because this approach allows each person to define their own identities for themselves instead of being constrained by previously selected terms, we recognize that this might be the most inclusive method of asking about sexual orientation. If a researcher opts for this approach, Suen et al. (2020) recommended using the term “sexual orientation” rather than “sexual identity” because (even among a sample of sexual and gender minorities) the term “sexual identity” was unclear to some focus group participants. In pilot data we collected for another project, many respondents misunderstood what we meant when using the phrase “sexual identity” in an open-ended question, providing their gender instead of their sexual orientation. “Sexual orientation” is more familiar to respondents and less likely to be misinterpreted.

Although an open-ended question has obvious benefits for the purpose of inclusive data collection, we offer a note of caution that points to a possible strength of categorical approaches like Figure 36. In pilot data collected for a different project (Hughes & Rouse, 2021), we noted that fewer MTurk respondents self-identified as sexual minorities when asked an open-ended question like Figure 37 than...
when asked a categorical question like Figure 36. When we surveyed the same MTurk workers multiple times, several people who selected a nonstraight identity on a categorical question either skipped or wrote “straight” or “heterosexual” on an open-ended question. We can only speculate about the reason. However, given the level of stigma and self-stigma surrounding sexual orientation, it is possible that a person who is not fully comfortable with a nonstraight identity might find it less threatening to simply click a checkbox than to write out a phrase that they have not fully embraced. Because of the method of recruitment used by Suen et al. (2020), it is likely that most of their participants took pride in their sexual identities, and the opportunity to put their own descriptors in their own terms in an open-ended question was attractive; those who are not at the same level of confidence or comfort in their identities may find a categorical system to be an easier and safer way to self-identify.

Conclusion
Our hope is that this article will help researchers think more critically about the types and formats of demographic questions they could use in their research surveys. As mentioned in the original article, it is important that researchers continuously evaluate the questions they use to make sure that they are using current and inclusive terminology that is representative of people’s identities (Hughes et al., 2016). The questions are formulated, edited, expanded, and revised to reflect the current time. These questions provide the guidance needed to adequately phrase demographic inquiries in studies. Of course, in the future, these will need updating again to reflect societal level changes in how people perceive themselves and others in terms of demographic characteristics. But in the meantime, these questions are designed to assist researchers to collect data that are as accurate as possible. Accuracy here means that what researchers find in the sample of a study really does reflect the population and studies and affect the generalizability of the findings (Hughes et al., 2016). So, these data in turn will impact the conclusions made from the studies and affect the generalizability of the findings (Hughes et al., 2016). More importantly, scientists investigating human behavior have an ethical obligation not only to avoid but also to correct themselves of inaccuracies so as not to harm others (Hughes et al., 2016).

References
Author Note. Jennifer L. Hughes
https://orcid.org/0000-0002-7978-5650
Abigail A. Camden
https://orcid.org/0000-0001-8231-836X
Tenzin Yangchen
https://orcid.org/0000-0002-7913-8568
Gabrielle P. A. Smith
https://orcid.org/0000-0001-5725-5555
Melanie M. Domenech Rodriguez
https://orcid.org/0000-0002-6610-6890
Steven V. Rouse
http://orcid.org/0000-0002-1080-5502
C. Peeper McDonald
https://orcid.org/0000-0003-1979-8959
Stella G. Lopez
https://orcid.org/0000-0003-1669-8485

We have no known conflict of interest to disclose.
Jennifer L. Hughes is the Charles A. Dana Professor of Psychology, Psi Chi Advisor at Agnes Scott College, and an Associate Editor of Psi Chi Journal of Psychological Research. Abigail A. Camden and Tenzin Yangchen are current doctoral students. Gabrielle P. A. Smith is an Associate Professor of Psychology and is serving as the Diversity Director for Psi Chi. Melanie M. Domenech Rodriguez is a Professor of Psychology, served as the President of Psi Chi, and is a former Editor of the Psi Chi Journal of Psychological Research. Steven V. Rouse is a Professor of Psychology and is the current Editor of the Psi Chi Journal of Psychological Research. Peeper McDonald is an Assistant Professor of Psychology. Stella Lopez is an Associate Professor of Psychology and Associate Editor of Psi Chi Journal of Psychological Research.

Correspondence concerning this article should be addressed to Jennifer L. Hughes, Department of Psychology, Agnes Scott College, 141 E. College Ave., Decatur, GA, 30030, United States. Email: jhughes@agnesscott.edu

https://www.census.gov/programs-surveys/cps/technical-documentation/subject-definitions.html#employedpeople
Bicultural Identity and Social Support Seeking Processes: The Effects of Cultural Priming Among East Asian Americans

Maggie Yao, Delancey C. Wu, and Nancy L. Collins
Department of Psychological & Brain Sciences, University of California, Santa Barbara

ABSTRACT. Research has discovered that those from individualistic cultures are more likely to seek social support and view it as effective when coping with stress compared to those from collectivistic cultures. However, many individuals have a bicultural background rooted in both individualistic and collectivistic values, which can enable them to adapt their support seeking behavior to different contexts, depending on which cultural values are salient. The present research evaluated the effects of identity priming on support seeking and support expectations among East Asian Americans. Participants (N = 146) recruited from undergraduate populations were randomly primed with either American (individualistic) or East Asian (collectivistic) cultural icons, followed by questionnaires evaluating their support seeking behaviors and perceptions of support for a current stressor. Generational status and cultural identification were assessed as potential moderators. Contrary to our hypothesis, cultural prime had no impact on support seeking behavior, but did affect perceptions of support. Participants primed with East Asian (vs. American) icons perceived significantly higher overall support (b = –0.42, β = –.21, p = .014) and family support (b = –0.49, β = –.18, p = .038). We also found that those who strongly identified with their American identity reported significantly higher overall perceived support (b_{simple} = –0.71, β_{simple} = –.73, p = .003, ΔR^2 = .030) and friend support (b_{simple} = –0.72, β_{simple} = –.68, p = .005) when primed with East Asian (vs. American) icons. This study considers the within-group differences among East Asian Americans and expands understanding of the relationship between culture and social support.

Keywords: social support, culture, cultural identification, bicultural, priming

When people experience stressful events, they can utilize a wide range of coping strategies. Among these strategies, seeking and receiving social support has been shown to act as a buffer against the adverse mental and physical effects of stress (Cohen & McKay, 2020; Seeman, 1996). Social support is defined as information from others that leads one to feel genuinely cared for, respected, and
included in a social network of reciprocal obligations (Cobb, 1976). A large body of research has shown the significant role of social support in fostering health, well-being, and relationship satisfaction (e.g., Cohen & Syme, 1985; Collins & Feeney, 2000; Uchino, 2008). Nonetheless, a growing body of research has identified important cultural variation in support seeking behavior and perceptions.

One important cultural dimension that can moderate the effect of social support is individualism vs. collectivism (Wu et al., 2021). Individualistic cultures value personal autonomy whereas collectivistic cultures value interdependence (Markus & Kitayama, 1991), and these differences in self-concept influence how individuals from these cultures perceive social support. However, bicultural East Asian Americans are exposed to both individualistic and collectivistic values depending on the objective and subjective context they are in. Although research has examined cross-cultural differences in support seeking behaviors, there is minimal understanding of how identity priming can affect specifically bicultural individuals. Hence, the purpose of this study was to investigate how priming East Asian Americans with their ethnic (i.e., East Asian) or national (i.e., American) identity would affect their support seeking behaviors, with generational status and cultural identification as possible moderators to consider differences in level of cultural exposure.

**Social Support**

When faced with stressors, individuals often confide in members of their social circle for support. Past literature has shown the positive implications of social support on the mental and physical well-being of support seekers (e.g., Cohen & Syme, 1985; Uchino, 2008). For example, emotional and instrumental support from family can significantly reduce caregiver burden, which has been linked to severe health issues (Mashinchi et al., 2021). Some studies have also suggested that social support is effective in aiding recovery from disease, adaptation to chronic illness, and degree of psychiatric disturbance (Cohen & McKay, 2020; Uchino, 2008).

Although the positive effects of social support are well-documented, some studies have shown that there can also be negative consequences of seeking support. Bolger et al. (2000) found that participants who were aware that they received support can suffer an emotional cost, potentially as a result of feeling inequity within the relationship or that they are being pitied for not being able to support themselves (Dunbar et al., 1998). Explicitly seeking support may also negatively impact close relationships, especially when individuals expect that support should be provided without directly asking for it (Mills & Clark, 1982). These findings suggest that, although the benefits of social support seem to be universally applicable (Taylor, 2011), other factors can moderate the effectiveness of social support, with an important one being culture.

**Culture and Relationships**

There are established differences in ideologies and values regarding the self and others between individualistic (e.g., United States of America) and collectivistic cultures (e.g., China, Japan). Individualistic cultures emphasize an independent model of the self that regards individuals as possessing the freedom to make their own decisions and express their beliefs without many social obligations (e.g., Adams & Plaut, 2003; Markus & Kitayama, 1991). This emphasis on autonomy allows individuals to freely choose their relationships and, therefore, face less severe consequences when the harmony of the group is not maintained. Collectivistic cultures, on the other hand, emphasize an interdependent model of the self that considers individuals to be connected and bound to others with a responsibility to prioritize the goals of the group above individual needs (e.g., Kim et al., 2001; Markus & Kitayama, 1991). Relationships in these cultures are assigned rather than voluntarily chosen, meaning there is a great sense of duty associated with them. As a result, those from collectivistic cultures face severe consequences when the harmony of the group is disrupted, thereby hindering their tendencies to self-disclose their personal stressors and conflicts to others (Schug et al., 2010). In short, personal comfort is often prioritized over group harmony within individualistic cultures, whereas the reverse is the case within collectivistic cultures.

**Culture and Social Support**

Past research has discovered cultural differences in the way that social support is sought and perceived (Taylor et al., 2004). Generally, those from individualistic cultures are more likely to seek social support as a coping mechanism and appraise it to be helpful, whereas those from collectivistic cultures are less likely to seek support and tend to view social support as less effective (Kim et al., 2006; Taylor et al., 2004). Other work has shown that Asian Americans are less likely than European Americans to seek social support because they are more concerned about relational complications, such as disturbing group harmony and losing face (Taylor et al., 2004). Overall, individualistic cultures are characterized by a focus on the self and emotional expression, whereas collectivistic cultures emphasize group harmony and emotional inhibition.

Although we might expect that East Asian Americans
would lean more toward collectivistic values and seek less social support, many Asian Americans have a bicultural background rooted in both individualistic and collectivistic values. This dichotomy may place them in a difficult position to regulate contrasting, and sometimes conflicting, goals. In these circumstances, social contextual cues could guide their behavior. Although an abundance of research has examined cross-cultural differences in social support processes, little research has examined within-culture variation and how different identities within the self can play a role in shaping social support behavior and expectations for Asian Americans. Therefore, in the current study, we utilized a priming methodology to investigate if manipulating the saliency of one's American or East Asian identity can influence support seeking behavior and perceptions. Addressing this gap in the literature will help overcome the common limitation in cultural research of over-aggregating Asian Americans into a single category. Grouping all Asian Americans together does not consider important within-group variation of different Asian and American cultures, which ultimately contributes to the misrepresentation and underrepresentation of this population in the literature (Shah & Kandula, 2020). Thus, examining within-group differences among bicultural East Asian Americans using experimental priming can provide a more accurate understanding of how culture shapes social support processes and promote greater cultural representation and equity in research. Another important cultural angle to examine is the effects of acculturation and enculturation.

The Role of Acculturation and Enculturation
When evaluating social support behaviors among Asian Americans, it is important to note that their values could differ depending on the degree of their acculturation to American culture or enculturation to Asian culture. Acculturation is defined as the degree to which people are assimilating to the norms of a dominant culture, while still upholding values of their heritage culture (Redfield et al., 1936). On the other hand, enculturation means the acquisition and preservation of the norms of one's heritage culture (Miller et al., 2011). Generational status refers to one's place of birth and the age at which one immigrates to a new region, and is often used as a proxy measure for acculturation and enculturation levels. For instance, one study found that foreign-born participants reported higher levels of behavioral enculturation (e.g., enjoying Asian language TV), whereas U.S.-born participants reported higher levels of behavioral acculturation (e.g., enjoying English language TV), suggesting that first-generation Asian Americans emphasize the values of their heritage culture more, and thus, may seek less support than later-generation Asian Americans (Miller et al., 2011). Furthermore, there are generational differences in the use of social support when coping with stress, in that Asian nationals and immigrants reported significantly less support seeking behavior than later-generation Asian Americans (Kim et al., 2008; Taylor et al., 2004). Similar results were found in a study about mental health services, where Asian immigrants utilized mental health services at significantly lower rates compared to U.S.-born Asian Americans (Abe-Kim et al., 2007). This line of research suggests that generational status can influence help-seeking behaviors and attitudes toward social support and mental health care, which can be critical in maintaining well-being. Consequently, it is important to investigate how generational status could influence support seeking behavior and moderate the effects of the priming manipulation in the present study.

Although generational status can be a useful proxy for one's level of acculturation and enculturation, cultural identification and cultural orientation may also be useful. Cultural identification can be defined as how much individuals identify strongly (vs. weakly) with their culture. Research has found that people with stronger cultural identification are more likely to endorse and conform to that culture's values and norms (e.g., Jetten et al., 2002; Wan et al., 2007). Specifically, Jetten et al. (2002) found that, within an individualist culture, those who strongly (vs. weakly) identified with their national identity were more individualist in their internalized beliefs and behaviors, whereas the opposite effect occurred within a collectivist culture. Moreover, a study that utilized a similar priming approach as ours concluded that, among bicultural Asian Americans who strongly (vs. weakly) identified as American, priming them with American icons led to assimilative responses, such that their judgments shifted toward the norms of American culture (Zou et al., 2008). This suggests that bicultural individuals’ responses to social cues can depend on their cultural identity motives and how personally connected they feel toward their cultural group. Cultural orientation is defined as minority individuals’ degree of affiliation with their ethnic and majority cultures (Ying, 1995); in the case of our study and based on past research, we wanted to focus on Asian Americans’ individualistic-collectivistic orientation. In summary, because generational status may not tell the whole story of one's personal connection to one's culture, we also investigated the roles of more subjective measures of acculturation (cultural identification and cultural orientation) as exploratory moderators.
Present Study
This study investigated the impact of cultural identity priming on support seeking behavior among East Asian American undergraduates. We utilized priming methods to experimentally manipulate the saliency of one’s national or ethnic identity. We decided to use priming based on past studies that have found that individualistic and collectivistic values can be primed with American and Asian imagery, respectively (Hong et al., 1997, 2000). Further, similar approaches have been found to affect social support processes. Specifically, Kim et al (2006) found that Asian Americans sought more support when primed with self-goals (which were intended to activate aspects of individualism), whereas there were no effects of priming on European Americans. This supports the idea that making individualistic and collectivistic values salient can motivate one to behave according to those values.

We then examined how identity (experimentally primed) shaped help-seeking behaviors in terms of how likely participants would be to seek instrumental and emotional support and how much they perceived it to be available and effective in the face of a stressor. We also examined individual differences that might moderate the effect of priming on support seeking behavior and perceptions. Generational status was assessed as a potential moderator to consider how environmental exposure can influence social behaviors. Cultural identification was another potential moderator, serving as a proxy for enculturation and acculturation levels. Finally, we examined cultural orientation (individualism and collectivism) as exploratory moderators.

We hypothesized that priming East Asian Americans with their Asian identity would make aspects of collectivism salient and lead them to seek and perceive less social support. On the other hand, priming them with their American identity would make aspects of individualism salient and, therefore, prompt them to seek and perceive more social support. Generational status was also expected to predict support seeking behavior and expectations. Later generations, who have acculturated to be more individualistic, should seek and perceive more social support in comparison to earlier generations. In addition, we predicted that generational status would moderate the effect of the primes. Specifically, we predicted that priming would be less effective in later generations because their American identity is more likely to be dominant and cognitively accessible (and their Asian identity is less accessible), making them less impacted by the Asian primes. Therefore, later generations should show less change in their behavior (as a function of the primes) compared to earlier generations. Similar patterns were expected for cultural identification, such that those who identify strongly with their American identity should be more acculturated and, thus, seek and perceive more support in comparison to those who identify strongly with their Asian identity. For the same reasons as generational status, the primes should also be less effective among participants with a stronger American (vs. Asian) identity. Given the history of underrepresentation of this population in the psychology literature (Shah & Kandula, 2020), we sought to conduct vital research that primarily focuses on the behaviors and perspectives of a specific Asian subgroup and how those may vary across individuals, despite their shared cultural background, as well as offer additional insight on generational status and cultural identification as potential moderators.

Method
Participants
A total of 146 undergraduate students (female = 65.1%; age, $M = 20.29$, $SD = 1.68$, range = 18–25) who self-identified as East Asian American took part in the study. We restricted the age range to focus on young adults (undergraduates), as this is a critical developmental period that is characterized by identity formation (e.g., Erikson, 1993), and to reduce noise that may be caused by including participants outside this age range. A sensitivity analysis conducted in G*Power (Faul et al., 2007) for multiple regression [fixed model, $R^2$ increase, $N = 146$, $α = .05$, power = .80, tested predictors = 1, total predictors = 4], indicated that our sample size was well-powered for detecting relatively small effects ($f^2 = .06$).

Only 21.9% of participants were born outside the United States. Most classified themselves as second generation (55.5%) or beyond (17.8%). Participants were Chinese/Chinese American (51.4%), Japanese/Japanese American (14.4%), Korean/Korean American (13.7%), Taiwanese/Taiwanese American (12.3%), Mongolian/Mongolian American (0.7%), and Other (7.5%). See Table 1 for demographic information. An additional 44 participants enrolled in the study but were excluded because they did not meet eligibility requirements (i.e., age, East Asian background, undergraduate status; $n = 16$), failed to complete the study ($n = 24$), or showed a pattern of fabricated responses (e.g., marking “2” on every single scale item; $n = 4$).

Participants were recruited through the SONA undergraduate research pool ($n = 83$) and paid research pool ($n = 63$) at a public university. Participants received 0.5 course credits or a $5 Amazon e-gift card. To further incentivize participation, participants were able to enter a random drawing for two $50 Amazon e-gift cards. Additional recruitment was done through social media platforms (i.e., Facebook, GroupMe, Instagram).
Procedure
After obtaining approval from the University of California, Santa Barbara Institutional Review Board, participants were recruited for an online survey study about identity and social reactions. After providing informed consent, participants were randomly assigned to the American priming condition or the East Asian priming condition and presented with the associated set of cultural icons. Following the priming procedure, participants were asked to describe a current stressor (i.e., health, social, academic, financial, other) they were facing in an open-ended format for two minutes. If they were coping with multiple stressors, we asked them to select the most distressing one. After describing their stressor, their expected coping strategies (instrumental and emotional support seeking) were assessed. Participants then reported their expectations of their friends and family’s willingness to support them, as well as the perceived effectiveness of the support they would receive from friends and family. The final survey measured their cultural orientation. At the end of the study, participants completed demographic questions and were then given a debriefing statement and asked for permission to keep their data for research purposes.

Measures and Materials
Priming Materials
Participants were primed using an adaptation of the priming method from Hong et al. (2000), which exposes participants to a set of cultural icons created to activate the individualistic or collectivistic values. Ten types of icons were used per priming condition (20 total), such as symbols (i.e., the American flag vs. a Chinese dragon) and landmarks (i.e., the Statue of Liberty vs. the Genghis Khan monument). Pilot testing was conducted to ensure that the icons were culturally relevant. Participants were randomly assigned to be primed by either American or East Asian cultural icons. In both conditions, the 10 icons were shown all at once on the same page. After viewing the icons, participants were asked to select the one picture they felt most connected to and describe in an open-ended format how they identified with the picture, why they picked the picture, and/or what the picture made them think of in two minutes. As a manipulation check, they were asked to choose if the icons were representative of American or Asian culture. All participants chose the correct cultural category.

Use of Social Support
To assess differences in social support seeking, participants completed an 8-item questionnaire (adapted from the Brief COPE scale; Carver, 1997), which assessed their use of various coping strategies in response to a stressor. We utilized the adapted scale used by Kim et al. (2006). The primary outcomes of interest included instrumental support (α = .84) and emotional support (α = .89), which were both measured with 4-item subscales. Participants were given the following instructions: “With the stressor you have just described in mind, please rate the following statements based on how likely you will utilize that strategy to cope with the stressor.” Responses were given on a 5-point scale ranging from 1 (not at all) to 5 (very much). Sample items included “I will talk to someone about how I feel” for emotional support and “I will ask people who have had similar experiences what they did” for instrumental support.

Overall Perceived Support
To assess participants’ expectations about their friends’ and family’s willingness to provide support and the effectiveness of that received support, we used a 4-item scale adapted from Kim et al. (2006). Participants reported how much they expected their friends and family to be willing to support them in coping with their stressor (e.g., “With the social stressor you have just described in mind, how much do you expect your friends to be willing to help you in coping with it?”) and how effective that support would be in resolving their stressor (e.g., “With the stressor you have just described in mind, how effective do you expect your family to be in helping you overcome it?”). Responses were given on a 7-point scale from 1 (not at all willing/effective) to 7 (very willing/effective). We computed an overall perceived support index by averaging all four items (α = .64). However, the reliability was somewhat low because the friend items were not consistently correlated with the family items (r’s ranging from .07 to .62). Thus, we also computed two 2-item subscales to measure perceived friend support (α = .69) and family support (α = .72) due to the low reliability of the combined index and past literature demonstrating differences between friend and family support (e.g., Poulin et al., 2012).

Generational Status and Cultural Identification
To assess generational status, participants were asked to identify their generation, with higher numbers indicating that their family has been in the U.S. longer. Although generational status is typically coded as a dichotomous categorical variable, we retained it as a continuous, interval variable to capture the full variance and nuance of exposure to American culture. Seven participants who classified themselves as “Other” were excluded in all analyses involving generational status. See Table 1 for the distribution of these responses.

To measure cultural identification, we asked participants to rate their level of identification with their
American or East Asian culture (i.e., “Thinking about your feelings towards American and East Asian culture, how much do you identify with each?”). Responses were given on a 7-point scale from –3 (mostly identify with East Asian culture) to 3 (mostly identify with American culture), where higher numbers represented identifying more strongly with American culture. Generational status and cultural identification were positively correlated with each other (r = .42, p < .001). As would be expected, participants with later generational status reported greater identification with American (vs. Asian) culture.

**Cultural Orientation**
Cultural orientation was measured as an additional exploratory moderator variable using items adapted from Oyserman et al. (2002) and using the adaptation from Kim et al. (2016). This 14-item scale assessed individuals’ orientation toward individualism (α = .78) or collectivism (α = .73). Participants responded to each item on a scale from 1 (strongly disagree) to 7 (strongly agree). Sample items included “It is better for me to develop my own personal style” for individualism and “Family is more important to me than almost anything else” for collectivism. Individualism and collectivism were not significantly correlated with each other (r = .13, p = .114)

**Results**

**Analysis Plan**
Table 2 shows means, standard deviations, and intercorrelations for all study variables. To test the main hypotheses, we conducted moderated regressions with priming condition as a binary variable (coded 0 = East Asian, 1 = American) and generational status (centered) or cultural identification (centered) as continuous moderator variables. The main effects of priming condition and generational status (or cultural identification) were entered on Step 1, and the two-way interaction was entered on Step 2. We included gender (coded 0 = M, 1 = F) as a covariate at Step 1 because it was unevenly distributed between priming conditions, χ²(1) = 3.85, p = .05, and because there were gender differences on some of the outcome variables (see Table 2). We used the PROCESS macro (Hayes, 2017) in SPSS to follow up on significant interactions by computing simple slopes at 1 SD above and below the mean for generational status (or cultural identification). See Tables 3 and 4 for these results.

In addition to these primary analyses, we conducted a comparable set of regression analysis in which we entered cultural orientation—individualism and collectivism—as exploratory moderator variables in place

---

**TABLE 1**
Demographic Information for the Sample

<table>
<thead>
<tr>
<th>Variable</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
</tr>
<tr>
<td>Female (including cis women and transgender men)</td>
<td>65.1%</td>
</tr>
<tr>
<td>Male (including cis men and transgender men)</td>
<td>32.2%</td>
</tr>
<tr>
<td>Other or Prefer not to answer</td>
<td>2.8%</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
</tr>
<tr>
<td>Chinese/Chinese American</td>
<td>51.4%</td>
</tr>
<tr>
<td>Japanese/Japanese American</td>
<td>14.4%</td>
</tr>
<tr>
<td>Korean/Korean American</td>
<td>13.7%</td>
</tr>
<tr>
<td>Taiwanese/Taiwanese American</td>
<td>12.3%</td>
</tr>
<tr>
<td>Mongolian/Mongolian American</td>
<td>0.7%</td>
</tr>
<tr>
<td>Other</td>
<td>7.5%</td>
</tr>
<tr>
<td>Generational Status</td>
<td></td>
</tr>
<tr>
<td>First generation (born in Asian country, came to U.S. as adult)</td>
<td>10.3%</td>
</tr>
<tr>
<td>1.5 generation (born in Asian country, came to U.S. as child or adolescent)</td>
<td>11.6%</td>
</tr>
<tr>
<td>Second generation (born in U.S., parents born in Asian country)</td>
<td>55.5%</td>
</tr>
<tr>
<td>Third generation (born in U.S., parents born in U.S., grandparents born in an Asian country)</td>
<td>13.7%</td>
</tr>
<tr>
<td>Fourth generation (born in U.S., parents born in U.S. and at least one grandparent born in an Asian country)</td>
<td>2.7%</td>
</tr>
<tr>
<td>Fifth generation (born in U.S., parents and grandparents born in U.S.)</td>
<td>1.4%</td>
</tr>
<tr>
<td>Other</td>
<td>4.8%</td>
</tr>
</tbody>
</table>

Note. N = 146. Other responses: Chinese/Vietnamese American; Mixed (Chinese and White); Cantonese/HK-American; Adopted.

---

**TABLE 2**
Intercorrelations and Descriptive Statistics for Main Study Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>M</th>
<th>SD</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Priming Condition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0 = Asian, 1 = American)</td>
<td>–</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0 = Male, 1 = Female)</td>
<td>–</td>
<td>–</td>
<td>–17**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Generational Status</td>
<td>2.91</td>
<td>0.97</td>
<td>.16</td>
<td>–10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Cultural ID</td>
<td>0.34</td>
<td>1.61</td>
<td>.07</td>
<td>.01</td>
<td>.42***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Overall Support</td>
<td>3.95</td>
<td>0.97</td>
<td>–15*</td>
<td>–10</td>
<td>.09</td>
<td>.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Friend Support</td>
<td>4.17</td>
<td>1.07</td>
<td>–13</td>
<td>.07</td>
<td>.06</td>
<td>.06</td>
<td>.72***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Family Support</td>
<td>3.73</td>
<td>1.39</td>
<td>–11</td>
<td>–19*</td>
<td>.08</td>
<td>.03</td>
<td>.85***</td>
<td>.24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Instrumental</td>
<td>3.39</td>
<td>1.10</td>
<td>–15</td>
<td>.02</td>
<td>.09</td>
<td>.31***</td>
<td>.21</td>
<td>.27</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Emotional</td>
<td>3.64</td>
<td>1.08</td>
<td>–03</td>
<td>.03</td>
<td>.06</td>
<td>.05</td>
<td>.34***</td>
<td>.38**</td>
<td>.18</td>
<td>.60***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Individualism</td>
<td>5.63</td>
<td>0.93</td>
<td>.14</td>
<td>–20*</td>
<td>.03</td>
<td>.07</td>
<td>.01</td>
<td>.05</td>
<td>.02</td>
<td>.19</td>
<td>.19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Collectivism</td>
<td>4.72</td>
<td>0.90</td>
<td>–15</td>
<td>–15</td>
<td>.15</td>
<td>–25**</td>
<td>.31***</td>
<td>.20</td>
<td>.28***</td>
<td>.26</td>
<td>.22**</td>
<td>.13</td>
<td></td>
</tr>
</tbody>
</table>

Note. N = 146. Cultural ID = Cultural Identification scale, higher numbers reflect stronger identification with American culture; Overall Support = Overall Perceived Support scale; Friend Support = Perceived Friend Support subscale; Family Support = Perceived Family Support subscale; Instrumental = Instrumental Support Seeking subscale; Emotional = Emotional Support Seeking subscale. †p < .10, ‡p < .05, ‡‡p < .01, ‡‡‡p < .001.
of generational status and cultural identification. These analyses allowed us to explore whether differences in specific cultural dimensions moderate the impact of the prime on support behavior and perceptions.

### TABLE 3

**Summary of Regression Analyses for Priming Condition and Generational Status**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1 Gender</td>
<td>–0.22 (–.11)</td>
<td>0.18 (.08)</td>
<td>–0.62 (–.21)</td>
<td>–0.37 (–.17)</td>
<td>0.07 (.03)</td>
</tr>
<tr>
<td></td>
<td>[–0.57, 0.13]</td>
<td>[–0.22, 0.57]</td>
<td>[–1.11, –0.13]</td>
<td>[–0.75, 0.01]</td>
<td>[–0.31, 0.46]</td>
</tr>
<tr>
<td>Priming Condition</td>
<td>–0.42 (–.21)</td>
<td>–0.34 (–.16)</td>
<td>–0.49 (–.18)</td>
<td>–0.20 (–.10)</td>
<td>–0.13 (–.06)</td>
</tr>
<tr>
<td>(Asian = 0,</td>
<td>[–0.75, –0.09]</td>
<td>[–0.71, 0.04]</td>
<td>[–0.96, –0.03]</td>
<td>[–0.57, 0.16]</td>
<td>[–0.50, 0.24]</td>
</tr>
<tr>
<td>American = 1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Generational Status</td>
<td>0.11 (11)</td>
<td>0.10 (.09)</td>
<td>0.12 (0.09)</td>
<td>0.02 (0.02)</td>
<td>–0.05 (–0.05)</td>
</tr>
<tr>
<td></td>
<td>[–0.06, 0.28]</td>
<td>[–0.09, 0.29]</td>
<td>[–0.12, 0.36]</td>
<td>[–0.17, 0.20]</td>
<td>[–0.24, 0.14]</td>
</tr>
<tr>
<td>ΔR²</td>
<td>.062*</td>
<td>.034</td>
<td>.079*</td>
<td>.037</td>
<td>.008</td>
</tr>
<tr>
<td>Step 2 Priming Condition ×</td>
<td>–0.20</td>
<td>–0.12</td>
<td>–0.27</td>
<td>0.07</td>
<td>–0.13</td>
</tr>
<tr>
<td>Generational Status</td>
<td>[–0.54, 0.14]</td>
<td>[–0.51, 0.26]</td>
<td>[–0.76, 0.21]</td>
<td>[–0.31, 0.44]</td>
<td>[–0.51, 0.26]</td>
</tr>
<tr>
<td>ΔR²</td>
<td>.009</td>
<td>.003</td>
<td>.009</td>
<td>.001</td>
<td>.003</td>
</tr>
</tbody>
</table>

Note: Unstandardized coefficient (b) listed first, standardized coefficients (β) are shown in parentheses only for main effects and are not shown for interactions, 95% CI for b are shown in brackets. *p < .10, †p < .05.

### TABLE 4

**Summary of Regression Analyses for Priming Condition and Cultural Identification**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1 Gender</td>
<td>–0.20 (–.10)</td>
<td>0.15 (.07)</td>
<td>–0.54 (–.19)</td>
<td>–0.34 (–.15)</td>
<td>0.08 (.03)</td>
</tr>
<tr>
<td></td>
<td>[–0.54, 0.14]</td>
<td>[–0.23, 0.52]</td>
<td>[–0.73, –0.06]</td>
<td>[–0.71, 0.04]</td>
<td>[–0.31, 0.46]</td>
</tr>
<tr>
<td>Priming Condition</td>
<td>–0.36 (–.19)</td>
<td>–0.29 (–.14)</td>
<td>–0.43 (–.16)</td>
<td>–0.16 (–.07)</td>
<td>–0.11 (–.05)</td>
</tr>
<tr>
<td>(Asian = 0,</td>
<td>[–0.68, –0.04]</td>
<td>[–0.65, 0.07]</td>
<td>[–0.89, 0.03]</td>
<td>[–0.52, 0.20]</td>
<td>[–0.48, 0.26]</td>
</tr>
<tr>
<td>American = 1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cultural Identification</td>
<td>&lt;0.01 (–0.1)</td>
<td>–0.04 (0.06)</td>
<td>0.03 (0.04)</td>
<td>0.05 (0.07)</td>
<td>0.02 (0.04)</td>
</tr>
<tr>
<td></td>
<td>[–0.10, 0.09]</td>
<td>[–0.15, 0.07]</td>
<td>[–0.11, 0.17]</td>
<td>[–0.06, 0.16]</td>
<td>[–0.09, 0.14]</td>
</tr>
<tr>
<td>R²</td>
<td>.044</td>
<td>.027</td>
<td>.059*</td>
<td>.032</td>
<td>.005</td>
</tr>
<tr>
<td>Step 2 Priming Condition ×</td>
<td>–0.21*</td>
<td>–0.26*</td>
<td>–0.16</td>
<td>0.15</td>
<td>0.05</td>
</tr>
<tr>
<td>Cultural Identification</td>
<td>[–0.40, –0.01]</td>
<td>[–0.46, –0.04]</td>
<td>[–0.44, 0.13]</td>
<td>[–0.07, 0.37]</td>
<td>[–0.17, 0.28]</td>
</tr>
<tr>
<td>ΔR²</td>
<td>.030*</td>
<td>.038*</td>
<td>.008</td>
<td>.012</td>
<td>.001</td>
</tr>
</tbody>
</table>

Note: Unstandardized coefficient (b) listed first, standardized coefficients (β) are shown in parentheses only for main effects and are not shown for interactions, 95% CI for b are shown in brackets. *p < .10, †p < .05.

### Primary Analyses

#### Priming and Generational Status

**Overall Perceived Support.** At Step 1, there was a significant main effect of priming condition on overall perceived support (b = –0.42, β = –.21, p = .014). Participants who saw the American primes perceived lower overall support from friends and family compared to those who saw the Asian primes. There was no significant main effect of generational status. At Step 2, the interaction between priming condition and generational status was not significant, so the effect of priming condition did not depend on generational status (b = –0.20, β = –.35, p = .25).

**Perceived Friend Support.** At Step 1, there was a marginally significant main effect of priming condition on perceived friend support (b = –0.34, β = –.16, p = .08). On average, participants who saw the American primes reported marginally lower perceived support from friends than those who saw the Asian primes. There was no significant main effect of generational status. At Step 2, the interaction between priming condition and generational status did not add significant variance to explaining perceived friend support.

**Perceived Family Support.** At Step 1, there was a significant main effect of priming condition on perceived family support (b = –0.49, β = –.18, p = .038). Participants primed with American icons perceived less family support compared to those primed with Asian icons. There was no significant main effect of generational status. At Step 2, the interaction between priming condition and generational status was not significant.

**Instrumental and Emotional Support Seeking.** At Step 1, there were no significant main effects of priming condition or generational status on either instrumental or emotional support seeking. At Step 2, the interaction effects were also not significant.

#### Priming and Cultural Identification

**Overall Perceived Support.** At Step 1, there was a significant main effect of priming condition on overall perceived support (b = –0.36, β = –.19, p = .028). On average, participants in the American condition perceived lower support from friends and family, compared to those in the East Asian condition. There was no significant main effect of cultural identification. At Step 2, there was a significant interaction (b = –0.21, p = .037; ΔR² = .030, F(1, 137) = 4.42, p = .037) that explained 3.0% additional variance in overall perceived support. Simple slopes analyses revealed a significant effect of priming condition, but only for those who identified more with their American identity (b_{simple} = –0.71, β_{simple} = –.73, p = .003, 95% CI [–1.16, –0.25]), such that when primed with East Asian (vs. American) icons, they reported significantly higher perceived support (see Figure 1).

Yao, Wu, and Collins
**Perceived Friend Support.** At Step 1, there was no main effect of priming condition or cultural identification on perceived friend support. At Step 2, there was a significant interaction ($b = -0.26, p = .019; \Delta R^2 = .038, F(1, 137) = 5.61, p = .019$) explaining 3.8% additional variance in friend support. Simple slopes analyses displayed a significant effect of priming condition among those who identified more with their American identity, such that they reported significantly higher perceived support when primed with East Asian (vs. American) icons ($b_{\text{simple}} = -0.72, \beta_{\text{simple}} = -.68, p = .005, 95\% \text{ CI} [-1.23, -0.22]$; see Figure 2).

**Perceived Family Support.** At Step 1, there was a marginal main effect of priming condition ($b = -0.43, \beta = -.16, p = .064$). Participants assigned to the American condition perceived lower family support than those assigned to the East Asian condition. There was no significant main effect of cultural identification. At Step 2, the interaction was not significant.

**Instrumental and Emotional Support Seeking.** At Step 1, there were no significant main effects of priming condition or cultural identification on either instrumental or emotional support seeking. At Step 2, the interaction effects were also not significant.

**Secondary Analyses**
In addition to the primary analyses, we explored whether individual differences in cultural orientation—individualism and collectivism—would predict support seeking behavior and perceptions, or moderate the effect of the priming manipulation on these support variables. First, we examined the associations between individualism and collectivism and all dependent variables. As shown in Table 2, individualism was positively associated with instrumental and emotional support seeking behavior, but not with perceived support from friends and family. Collectivism was positively associated with all outcome variables; those who were high (vs. low) in collectivism had more positive expectations of support from friends and family and were more likely to seek instrumental and emotional support. Next, to explore whether cultural orientation moderated the effects of the priming manipulation on our dependent variables, we ran two sets of regression analyses: (a) priming condition and individualism, and (b) priming condition and collectivism. We found significant main effects of collectivism and individualism consistent with the significant correlations described above and marginal interactions with priming condition (see supplemental materials for complete results). These findings show that participants high in collectivism (or those primed with East Asian icons) were more likely to seek support and to hold positive expectations of support from friends and family. The interactions showed a similar pattern, such that among those in the Asian prime condition, participants who were high (vs. low) in both individualism and collectivism marginally perceived higher friend support. These findings further highlight the within-group variability among East Asian Americans.

**Discussion**
In the present study, we evaluated the effects of identity
Bicultural Identity and Social Support Seeking | Yao, Wu, and Collins

priming on East Asian Americans’ social support seeking behaviors and perceptions of support from family and friends. Our goal was to extend prior work showing that people from collectivistic cultures are less likely to seek social support than those from individualistic cultures (Kim et al., 2006; Taylor et al., 2004). However, contrary to this literature and to our hypotheses, results showed that participants primed with East Asian icons perceived significantly higher overall support and family support than those primed with American icons. A similar pattern was found among participants who strongly identified with their American identity. Within our secondary analyses, we found a conceptually similar pattern in that higher levels of collectivism were linked to more support seeking behavior and higher perceived support. Overall, these results differ from prior research showing that those from collectivistic cultures are less inclined to seek support and view it as less effective when coping with stress (Kim et al., 2006, 2008; Taylor et al., 2004). However, one important difference between the current study and prior studies is that we examined variability within a sample of East Asian individuals, whereas prior studies have typically examined variability between Asian and American samples. Our findings suggest that, among those who share East Asian cultural heritage, those who are reminded of their Asian identity, or who hold more collectivistic values, are more likely to perceive and seek support.

Although our main results were unexpected, prior work primarily focused on the act of seeking social support (Kim et al., 2006, 2008; Taylor et al., 2004), rather than perceptions of social support. Despite Asian Americans’ desire for emotional reassurance (Chang, 2015), they are hesitant to seek support due to social strains that come with certain collectivistic values, such as norms of reciprocity (Kim et al., 2001). However, there are other types of support and coping mechanisms that align with collectivistic values. For example, implicit support has been shown to benefit Asian Americans more than European Americans because implicit support conveys support and availability without explicitly drawing attention to the self or burdening the relationship (Taylor et al., 2007; Uchida et al., 2008), and upholding more Chinese values was associated with stronger perceived support availability and can be an effective coping mechanism for Asian Americans (Lee et al., 2012). In the context of this literature, our findings on perceived support complement work that explores more nuanced aspects of social support and highlight the importance of considering the complexities of within-culture variability and how cultural norms guide social behaviors.

We did not find any significant results regarding generational status, possibly because most of our participants were second generation and above, thus our sample did not have adequate variation in cultural exposure. However, our analyses with cultural identification showed that those who self-reported a stronger American identity perceived significantly higher support from close others (particularly friends) when primed with East Asian (vs. American) icons. It was expected that those who strongly identified with American culture would view social support more favorably because of their greater inclination toward individualism, but this was only the case when these participants were exposed to East Asian icons. Although our results do conflict with past research (Zou et al., 2008), these findings may be explained by a combination of our participants’ tendency to lean toward individualistic values of self-directedness, while also being reminded of collectivistic values that are tied to closeness and interdependence (Markus & Kitayama, 1991). Huynh et al. (2011) also concluded that bicultural individuals who rate low in bicultural identity integration (BII) tend to perceive their cultures as polarizing and thus respond to cultural cues by displaying contrasting behaviors (e.g., behaving in American ways after being primed with Asian culture). Given this, it may be that our participants with a stronger American identity also had lower BII between their two identities, leading them to exhibit primereistant behaviors. We also unexpectedly found that both individualism and collectivism were positively correlated with instrumental and emotional support seeking. This could have been due to our items for instrumental and emotional support lacking differentiation between friend and family support.

Although our results did not support our hypotheses, the overall pattern of findings was clear across multiple measures and in experimental and correlational findings. Within our sample of East Asian American participants, East Asian cultural icons and collectivistic values were associated with more support seeking behavior and higher perceived support. At first glance, these findings differ from prior work. However, prior studies have primarily compared Asian participants to European participants and have not yet fully explored variability within Asian American participants. Moreover, past research have demonstrated the complexities within collectivism itself, such as how collectivistic cultures outside of East Asia (e.g., Latinx) embody other forms of collectivism (e.g., convivial collectivism) that are associated with higher emotional expression and support seeking (Campos & Kim, 2017). This could also serve as a potential explanation as to why we found a positive correlation between collectivism and instrumental and emotional support seeking. Thus, in the context of the broader culture and support literature, our findings can be understood as highlighting the importance of
exploring within- and between-culture variation in social support processes, and the impactful role of bicultural identity.

**Limitations and Future Directions**

Several limitations of this study should be noted. First, we relied on a relatively homogenous population of young undergraduate students from East Asian regions. Given that the main objective of our study was to explore within-group differences, we narrowed our population to those who identify as East Asian American. In addition, most of our sample were female, second generation, Chinese American, and all college students. Thus, our findings may not generalize to the diverse Asian American community (e.g., South Asians, older adults, recent immigrants) of different educational or socioeconomic backgrounds, which may hold more traditional or different cultural identities. Future studies should attempt to replicate these findings with larger and more diverse samples of Asian Americans, as well as additional comparison groups from cultures that vary in collectivism and individualism (Campos & Kim, 2017; Wu et al., 2021).

Second, our research design would have benefited from additional control groups. For example, in future studies, it would be helpful to include a no-prime or noncultural prime condition to provide a baseline assessment as a comparison for our East Asian American participants. This could highlight whether our participants' default schemas aligned more with American or Asian culture, as well as improve the internal validity of our priming methodology.

Third, although research has shown that cultural priming activates cultural constructs in memory (Hong et al., 2000), we cannot be certain that cultural values were the only causal factor in producing changes in support perceptions in the present study. The cultural icons we used may also not be representative of the priming that occurs in everyday life, as well as inadvertently prime other constructs beyond culture, such as masculine or feminine gender role norms. Nonetheless, studies have shown that priming is an effective method for activating specific cultural schemas associated with American and East Asian identities (Hong et al., 1997, 2000; Morris & Mok, 2011). Replicating this study with a larger sample size and in a laboratory setting with manipulation checks or pre-tests for other constructs (e.g., perceptions of masculinity and femininity) would be a valuable next step in this work.

Finally, although we did find several statistically significant results, our effect sizes (R2) were modest. One reason for the modest results may be that our priming manipulation was somewhat subtle. In addition, participants completed the priming activity online at home (during the COVID-19 pandemic), where they might have spent less time and effort concentrating on their task than if they had come to the lab. Additionally, our effects might have been stronger if we had been able to recruit more recent immigrants (below second generation), allowing greater variability in bicultural identity and a greater impact of the primes.

**Implications and Conclusions**

With research showing important disparities in the use of mental health care within the Asian American community (Abe-Kim et al., 2007), understanding the underlying motivations and barriers to seeking support may provide insight on ways to foster greater help-seeking behavior. Although social support is not parallel to mental health services, studies have found that poorer perceived quality of support is associated with increased mental health issues (Hefner & Eisenberg, 2009; Wang et al., 2018). Thus, perceiving and receiving higher quality social support may act as a cushion against the detrimental effects of mental illness, which could then help reduce the disproportionate gap between Asian American patients and mental health services. Moreover, Asian Americans’ degree of acculturation has been shown to be positively associated with their attitudes toward seeking professional psychological help (Hamid et al., 2009), perhaps due to having more individualistic values and more favorable attitudes regarding support. Our data adds to this finding by showing that Asian Americans who identified strongly (vs. weakly) with their American identity perceived higher support when reminded of their East Asian (vs. American) heritage, indicating that acculturation can affect perceptions of social support seeking for professional or peer support. Thus, it is pivotal that providers consider acculturation, whether it is a patient’s generational status or cultural identity, when developing their theoretical orientation for their practice. Understanding how bicultural experiences intersect with social behaviors, such as support seeking, can ultimately help clinicians develop a strong therapeutic relationship that is built on mutual trust and cultural sensitivity.

The current research shows that, although all humans have a fundamental need to belong (Baumeister & Leary, 1995), people can perceive relationships differently depending on their cultural background. Although social support may foster well-being (e.g., Cohen & Syme, 1985), the processes of deliberately seeking support can differ depending on cultural context (Kim et al., 2006; Taylor et al., 2004). Unlike previous research that focuses on cross-cultural analyses, the present study considered within-group differences among East Asian
Americans, who are exposed to contrasting cultures, making it valuable to obtain an understanding of how that exposure can influence behaviors. Although our hypotheses were not supported, our results reveal that cultural values play an important and complex role in shaping interpersonal behavior and perceptions. Considering these differences helps increase accurate representation of this group by not assuming that Asian Americans are a homogenous category. We accomplished this by thoroughly examining bicultural East Asian American individuals’ unique experiences and how environmental cues influence their usage and evaluation of social support. Further research exploring these individuals can better inform professionals and researchers on ways to adjust their treatment and research approaches to suit the diverse needs of these populations.

References


misrepresentation and underrepresentation in research. Ethnicity & Disease, 30(3), 513–516. https://doi.org/10.18865/ed.30.3.513


https://doi.org/10.1016/j.jesp.2008.02.001

Author Note

Maggie Yao https://orcid.org/0000-0002-4749-2204
Delaney C. Wu https://orcid.org/0000-0002-6613-3012
Nancy L. Collins https://orcid.org/0000-0003-3670-6745

All materials and data for this study can be found at https://osf.io/fqyb4/. We have no conflict of interest to disclose. This study was supported by the University of California, Santa Barbara Undergraduate Research and Creative Activities Grant.

Positionality Statement: Maggie identifies as a cisgender, bisexual woman. She also identifies as a second-generation Chinese American and first-generation college graduate, and as part of the LGBTQ+ and the Asian American and Pacific Islander communities.

Delaney identifies as a cisgender, heterosexual woman, who is currently a 5th-year PhD candidate studying social support processes in the context of cultural psychology and close relationships. She also identifies as a second-generation, Taiwanese American who grew up in the United States and as part of the Asian American and Pacific Islander community.

Nancy identifies as a third-generation American of European and Jewish ancestry, and as a first-generation college graduate. She also identifies as a cisgender, heterosexual woman. She has studied close relationships and health psychology from a social psychological perspective for almost three decades.

All authors are nondisabled and acknowledge that their perspective is influenced by their position within all of these dimensions of identity.

Correspondence concerning this article should be addressed to Maggie Yao. Email: maggiyao26@gmail.com
Responses to Shame: Influences of Adherence to Masculinity Norms

Zachary Stine, Taylor Broughton, Glena Andrews, and Chris Spromberg
George Fox University, School of Clinical Psychology

ABSTRACT. Managing negative emotions is challenging even without the overlay of socially constructed gender role norms. We were curious about the effects of being confronted with shame on young men who either adhere strongly to traditional masculine gender roles norms or are nonconforming to these norms. We evaluated the shame responses of college men (N = 23), using the Thurston-Craddock Test of Shame (TCTS; Thurston & Craddock, 1998) as our stimulus while measuring physiological and verbal responses. Men in the traditional masculinity norms conforming group scored significantly higher (p = .001, η² = .58) on 7 of the 11 subscales from the Conformity to Masculine Norms Inventory (CMNI; Mahalik et al., 2003) than men in the nonconforming group. When presented with shame-based stimuli, we found differences in heart rate variability depending on the level of overt shame in the stimulus with the two groups, p = .038, η² = .12. The use of aggressive language when confronted with overt examples of shame was not different for our comparison groups. Our results indicate that men who adhere to traditional masculine norms hold different values for interacting with women, physical confrontation, and being in control of situations and others than men who have not internalized the traditional values of American masculinity. Understanding the continuum of values and responses to difficult emotions continues to be a valuable area to study.

Keywords: masculinity, traditional men, shame, RMSSD, heart rate variability

Masculinity is a social construct with a spectrum of expressions from very traditional to nontraditional (Mahalik et al., 2003). Some men strongly conform but some men do not conform to traditional masculine gender-role ideals and norms of American society. These socially constructed gender roles, learned from a person’s culture and upbringing, provide direction and boundaries for normative behavior (Kivel, 2010). Traditional masculinity often emphasized emotional restrictiveness, a drive for success, and stoicism. Whereas nontraditional masculinity viewed emotional expression as less threatening (Reilly et al., 2014). Terms such as “stoic” and “dominant” have formed a narrow, rigid framework for identifying as “a man” in American society (Kivel, 2010), a socially constructed box defining values, beliefs, and behaviors associated with “masculinity”. Genuchi and Valdez (2015) found that men who tended to be more traditional in their expression of masculinity and experience of anger, both trait and state, were more at risk for depression. Men struggled with balancing their own personal experiences and feelings with values of what is held as socially acceptable within traditional masculine gender-role norms. Men holding traditionally masculine values, tended to use distracting coping mechanisms as well as externalizing to manage their anger (Genuchi & Valdez, 2015).

Acknowledging the experience of vulnerability in times of distress goes against male socialization, leading to an internalization of emotions (Reilly et al., 2014). Rice et al. (2018) found that men who exhibited more externalizing behavior were at greater risk for substance use, anger, and risk-taking. Men who used externalization to manage emotions were more likely to have had a recent suicide plan or attempt. Substance use and aggressive behavior tended to interfere with...
positive interpersonal relationship for men (Smith et al., 2018). Men in the United States currently lead women in seven of the 10 leading causes of death and are four times more likely to complete suicide than women (Center for Disease Control and Prevention [CDC], 2019). Men are more likely to hold negative attitudes toward seeking help (Addis & Mahalik, 2003). Men who follow more traditional masculine gender norms pose the greatest risks to their own health and life (Mahalik et al., 2007).

Shame and guilt are often used interchangeably in common vernacular; however, they are not identical. Brown (2013) defined shame as feeling negatively toward oneself and feeling flawed (e.g., I am a bad person), whereas guilt is a person realizing they have done something negative or hurtful (e.g., I did something bad) and can be adaptive and helpful (Brown, 2013). Kindlon and Thompson (2000) suggested that shame is a commonly felt yet unregulated emotion for emerging men. Personal attacks (real or perceived) can result in an aversion to experiencing shame, lowering a man’s ability to negotiate both psychological distress and vulnerable emotions promoting empathy and self-kindness (Sabatino, 1999). This distress leads the masculine identity to value stoicism, to endure difficult emotions without expression, in order to avoid felt or perceived shame. Cycles of distancing oneself from emotion becomes adaptive to alleviate the gender-role strain and policing of masculinity. The resulting internalized shame is a debilitating experience, leaving a man to believe he is defective and unworthy of kindness (Reilly et al., 2014), and has been found to correlate with externalizing behaviors, such as aggression toward women (Dutton & Golant, 1995).

Stanaland and Gaither (2021) determined that men, especially younger men, respond predominantly with aggression when they perceive that their masculinity is threatened. Their research indicated that this was particularly true for men who value the opinions of others and attempt to adhere to strict traditional gender norms. Young men who experienced perceived threats to their masculinity in the form of statements from others were most likely to respond with verbal aggression.

Heart rate variability (HRV) has been linked to self-regulation for cognitive, emotional, and social processing, as well as physical well-being (Laborde et al., 2017). Heart rate variability is the measure of the adaptation of the physiological and emotional system to the changing environment (Bernston et al., 2007). Porges (2007) developed the polyvagal theory and assumed a higher vagal tone was associated with better social functioning. When stressful emotional stimuli, such as shame, are experienced, the autonomic nervous system responds with sympathetic nervous system activation or parasympathetic nervous system inhibition, causing shifts in vagal tone or vagal withdrawal (Bernston et al., 2007; Spromberg, 2019). Cardiac vagal tone represents the part of the parasympathetic system related to heart regulation (Scott & Weems, 2014). Adjusting between the sympathetic and parasympathetic systems to stressful stimuli affected the length of time between consecutive heartbeats, creating variability (Appelhans & Luecken, 2006). Appelhans and Lueken (2006) referred to the increased sympathetic activity or decreased parasympathetic activity as vagal withdrawal. Porges et al. (2007) suggested that vagal tone is a measure of a person’s ability to regulate emotional states. Vagal tone during stress may be a maladaptive response to the situation (Scott & Weems, 2014).

The focus of this research was to explore responses to shame-based stimuli in men by investigating those with traditional masculine values adherence or lack thereof, verbal responses to shame-based stimuli, and utilizing physiological measurements to explore the polyvagal theory of emotional response (Porges et al., 2007). If men who conform to traditional masculine values tend to externalize negative emotional experiences and respond with anger (Genuchi & Valdez, 2015), we expected these men to have higher scores for the Conformity to Masculine Norms Inventory (CMNI; Mahalik et al., 2003)) subscales for Power Over Women, Violence, and Dominance, along with using aggressive language in response to the TCTS cards. If their verbal behavior was consistent with their internalized values, we expected to see HRV, specifically the Root Mean Square of Successive Differences (RMSSD), lower in men who conform to traditional masculine values compared to men who do not conform because they are acting consistent with gender role values. We were interested in using heart rate variability and verbalization of aggression as indicators of conscious and subconscious reactions to negative social situations. Two cards from the Thurston-Craddock Test of Shame (TCTS, Thurston & Craddock, 2009) were selected for this portion of the study; one depicting a woman spanking a child in front of peers (#6), and the second was a male coach yelling at a female player (#3) primarily because the cards depict overt aggression whereas other TCTS cards were more indirect and subtle in the negative emotion they portrayed. Two groups of men (high or low conformity on the CMNI) were the comparison groups.

We hypothesized that (a) conforming men would score higher on several of the subscales of the CMNI, including Power Over Women, Violence, and Dominance. We hypothesized that (b) conforming men would have a higher frequency of aggressive language in response to the TCTS cards selected than nonconforming men. We hypothesized that (c) men high in conformity to traditional masculine norms would have lower RMSSD.
scores in response to the TCTS cards, reflecting their vagal tone or low arousal as a result of holding traditional masculine gender role values and using distracting coping mechanisms (Genuchi & Valdez, 2015).

Method

Participants
Participants were single, undergraduate men (N = 23; 18–29 years of age; 83% White, 17% men of color; Asian/Pacific Islander = 2, Black/African American = 1, Latino = 1). Participants spanned all undergraduate grade levels with the majority of the sample from the first and second years. College men were invited to participate in a study on masculine values. For the initial phase of the study all participants (N = 233) completed the CMNI. Participants in the present study, second phase, were from the top quartile (high conformity to masculine norms, n = 12) or the lowest quartile (low conformity to masculine norms, n = 11).

Materials
The Conformity to Masculine Norms Inventory (Mahalik et al., 2003) is a 94-item rating scale with 11 subscales measuring traditional masculine norms. Seven of the 11 subscales of the CMNI were significantly different between men who were high in conforming to traditional masculine norms and men who scored low (see Table 1). Power Over Women had the largest effect size. Power Over Women indicates the attitude of the participant toward male–female relationship status. Violence and Dominance were the next two highest CMNI subscales. Violence is designed to reflect the tendency to engage in physical confrontations, and Dominance refers to the desire to have personal control of situations (Mahalik et al., 2013). CMNI has internal consistency for men, coefficient alpha of .94 for the total CMNI score. For the Masculinity Norms subscales, alphas ranged from .72 for Pursuit of Status to .91 for Emotional Control.

The Thurston-Cradock Test of Shame (Thurston & Cradock, 2009) is a 10-card projective test to evoke shame-based themes with dichotomous scoring. Rote (2002) evaluated the construct validity of the TCTS using the 16 Personality Factor Adolescent Personality Questionnaire (16 PF-APQ). He found correlations between several of the domains of the 16 PF with scores on the TCTS such as Aggression, Deflation/Withdrawal, and Inflation/Contempt.

Two TCTS cards were the focus of this paper. Card 3 depicted an angry male coach yelling at a female basketball player. Card 6 showed a woman holding a boy over her knee, spanking him in front of peers. Both cards show overt aggression, either physical or verbal. These cards were selected after conducting a Spearman rho with all 10 cards and Power Over Women subscale score. Card 3 was significantly positively correlated with Aggression of the TCTS, r(23) = .50, p = .018, Card 6 was the least correlated with Power Over Women and in a negative direction, r(23) = –.32, p = .14. The Aggression subscale measures whether the participant’s story for the card included aggressive language.

Electrocardiogram (ECG) measures were gathered using Biopac MP150 and Acqknowledge software (n.d.) as participants viewed TCTS stimulus cards via SuperLab (n.d.). The visual stimuli were projected via SuperLab included in the following sequence: an initial screen with instructions for the participant to prepare to rest; a rest screen (image of a mountain) that was timed to 180 seconds; an instruction screen that prompted the participant to think of a story that had a beginning, middle, and end and imagine what each character was thinking and feeling upon seeing the image; the 10 TCTS cards presented in order 1–10; and an ending rest slide. The participant was instructed to view the card and think of the story for 30 seconds. Then the participant was instructed to verbalize his story and it was recorded. Each card was presented through SuperLab for the amount of time needed for each participant to share their projected story. Timing was recorded by the administrator for each card change: 30 seconds of thinking of a response, the start of telling the story, the end of story, and a new card. For this study, we elected to use the Time-domain for heart rate variability (HRV; Laborde et al., 2017). There are a variety of variables available for HRV, including RMSSD. Because we were interested in vagal tone, we selected RMSSD for our dependent variable (Laborde et al., 2017). HRV was gathered with Acqknowledge software and uploaded into

<table>
<thead>
<tr>
<th>TABLE 1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Statistical Information for MANOVA of Conformity to Masculine Norms Inventory Subscales</strong></td>
</tr>
<tr>
<td>Subscale Names</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>Winning</td>
</tr>
<tr>
<td>Emotional Control</td>
</tr>
<tr>
<td>Risk Taking</td>
</tr>
<tr>
<td>Violence</td>
</tr>
<tr>
<td>Power Over Women</td>
</tr>
<tr>
<td>Dominance</td>
</tr>
<tr>
<td>Playboy</td>
</tr>
<tr>
<td>Self-Reliance</td>
</tr>
<tr>
<td>Primacy of Work</td>
</tr>
<tr>
<td>Disdain for Homosexuality</td>
</tr>
<tr>
<td>Pursuit of Status</td>
</tr>
</tbody>
</table>

Note. This table lists the statistical significance information from comparing scores of men in the conforming group to those in the nonconforming group.
Kubios software (n.d.) for conversion into RMSSD (Root Mean Square of Successive Differences) scores (Laborde et al., 2017). Kubios was used with the initial study and continued for continuity of calculations.

**Procedure**

Institutional Review Board approval was granted from the George Fox University committee prior to the start of this study. Male undergraduate students from a private university were invited to participate in a study on masculine values via an online research platform. The original sample data (N = 233) were analyzed for demographic information. Men scoring in the first and fourth quartiles of the CMNI were invited via email to participate in a physiological study in the EEG lab. Participants were unaware that the studies (online and lab) were connected. Participants arrived at the lab where ECG electrodes were placed on the chest. TCTS cards were presented to participants via SuperLab software (n.d.) sequentially. ECG measures were gathered as participants viewed TCTS stimulus cards. The recordings of the stories were transcribed by a person who was not part of gathering data. The transcriptions were identified by a code number with no other identifying information. Two students scored each story after being trained in scoring of the TCTS by one of the authors of the test. The two students scoring verified their scores with one another and discussed any difficulties. Debriefing was done through presentations of the results of the larger study.

**Results**

We analyzed the CMNI scores for the 11 subscales with a MANOVA comparing men high in conforming and men low in conforming to traditional masculine norms. Seven of the 11 subscales of the CMNI were found to be significantly higher for men in the conforming group than for those in the nonconforming group (see Table 1). This supports our first hypothesis.

A main effect was found for Power Over Women subscale score, \( F(1, 19) = 26.02, p = .001, \eta^2 = .58 \), with conforming men scoring higher (\( M = 10.82, SD = 2.31 \)) than nonconforming men (\( M = 3.67, SD = 2.06 \)). A main effect was found for Dominance subscale, \( F(1, 19) = 20.74, p = .001, \eta^2 = .92 \), with conforming men scoring higher (\( M = 7.55, SD = 1.86 \)) than nonconforming men (\( M = 4.08, SD = 1.78 \)). A main effect was also found for Violence subscale, \( F(1, 19) = 18.19, p = .001, \eta^2 = .46 \), with conforming men scoring higher (\( M = 15.81, SD = 3.87 \)) than nonconforming men (\( M = 2.83, SD = 2.83 \); see Figure 1).

An independent-samples \( t \) test was used to analyze any differences in Power Over Women scores for conforming and nonconforming men for Card 3. Nonconforming men had significantly lower Power Over Women scores (\( M = 4.11, 1.96 \)) than the conforming men, \( t(18) = 6.89, p < .001, d = 3.10 \). Nonconforming men (\( M = 4.11, 1.96 \)) than the conforming men, \( t(18) = 6.89, p < .001, d = 3.10 \).Nonconforming men had significantly lower Power Over Women scores (\( M = 4.11, 1.96 \)) than the conforming men, \( t(18) = 6.89, p < .001, d = 3.10 \). Nonconforming men had significantly lower Power Over Women scores (\( M = 4.11, 1.96 \)) than the conforming men, \( t(18) = 6.89, p < .001, d = 3.10 \).

We calculated TCTS scores following standardized instructions as indicated in the TCTS manual, to determine subscale scores for all 10 cards (Thurston & Craddock, 2009). We focused on the use of aggressive

![FIGURE 1](image-url)

**FIGURE 1**

**Mean Scores for Conformity to Masculine Norms Inventory Subtests Between Two Groups**

Note. This figure provides the M of the conforming and nonconforming groups for each subscale of the Conformity to Masculine Norms Inventory.
language in the verbal response resulting in a dichotomous score of present or not present. For Card 3, 100% of conforming men included aggressive language in their responses, and 75% of nonconforming men included aggressive language in their responses. This difference was not significant, $\chi^2(n = 23) = 3.16, p = .08$. In response to Card 6, 64% of conforming men included aggressive language in their response, and 92% of nonconforming men included aggressive language. This difference was not significant, $\chi^2(n = 23) = 32.65, p = .10$. Thus, our second hypothesis was not supported.

Interactions between comparison groups and aggression categories for Power Over Women on Card 3 could not be calculated due to 100% of conforming men having aggressive language present. There was not a significant interaction between comparison groups and voicing aggression for Card 6 for Power of Women, $F(1, 19) = .013, p = .91, \eta^2 = .001$.

Looking at Card 6 with a focus on the scores for Violence and Dominance (from the CMNI), we found a main effect for Violence, $F(1, 19) = 10.43, p = .004, \eta^2 = .35$ and Dominance, $F(1, 19) = 8.65, p = .008, \eta^2 = .31$; see Figure 2).

We were interested in HRV of conforming and nonconforming men especially when viewing stimuli to elicit feelings of shame. We used the HRV variable RMSSD for our measurement. We used a repeated-measure ANOVA to analyze the RMSSD value across all 10 stimulus cards of the TCTS. Mauchly’s Test of Sphericity indicated that we could not assume sphericity; therefore, we utilized Greenhouse-Geiser. There was a main effect for cards, $F(3.677, 77.22) = 2.74, p = .04, \eta^2 = .12$; RMSSD was lowest for Card 7 ($M = 42.99, SD = 22.37$) and had the least amount of variability. Card 2 had the highest RMSSD ($M = 61.70, SD = 41.96$) and greatest variance. Post hoc analysis indicated that Card 2 and Card 7 were significantly different ($p = .005$), Card 2 and Card 1 were significantly different ($p = .03$). Of most interest for this study is that Card 3 had significantly higher RMSSD than Card 7 ($p = .003$), and Card 6 was significantly higher RMSSD than Card 7 ($p = .03$). Cards 3 and 6 were not significantly different from one another ($p = .88$).

We used a repeated-measures ANOVA to evaluate our change from initial rest, to the specific card reaction, to the final rest. There was no significant difference for Card 7 between initial rest, stimulus, and final rest, $F(1.53, 32.02) = 2.11, p = .15, \eta^2 = .09$. We found a significant difference for Card 2 between initial rest, stimulus, and final rest, $F(1.46, 30.70) = 7.68, p = .004, \eta^2 = .27$.

No main effect was found for our comparison groups for RMSSD across all TCTS cards, $F(1, 21) = 1.19, p = .29, \eta^2 = .05$ There was not a significant interaction between the groups and cards for RMSSD, $F(3.68, 77.22) = 0.37, p = .81, \eta^2 = .02$. Thus, our third hypothesis was not supported.

**Discussion**

Knowledge about the dynamic characteristics of masculinity norms and their effects are crucial for understanding the internal world of men. Social norms are learned as one engages with various groups, including family, friends, community, and the broader society. Decisions to adhere to the social norms are determined via internal choices and managed by internal and external reinforcement (Kivel, 2010). Investigating where a man falls on the continuum of traditional masculine norms can provide insight into how he approaches others and makes decisions about interacting with society. We examined several elements involved in the level of conformity to social norms of masculinity including the level of adherence to traditional norms based on the CMNI, verbal responses to stimuli evoking shame reactions, use of aggressive language, and physiological responses.

To understand the influence of conforming to social norms, we decided we needed to compare men at the extremes of adherence to social norms for traditional values of American masculinity and compare them to men who responded in a way that indicated they were extreme in their lack of adherence to traditional masculine values. Research suggests that emotional responses
to stimuli were both internal and external with men reported to manage their emotions with more negative behaviors (Reilly et al., 2014; Smith et al., 2018).

Our hypothesis that conforming men would score significantly higher on subscales of the CMNI such as Power Over Women, Violence, and Dominance was supported. This is not especially surprising because we were comparing groups from the first and fourth quartile of the sample, yet the two groups were not significantly different on all the CMNI subscales. Power Over Women, Violence, and Dominance had the largest effect sizes suggesting they may contain elements that are central to the group of men who embody traditional masculine norms most fervently. Conforming men appear more likely to consider that having power over women in a variety of situations is acceptable especially in male–female relationships. This viewpoint can interfere with successful interactions in work and personal relationships. A potential toward physical confrontation is also a value that was significantly different between the men who adhere strongly to traditional norms and men who do not hold these values. If managing emotional content is dealt with using physical confrontation, these men can find themselves in difficult situations (Genuchi & Valdez, 2013). These men were more at risk for depression (Genuchi & Valdez, 2013), substance use, and suicide (Rice et al., 2014). Finally, a desire to be in control in all situations (Dominance) is significantly different between the two groups. Here again is a situation that can lead to interpersonal difficulties in personal and work relationships. This is consistent with findings from Smith et al. (2018). There are also potential negative consequences for men who are least likely to conform to traditional masculine values. If society sets the expectation for men to behave and react to emotions in specific ways, being on the opposite side of these norms and not expecting control, not responding with physical confrontation, and not expecting control in male–female relationships can be confusing.

Utilizing the TCTS cards enabled us to evaluate the verbalization of aggression expressed by the participants. If the gender role for men is to respond to negative emotions with externalizing behaviors such as control of others, physical confrontation, and dominance, we might expect aggressive language when faced with a stimulus depicting overt shame or embarrassment. Our second hypothesis was not supported. When viewing the stimulus of a male coach yelling at a female athlete, conforming men included aggressive language in their responses as did many of the non-conforming men. It seems that overt verbal aggression was recognized by men regardless of their conforming or nonconforming values. This likely points to the universality of the aggression presented in the card.

Both depictions of shame used in this study can have various responses, including how coaches react to athletes in the midst of the game. Is yelling at a player for the good of the team? Is this acceptable behavior? With nearly all our sample using aggressive language in this instance, it appears the verbal aggression in this situation was accepted. When viewing a male child being spanked by a female adult while peers looked on, neither conforming nor nonconforming groups were in full agreement that this was aggression. There appears to be more varied opinions about a boy being spanked by a woman. We do not know if it is the act of physical punishment or a woman demonstrating power over a young boy that leads to the varied responses from our sample of men. These results may be consistent with Stanaland and Gaither’s (2021) findings especially for young men when they feel that their masculinity is threatened. Aggressive language appears to be a reaction to experiencing shame-induced negative emotions.

HRV is affected by emotional stimuli measuring both sympathetic and parasympathetic nervous system activation and inhibition (Bernston et al., 2007). RMSSD was found to change across the different TCTS cards. When RMSSD is low, it is indicative of the activation of the parasympathetic system (Porges et al., 2015) and what has been termed vagal tone. When RMSSD is high, it indicates vagal withdrawal or less activation of the parasympathetic system and thus more of a sympathetic system response. Our participants had different RMSSD responses to the various cards with the greatest vagal tone (calm) to the card showing a person walking into a room where a small group is in a conversation. This card appeared to elicit less potential stress or shame than the other nine cards. This may be an indication that there was less of an emotional response to this situation by both conforming and nonconforming men. Participants responded to the card with a young boy doing math on a blackboard with the greatest variability, suggesting this situation evoked a variety of emotional responses from the men. This may have to do with their own individual experiences with performing academic problems in a public place. The cards with the highest or lowest RMSSD were not ones directly depicting dominance, violence, or power over women. Our sample responded with more vagal withdrawal to the two cards that were the focus for this study, each depicting direct forms of dominance and power. This would suggest that HRV provides some indication that a person is processing negative emotions. It appears that the emotional reaction does not change based upon the person’s level of adherence to
Masculinity and Shame

brain wave reactions, has potential benefit for men. Understanding differences between verbal and the more subconscious responses that provide information for those conducting therapy is beneficial. Future research in this area will continue to use physiological measures to understand reactions to emotions and subsequent reactions. Continuing research can be sensitive to the amount of time required by each participant. Utilizing longer measurements of HRV may be beneficial (Laborde et al., 2017).

There are always limits to using self-report inventories. The CMNI was used by many researchers, and it gathers the opinions of the men about their own values. The TCTS cards are shown to elicit shame responses but it is possible that it is not as strong of an emotional reaction as what would be elicited with video or in vivo situations. However, those situations would bring about potential ethical issues. Measuring and interpreting HRV is also challenging. We considered various options in the design of the study. We recognized that our measurement was short for HRV, yet we wanted to be sensitive to the amount of time required by each participant. Utilizing longer measurements of HRV may be beneficial (Laborde et al., 2017).

We were not able to look at differences in responses from men of various ethnicities because most men who responded were White. We were also pulling from a university population, and thus our age range was limited. It may be that men from a wider spread of generations would respond differently to the CMNI and the TCTS cards.

Future Research to Consider

Continuing to use physiological measures to understand responses to emotions and subsequent reactions is beneficial. Future research in this area will continue to provide information for those conducting therapy with men. Understanding differences between verbal responses and the more subconscious responses that can lead to changes in HRV, skin responses, as well as brain wave reactions, has potential benefit for men.

Conclusion

Men fall along the continuum of conformity to traditional American masculinity norms as developed through society and gender roles over the years. It is important to understand that this these values are on a continuum rather than dichotomous. When encountering negative emotions, men with high conformance with traditional norms will be more likely to value being in control, responding with confrontation, and believing they have a right to exhibit dominance within male-female relationships. When exposed to emotional situations with overt aggression, men tend to verbalize the aggression whether they conform or do not conform to masculinity norms. In addition, men will respond physiologically differently to overt shame than to subtle shame. This response was not different between men who hold traditional masculine values and those who do not conform to these values. It is also important to understand that research done with predominantly white, middle-class participants must be applied with caution because men from different ethnic backgrounds and races may respond in a variety of ways to the traditional masculine values than the sample of this study. Continuing research with an expanded sample that is more diverse in age, education, geographical area, ethnic background, and race will be valuable.

References


Stine, Broughton, Andrews, and Spromberg | Masculinity and Shame

https://digitalcommons.georgefox.edu/psyd/252/


Author Note

Glena Andrews https://orcid.org/0000­0002­8622­6101

Christopher Spromberg https://orcid.org/0000­0003­0370­3222

Taylor Broughton is now at Loma Linda Veteran’s Administration Medical Center, Loma Linda, CA.

Glena Andrews is now at RAF Lakenheath United States Air Force Medical Center, Brandon, England.

Christopher Spromberg is now at Livingston HealthCare – Billings Clinic Affiliate, Billings, MT.

We have no known conflict of interest to disclose.

Correspondence concerning this article should be addressed to Glena L. Andrews, PSC 41, Box 3567, APO, AE 09464. Email: GlenaAndrewsPhD@gmail.com
Examining Relationships Between Transgender Prejudice, Gender Essentialism, and Defining and Categorizing Transgender People

Bec Stargel and Angela C. Bell
Department of Psychology, Lafayette College

ABSTRACT. Some definitions and adoptions of the word “transgender” may be rooted in transgender prejudice and result in different conclusions about who is considered transgender. Transnormative definitions, fixed to the gender binary and medical transitions, could lead to categorical exclusions, whereas umbrella definitions, centered on people’s self-identification and often featuring a greater range of gender identities, could lead to broader inclusion. The present work explored whether the relationship between attitudes toward transgender people and transgender target categorization is associated with divergent definitions of the word transgender. We hypothesized that higher prejudice and gender essentialism would predict agreement with transnormative definitions that would then predict lower categorization of transgender targets. We also hypothesized that lower prejudice and gender essentialism would predict agreement with umbrella definitions that would then predict higher transgender categorization. Using a sample of MTurk Workers (N = 497), we tested these hypotheses with a parallel moderated-mediation analysis. Results support that gender essentialism moderated the relationship between prejudice and umbrella definitions, and umbrella definitions mediated the relationship between attitudes and categorization of transgender targets (index of moderated mediation = .46, BCa CI [.20, .76]). Transnormative definition agreement was not predicted by prejudice or gender essentialism, and did not mediate the relationship between these attitudes and categorization (index of moderated mediation = .00, BCa CI [–.06, .05]). These findings provide insight into the role of underlying attitudes on categorization, which has implications on transgender and gender-diverse people’s identity affirmation, social belonging, and healthcare access.

Keywords: transgender prejudice, social categorization, gender essentialism, transphobia

Growing awareness of transgender identity and representation highlights divergent conceptualizations of the word “transgender.” To some, transgender only applies to individuals who have received medical treatments to support transitioning from one gender to another. For others, transgender applies to those who self-identify as such, and self-identification is independent from medical interventions or the gender binary (Kuper et al., 2012; Tate et al., 2014). Henceforth, we use “transnormative definitions” to describe conceptualizations that rely on medicalized and binary narratives and “umbrella definitions” to describe...
conceptualizations that are predominantly focused on a person's self-identification (see Darwin, 2020, for full review of "binary and medicalized" and "umbrella" models). Although definitions of transgender continue to shift and evolve, these two thematic definitions are of particular theoretical interest due to their implications for who is perceived as transgender. Thus, the present work examined the extent to which transgender prejudices and gender essentialist (i.e., biologically based) beliefs predict agreement with transnormative and umbrella definitions, and who is categorized as transgender as a result of these beliefs.

Transnormativity refers to the structure of norms and expectations that transgender people may be held accountable to in order for their identity to be seen as authentic (Johnson, 2016). These perceptions often frame self-hatred, mental illness, and gender dysphoria (i.e., distress due to conflicting gender identity and sex assigned at birth) as core experiences to one's transgender identity (Johnson, 2016; Konnelly, 2021). Such expectations frame surgery and hormone replacement therapy as a necessity (Bilodeau & Renn, 2005; Johnson, 2016). Although many transgender people experience such distress and do choose to medically transition, expecting these as vital features of transgender identity can be limiting to those who do not pursue or have access to such medical care. Further, transnormativity reinforces directional gender binaries (e.g., transitioning from "male-to-female" or "female-to-male") that fails to emphasize current gender identity and to include people who identify with nonbinary genders (Factor & Rothblum, 2008; Kuper et al., 2012). Overall, transnormative definitions may fail to represent the wide range of transgender people's experiences with gender and may reflect underlying prejudices.

The metaphor of the umbrella frames transgender identity as a broad spectrum. For example, in a survey of 1,211 transgender people, there were 279 additional unique descriptors of gender identity (e.g., "genderqueer," "transvestite," "two-spirit," "butch" and "I am my own gender;" Beemyn & Rankin, 2011). Further, Puckett and colleagues (2020) found that transgender and gender diverse participants suggested a range of identities, such as "nonbinary," "genderfluid," and "agender," to be included in demographic questionnaires. The ability to choose one's transgender identity is said to provide greater agency to individuals (Darwin, 2020; Valentine, 2007). Umbrella definitions can be especially important if they counter transnormative definitions that are said to give the power of labeling to medical professionals (Spade, 2013; Stone, 1992).

Conceptions of the word transgender may be influenced by prejudice toward transgender people. Buck's (2016) analysis of open-ended definitions of the word "transgender" revealed that heterosexual, cisgender participants who referenced sex/gender changes had more negative attitudes toward transgender people, whereas those who referenced current gender identity reported more positive attitudes. These findings were supported by a recent replication study (Anderson, 2022). Moreover, negative feelings toward transgender people are associated with a higher endorsement of belief in the gender binary (Norton & Herek, 2013). Traditional gender views that influence outcomes of transgender prejudice may be due to gender essentialism, defined as the belief that "the differences between the sexes are of an intrinsic nature, closely associated with physical, physiological, and/or spiritual differences" (Crompton & Lyonette, 2005, p. 601). Gender essentialism has been linked to the justification of traditional gender norms, beliefs that gender is biological and unchangeable, and opposition to transgender civil rights (Brescoll et al., 2013; Wilton et al., 2019). Transgender prejudice and gender essentialist beliefs appear to similarly reflect themes of transnormative and umbrella definitions.

Social categorization helps people define and understand a complex world. Transgender categorization, however, may be more complicated for people who do not approach transgender identity through self-ascription. Previous work on transgender categorization has predominantly focused on how cisgender targets categorize the gender of transgender targets, and whether they categorize them by their stated gender identity or by the gender/sex they were assigned at birth (Gülgöz et al., 2018; Howansky et al., 2020, 2021; Stern & Rule, 2018). Other work has qualitatively explored transgender people's experiences—or fear—of exclusion from the category of transgender (Catalano, 2015; Darwin, 2017, 2020; Garrison, 2018). To our knowledge, research has yet to explore factors that influence whether or not people categorize individuals as transgender. Such categorization could impact one's ability to access certain forms of healthcare (Spade, 2013). Additionally, categorization could influence an individual's comfort with the label of transgender, or connection to transgender communities, both of which have important consequences on health and well-being (Barr et al., 2016; Testa et al., 2014). Therefore, it is important to explore how cisgender observers categorize gender diverse targets as transgender.

Past work on biracial and multiracial categorization might inform how transgender prejudice and gender essentialism could influence transgender categorization. Blascovich and colleagues (1997) found that participants high in racial prejudice (i.e., compared to those lower in prejudice) took longer to categorize "ambiguous" faces as either Black or White, supporting the conclusion that
people with higher prejudice may be more cautious about these categorizations. Additionally, Ho and colleagues (2015) found that the interaction between high racial essentialism and high interracial bias predicted that multiracial targets would be more likely to be categorized as Black. Finally, Chen and colleagues (2014) found that people who were internally motivated to control racial prejudice (i.e., more egalitarian) were more likely to categorize such targets as multiracial. These studies show how prejudice can be moderated by essentialism and may influence how outgroup members may categorize “ambiguous” social targets. Such research provides a potential framework through which transgender identity categorization can be explored.

Although race and transgender identity—and the underlying processes that systematically regulate them—cannot be conflated, it is possible that the relationships between prejudice, essentialism, and categorization may also exist within the domain of transgender identity. Using this framework as an exploratory mechanism, it is unclear what the directionality of such an effect might be. In the context of racial categorization, Krosch and colleagues (2013) found that people categorized multiracial targets as being members of the more socially subordinated racial group. This notion of “hypodescent” categorization may occur for other social groups. If a similar pattern follows for transgender categorization, would inclusion or exclusion from the category be seen as more socially subordinated? This is where transgender definitions may play an important mediating role between attitudes and categorization.

The Present Study
This investigation examined the ability of transgender prejudice and gender essentialism to predict transgender definition agreement and transgender target categorization (see Figure 1 for a conceptual model). We predicted that lower prejudice and gender essentialism would predict greater agreement with umbrella definitions that in turn would positively predict target categorization (i.e., resulting in targets categorized overall as more transgender). Then, we predicted that higher prejudice and gender essentialism would predict greater agreement with transnormative definitions of the word “transgender” that in turn would negatively predict target categorization (i.e., resulting in targets categorized overall as less transgender). We also anticipated that transgender prejudice and gender essentialism would indirectly predict categorization through their relation to these definitions.

Method
Participants and Procedure
Participants (N = 637) were recruited via TurkPrime and were over 18 years of age, spoke English, and resided in the United States. One hundred fifteen participants were excluded for duplicate responses (7), giving incorrect or nonsensical answers to attention checks (77), or responding affirmatively to the question “Is there any reason we should not use your data?” (31). Because the current study was primarily focused on the perceptions of cisgender people, we excluded 25 participants who identified as noncisgender. Of the remaining 497 participants, M_age = 39.34 years, 361 identified as White, 262 women, and 440 heterosexual. See Table 1 for full demographics.

Participants received directions to answer questions honestly based on their personal definitions and feelings. First, they rated their agreement with a series of definitions for the word “transgender” and categorized a series of written targets as transgender (these two tasks were counterbalanced). Participants then completed a measure of transgender prejudice, followed by a measure of gender essentialism. Finally, participants responded to demographic questions and were then debriefed and dismissed. Lafayette College’s Institutional Review Board approved all procedures. All materials, data files, and

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>N = 497</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, M (SD)</td>
<td>39.34 (12.51)</td>
</tr>
<tr>
<td>Gender, n (%)</td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>262 (52.7%)</td>
</tr>
<tr>
<td>Men</td>
<td>234 (47.1%)</td>
</tr>
<tr>
<td>Nonbinary or Genderqueer</td>
<td>1 (0.2%)</td>
</tr>
<tr>
<td>Sexuality</td>
<td></td>
</tr>
<tr>
<td>Heterosexual</td>
<td>440 (88.5%)</td>
</tr>
<tr>
<td>Bisexual</td>
<td>24 (4.8%)</td>
</tr>
<tr>
<td>Gay</td>
<td>10 (2.0%)</td>
</tr>
<tr>
<td>Lesbian</td>
<td>4 (0.8%)</td>
</tr>
<tr>
<td>Pansexual</td>
<td>3 (0.6%)</td>
</tr>
<tr>
<td>Asexual</td>
<td>4 (0.8%)</td>
</tr>
<tr>
<td>Queer</td>
<td>2 (0.4%)</td>
</tr>
<tr>
<td>Demisexual</td>
<td>1 (0.2%)</td>
</tr>
<tr>
<td>Selected two or more</td>
<td>9 (1.8%)</td>
</tr>
<tr>
<td>Race/Ethnicity</td>
<td></td>
</tr>
<tr>
<td>White/European American</td>
<td>361 (72.6%)</td>
</tr>
<tr>
<td>Black/African American</td>
<td>43 (8.7%)</td>
</tr>
<tr>
<td>Asian-descent/Asian American</td>
<td>40 (8.0%)</td>
</tr>
<tr>
<td>Hispanic/Latinx</td>
<td>22 (4.4%)</td>
</tr>
<tr>
<td>Indigenous/Native American</td>
<td>3 (0.6%)</td>
</tr>
<tr>
<td>Selected two or more races/ethnicities</td>
<td>23 (4.6%)</td>
</tr>
<tr>
<td>Another race/ethnicity</td>
<td>5 (1.0%)</td>
</tr>
</tbody>
</table>
supplementary analysis including all gender-diverse participants are available at https://osf.io/h87qj/.

Materials

Definition Agreement

The definition agreement task was developed by the first author for this study and was not pretested. Definitions reflect Darwin’s (2020) framework of the medical and binary model and the umbrella model. Four descriptions of the word “transgender” were provided in counterbalanced order. Two definitions reflected the transnormative definitions (e.g., “Transgender people are individuals who transition from male to female or from female to male”), and two reflected the umbrella definitions (e.g., “Transgender refers to anyone who identifies as transgender, or as a member of the transgender community, regardless of other characteristics”). Participants rated their agreement on a scale from 1 (strongly disagree) to 7 (strongly agree).

There was one index of transnormative definition agreement (M = 5.5, SD = 1.0) and one index of umbrella definition agreement (M = 5.2, SD = 1.3). These two-item indices had low reliability between the items (transnormative α = .38, umbrella α = .57). However, an exploratory factor analysis revealed two common factors consistent with our designated transnormative and umbrella definition agreement indices, and were extracted from the four items (KMO = .52; Bartlett's test of sphericity, χ²(6) = 131.06, p < .001). The eigenvalues (>1, also supported by the scree-plot) indicated that the two factors explained 37.52% (transnormative definitions) and 28.48% (umbrella definitions) amount of variance. PCA extraction with Varimax with Kaiser Normalization rotation revealed all items had primary loadings over 0.77 with no cross-loadings above 0.10. Due to the factor analysis, we created two composite scores from the two extracted factors based on the mean of the items, which had their primary loadings on each factor. Higher scores indicated greater definition agreement for each index. Data analysis that treated each definition agreement separately revealed no significant differences in final data analysis, and can be found in the supplementary materials.

Categorization Task

Participants rated a counterbalanced series of 23 target descriptors developed by the first author and based on past literature that presented a variety of descriptions of transgender people’s potential experiences (Beemyn & Rankin, 2011; Davidson, 2007; Factor & Rothblum, 2008). An example descriptor is, “A person who was assigned female at birth who identifies and presents as male, and who has not undergone hormone replacement therapy nor gender confirmation surgery, but wants to.” Ratings were made on a 0 to 100 scale adapted from a feeling thermometer measure from Bell et al. (2021), and included the following instructions:

FIGURE 1

Moderate-Mediation Model

<table>
<thead>
<tr>
<th>Transnormative definition agreement</th>
<th>Categorization score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transgender prejudice</td>
<td></td>
</tr>
<tr>
<td>a₁ = −0.19***</td>
<td></td>
</tr>
<tr>
<td>Gender essentialism</td>
<td></td>
</tr>
<tr>
<td>Prejudice x GE interaction = .08**</td>
<td></td>
</tr>
<tr>
<td>b₁ = 5.91***</td>
<td></td>
</tr>
</tbody>
</table>

| Indirect effect for Low GE = −1.29, a; b₁ = −1.69, BCa CI[−2.45, −1.06] |
| Indirect effect for Average GE = 0.00, a; b₁ = −1.10, BCa CI[−1.67, −0.58] |
| Index of moderated mediation for umbrella definition agreement = 0.46, BCa CI[0.20, 0.76] |

Note. Reported effects are unstandardized coefficients. Only statistically significant effects are included in this figure. See supplementary files on OSF for reporting of nonsignificant effects. ***p < .001; **p < .01; *p < .05.
Table 2: Pearson Correlations Between Variables (N = 497)

<table>
<thead>
<tr>
<th>Variable</th>
<th>M (SD)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Transnormative definition agreement</td>
<td>5.5</td>
<td>(1.0)</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>2. Umbrella definition agreement</td>
<td>5.8</td>
<td>(1.3)</td>
<td>.14</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>3. Transgender prejudice</td>
<td>4.1</td>
<td>(1.8)</td>
<td>-.02</td>
<td>-.25</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>4. Gender essentialism</td>
<td>4.2</td>
<td>(1.3)</td>
<td>-.01</td>
<td>-.14</td>
<td>.63</td>
<td>--</td>
</tr>
<tr>
<td>5. Target categorization</td>
<td>51.2</td>
<td>(17.4)</td>
<td>.10</td>
<td>.45</td>
<td>-.09</td>
<td>-.03</td>
</tr>
</tbody>
</table>

Note: *p < .05 **p < .01 ***p < .001.

Parallel Moderated Mediation Analysis
We tested whether transnormative and umbrella definition agreement simultaneously mediated the relationship between transgender prejudice and target categorization, and whether gender essentialism moderates the relationship between transgender prejudice and each definition agreement. The parallel moderated-mediation analysis was conducted through the PROCESS macro in SPSS (Model 7; Hayes, 2013; Preacher & Hayes, 2004). As seen in Figure 1, results indicated that transgender prejudice is indirectly related to transgender categorization through its relationship with umbrella definition agreement (as moderated by gender essentialism) but not through transnormative definition agreement. See Table 3 for $R^2$ values that highlight the proportion of variance in categorization scores explained by each model.
Evaluating Transnormative Definition Agreement as the Mediator

Transgender prejudice was not directly associated with transnormative definition agreement \( (a_1 = -.01, p = .73) \), transnormative definition agreement did not predict categorization \( (b_1 = .70, p = .31) \), and the relationship between transgender prejudice and transnormative definition agreement was not moderated by gender essentialism (interaction \( = -.00 p = .87 \)). A 95% bias-corrected confidence interval based on 10,000 bootstrap samples indicated that the indirect effect through transnormative definition agreement, holding the umbrella definition agreement mediator constant, did not fall above zero at any level of gender essentialism (GE\(_{\text{1SD}}\) = –1.29, \( a_1 b_1 = -0.01, BCa CI \[ -0.12, 0.10 \]; GE\(_{\text{0}}\) = 0.00, \( a_1 b_1 = -0.01, BCa CI \[ -0.10, 0.08 \]). \( \text{GE}_{-1\text{SD}} = 1.29, a_1 b_1 = -0.01, BCa CI \[ -0.14, 0.10 \] \)). The overall moderated mediation model was not supported with an index of moderated mediation for transnormative definition agreement as the mediator \( a_1 = -0.00, BCa CI \[ -0.06, 0.05 \].

Evaluating Umbrella Definition Agreement as the Mediator

Transgender prejudice was a significant and negative predictor of umbrella definition agreement \( (a_2 = -0.19, p < .001) \). Then, umbrella definition agreement significantly and positively predicted target categorization \( (b_2 = 5.91, p < .001) \). The effect of transgender prejudice on umbrella definition agreement was moderated by gender essentialism (interaction \( = 0.08 p = .001 \)).

To probe the moderating effect, conditional indirect effects were examined using the procedures of Aiken and colleagues (1991). The effects of transgender prejudice on umbrella definition agreement were examined by a simple slopes analysis at the mean as well as at +/-1 SD above and below the mean of gender essentialism (see Figure 2). When gender essentialism was low or at the mean, lower levels of transgender prejudice were associated with a higher umbrella definition agreement (GE\(_{\text{1SD}}\) = –0.29, CI \[ -0.38, –0.19 \]; GE\(_{\text{0}}\) = –0.19, CI \[ -0.26, –0.11 \]). In contrast, when gender essentialism was high, transgender prejudice no longer predicted umbrella definition agreement (GE\(_{\text{1SD}}\) = –0.09, CI \[ -0.19, 0.01 \]).

A 95% bias-corrected confidence interval based on 10,000 bootstrap samples indicated that the indirect effect through umbrella definition agreement, holding the transnormative definition agreement mediator constant, fell below zero when gender essentialism was at low and mean levels (GE\(_{\text{1SD}}\) \( c = 1.29, a_2 b_2 = -1.69, BCa CI \[ -2.45, –1.06 \]; GE \( c = 0.00, a_2 b_2 = -1.10, BCa CI \[ -1.67, –0.58 \]), but not when gender essentialism was high (GE\(_{\text{1SD}}\) \( c = 1.29, a_2 b_2 = -0.51, BCa CI \[ -1.11, 0.09 \]).

Finally, transgender prejudice did not predict greater target categorization when taking into account the moderating factor of gender essentialism and its indirect effect through both definition agreements \( (c' = 0.25, p = .52) \). The overall moderated mediation model was supported with an index of moderated mediation for umbrella definition agreement as the mediator \( = .46, BCa CI \[ .20, .76 \]. See Figure 1 for visualization of results.

Discussion

The present study assessed the interplay of transgender prejudice and gender essentialism to evaluate their influence on transgender definition agreement and subsequent categorization of transgender targets. The moderated-moderation regression analysis supported the predicted model only when umbrella definition agreement was treated as a mediator, but not when transnormative definition agreement was treated as a mediator. Although not all findings were as expected, the overall results have important implications considering how divergent narratives about transgender identity can reflect existing prejudices and influence perceptions of who is “trans enough.”

When umbrella definition agreement was the mediator, transgender prejudice did not directly predict target categorization. Umbrella definition agreement did positively predict categorization such that higher agreement with umbrella definitions was associated with higher categorization scores (as hypothesized). Moreover, the indirect effect of transgender prejudice on categorization via umbrella definition agreement was moderated by gender essentialism. Specifically, low
transgender prejudice predicted higher umbrella definition agreement, which led to a higher categorization, but only for those who are low-to-average in gender essentialism. These results align with work that found that definitions associated with identity were associated with lower prejudice (Anderson, 2022; Buck, 2016). Further, the findings reflect theoretical arguments, such as those by Ekins and King (2006), which assert that these definitions are historically associated with transgender activism and self-advocacy and would thus be associated with lower prejudice and a rejection of gender essentialist views. Although gender essentialism alone did not predict agreement with umbrella definitions, the combination of low prejudice and low gender essentialism predicted greater agreement with umbrella definitions. Additionally, umbrella definition agreement explained the relationship between prejudice and categorization among those with average or below levels of gender essentialism. This not only supported the hypothesis that the types of relationship identified in Ho and colleague's (2015) work on racial categorization may present a framework through which transgender categorization can also be explored, but also identified the specific mechanism of this relationship. For those with lower levels of prejudice and gender essentialism, categorizing targets as transgender appears to be based on conceptions of the category as broad, inclusive, and based on self-identity.

When transnormative definition agreement was treated as a mediator, transgender prejudice and gender essentialism did not predict transnormative definition agreement or target categorization. Transnormative definition agreement was positively correlated with categorization, but this relationship did not persist in the moderated-mediation model. One potential explanation for these findings is that the definition task did not fully measure belief in transnormative beliefs. Transnormative understandings not only center dysphoria and medical transition, but also enforce a hierarchy, which rejects all other narratives and experiences (Johnson, 2016). Umbrella understandings of transgender also recognize binary medical transitions and experiences with dysphoria as important aspects of many transgender people’s experiences, just as they also legitimize other experiences. Because participants were not required to choose between definitions or rank them, and because these definitions did not include explicitly exclusionary language, they may not necessarily reflect the exclusionary beliefs associated with transnormativity. This explanation could account for why the results of this aspect of the current study do not reflect those of Buck (2016) and Anderson (2022), who found that higher prejudice was associated with providing definitions, which focused on transition, rather than identity. This focus on transition is one aspect of transnormative definitions, in addition to expectations of binary gender and dysphoria. In Bucks’ (2016) study, as well as Anderson’s (2022), participants were asked to produce one definition which was likely the one that was most salient to them or the one they agreed with the most. In the current study, however, all participants rated several prewritten definitions and could agree with multiple definitions. A more exclusionary task might have been better suited to measuring exclusionary definitions.

Limitations and Future Directions
The present work sought to understand the processes by which underlying attitudes can influence agreement with transgender definitions and categorizing transgender people. The study is not without limitations in its development and selection of measures. Billard’s (2018) Attitudes Toward Transgender Men and Women Scale has been found to have strong validity, but is limited in its ability to evaluate explicit prejudice and does not measure attitudes toward transgender people with nonbinary genders who were represented in the categorization task. Future work could test other forms of transgender prejudice (i.e., implicit bias, ignorance or lack of education of transgender identity), and attitudes toward noncisgender people, to evaluate their relationships with definitional beliefs and subsequent categorization. For this study, definition and categorization tasks were developed by the first author and not pretested. Umbrella and transnormative definitions were based on existing and common conceptualizations of how transgender is defined but did not account for the full range of possible definitions or lay perceptions of the word. For example, some people do not view transgender as a valid category, and believe transgender people to be deceptive or pretending about their gender identity (Beauchamp, 2019). Others may believe that transgender people are confused about their gender identity (Gazzola & Morrison, 2014). In a recent study, Anderson (2022) found that, when people referenced themes of confusion in defining the word transgender, they also reported higher levels of prejudice toward transgender people. Such themes might present an additional category of transgender definitions for testing in follow-up work.

Additional work could also explore the reliability and validity of the categorization task because it was developed for this analysis and was not tested prior to this study. This categorization task also deviates from past literature on transgender categorization in which cisgender participants categorized transgender targets...
as either their gender or their gender assigned at birth (Gülgöz et al., 2018; Howansky et al., 2020; 2021; Stern & Rule, 2018), or in which transgender people reflected on feelings or experiences of inclusion and exclusion from the category (Catalano, 2015; Darwin, 2017; 2020; Garrison, 2018). This study was particularly focused on the extent to which cisgender participants regarded a target as transgender or not, which most similarly reflected past research on multiracial target categorization (Blascovich et al., 1997; Chen et al., 2014; Ho et al., 2015). Further evaluation of the categorization procedure could best determine whether targets (that varied by gender identity, gender presentation, gender assigned at birth, and references to medical transitions) were most representative of gender-diverse individuals. The targets developed for this study did not explicitly feature gender dysphoria, an important aspect of transnormativity that could affect people’s responses to target categorization, and the targets did not incorporate other social identities. In particular, transgender people of color have experienced a long and continued history of exclusion from transgender narratives, activism, and support systems (Skidmore, 2011; Snorton, 2017). Categorization could be uniquely influenced by identity multiplicity and intersectionality if gender diverse individuals are perceived as more or less transgender due to other identity factors, like race/ethnicity and socioeconomic status.

Regarding the sample and participants, a large subsample was excluded from analysis because participants failed one or more attention and quality assessments. The increasing presence of bots and MTurk “farmers” required intensive screening measures to identify and locate sources of such low-quality data in MTurk samples (Chmielewski and Kucker, 2020). Additionally, participants who qualified for final data analysis were mostly heterosexual and cisgender. Schudson and colleagues (2019) found that transgender people defined words related to gender, like man or woman, with a greater level of complexity than their cisgender counterparts. Following this work, we could expect that transgender people might differ in their agreement with, and conceptualizations of, transnormative and umbrella definitions. Because of the small proportion of respondents who identified as a label other than cisgender, such differences could not be explored in the current study. This sample limitation is particularly relevant because issues related to definition agreement and categorization have important impacts on acceptance within transgender spaces (Darwin, 2020; Sutherland, 2021). Future work should examine noncisgender perspectives regarding definitions and categorizations of transgender.

Due to the cross-sectional design, the analysis of the current project is limited in its ability to explore causal effects, and future research could experimentally test whether gender essentialism and transgender definitions could influence target categorization. Such research could be modeled after Ching and Xu (2018), who manipulated gender neuroessentialism (a subtype of gender essentialism), by providing fictional articles that attributed sex-based personality and behavior differences on either neurological or social explanations, before evaluating the extent of people’s transgender prejudices. They found that those exposed to the neuroessentialist condition reported higher levels of transgender prejudice compared to those exposed to the condition which also discussed social influences, as well as the control condition. Future work could similarly manipulate gender essentialism experimentally and measure its effects on agreement with different definitions of transgender as well as categorization process.

Implications and Conclusion

This project has laid groundwork for future exploration of perceived transgender categorization. Categorization was the outcome in this study but past work has found that categorization can consequently affect well-being and resilience for transgender people (Barr et al., 2016; Testa et al., 2014). For example, some nonbinary people do not identify with the label “transgender” because of perceived exclusion based on binary and medicalized definitions of the word (Darwin, 2017, 2020). Nonbinary people may also feel pressure to prove that they are “trans enough” in the face of transnormative narratives (Garrison, 2018). Then, categorization benchmarks utilized by medical gatekeepers, like doctors and therapists, could limit how transgender and gender-diverse individuals can access healthcare (Konnelly, 2021; Spade, 2013).

This study’s findings demonstrate lower prejudiced and essentialist attitudes, and umbrella definitional beliefs are positively related to transgender target categorization. To encourage greater transgender inclusion, it is critical to address individual attitudes and how transgender is defined to buffer against deleterious effects associated with categorization. Approaches to transgender categorization have important implications for psychological and physical well-being, healthcare experiences, community building, and activist movements. Wide adoption of umbrella definitions may improve education and intervention efforts to counteract prejudice and essentialist beliefs, and increase inclusion with broader categorization.
Defining and Categorizing Transgender People


https://doi.org/10.1080/00224495.2011.596564


https://doi.org/10.1037/194855609187514


https://doi.org/10.1080/15532739.2014.937041


Defining and Categorizing Transgender People

Materials and data for this study can be accessed at https://osf.io/h87qj/.

We have no known conflict of interest to disclose. The study was supported by funding from the Academic Research Committee at Lafayette College.

Preregistration for this study can be found at https://osf.io/2wvg5. This project was preregistered prior to the collection of any data. There were two changes between the preregistered plan and the final manuscript. First, an additional round of participants were sampled to increase statistical power, and received the same pre-registered materials and procedure. Second, the final analysis used a parallel moderated mediation model, rather than two separate moderated mediation models. This change was a better representation of testing the central research question and subsequently simplified the presentation of results. The analysis for the original separated models are described in the supplementary data analysis files available on OSF.

Correspondence concerning this article should be addressed to Bec Stargel, Department of Sociology, North Carolina State University, 10 Current Dr., Raleigh, NC, 27606. Email: rstarge@ncsu.edu

Author Note
Bec Stargel https://orcid.org/0000-0002-3549-3226
Angela C. Bell https://orcid.org/0000-0001-9870-8645
Stargel is now a PhD student at North Carolina State University.
Positionality Statement: Angela Bell identifies as a cisgender biracial (Black, White) woman, and is a trained social psychologist from the United States. Both authors acknowledge that, regardless of disclosure, their perspectives are influenced by their positions within dimensions of their identities.
The Patterns of Children’s and Caregivers’ Gender-Typed Exhibit Choices in a Pop-Up Children’s Museum

Brandon Garcia, Natassia Aleman-Teweles, and Jennifer Dyer-Seymour
Department of Psychology, California State University, Monterey Bay

ABSTRACT. A children’s museum is a place to explore and play. It allows children and caregivers to choose among a variety of exhibits. We were interested in visitors’ choices to play in exhibits that included activities typed for their gender. We observed 71 family groups for 15 minutes each and noted the extent to which children and caregivers chose exhibits typed for their gender. We found that, for the majority of the time during our observation, women, men, and boys chose exhibits that included activities not typed for their gender, whereas girls spent the majority of the time during our observation at exhibits that included activities typed for their gender. These findings suggest that museums may allow for a kind of freedom from certain expectations for children’s play.

Keywords: children’s museums, gender, gender stereotypes, exhibit choice

ABSTRACTO. Un museo para niños es un lugar para explorar y jugar. Permite a los niños y cuidadores elegir entre una variedad de exhibiciones. Estábamos interesados en las selecciones de los visitantes para jugar en exhibiciones que incluyan actividades tipificadas para su género. Observamos 71 grupos familiares durante 15 minutos cada grupo y notamos hasta qué punto los niños y los cuidadores eligieron exhibiciones escritas para su género. Descubrimos que, durante la mayor parte del tiempo durante nuestra observación, las mujeres, los hombres, y los niños varones eligieron exhibiciones que incluyan actividades no tipificadas para su género, mientras que las niñas pasaron la mayor parte del tiempo durante nuestra observación en exhibiciones que incluyan actividades tipificadas para su género. Estos hallazgos sugieren que los museos pueden permitir una especie de libertad de ciertas expectativas para el juego de los niños.

Palabras clave: museos para niños, género, estereotipos de género, elección de exhibición
Families in many communities in the United States have the privilege of leaving their distractions at home and visiting a space such as a children’s museum that was intentionally designed for exploration, play, and learning. Children and family members choose exhibits to play in and spend as little or as much time as they want in each one. The freedom to explore and play is fundamental to the children’s museum and is a part of how children learn there (Callanan et al., 2020; Dierking & Falk, 1994).

The aim of the current study was to measure the extent to which family members chose to visit exhibits that did and did not include activities that were stereotypically associated with their gender. We wanted to better understand visitors’ choices over time during their visit. Specifically, we wanted to know to what extent visitors choose and stay at exhibits with activities stereotypically associated with their gender and to what extent they explore exhibits with activities associated with other genders or that include gender-neutral activities. Gender schema theory (Martin & Halverson, 1981; Martin et al., 2002) suggests that children would identify the exhibits that were appropriate for their gender and be more likely to play in those exhibits than in those that appeared to be for other genders. Alternatively, it is possible that some of the advances in gender representation since 1990 in the United States (e.g., five women on the Supreme Court; the first ever woman who won the Democratic nomination for President of the United States in 2016; 15% of CEOs are female, up from 0% in 1995) might have made an environment where children of all genders feel more freedom to explore outside of gender norms. We wanted to measure whether such advances would reveal themselves in children’s choices in the museum and be different from similar investigations inside of children’s museums in the 1990s (Greenfield, 1995; Kremer & Mullins, 1992; Luria & Herzog, 1991). This investigation is important because it provides evidence about the extent to which children’s museums are places where visitors are constrained by gender stereotypes and the extent to which they feel some freedom to explore outside those stereotypes, perhaps showing whether societal changes over time show up in the museum. When institutions make it possible for people to step outside the norms that say people with a certain gender have to behave in certain ways, more advances in gender representation may be possible. For example, U.S. society continues to advance in its acknowledgment of people who identify as nonbinary on the gender spectrum. In this article, we attempted to support such an acknowledgment by using nonbinary language as much as possible. Institutions such as children’s museums might afford tighter or looser constraints on gendered behavior than has been observed in other settings such as schools, homes, and with toys.

Gendered Behavior in School and Home Settings
The play behaviors of children of different genders have been studied since at least the 1970s, and these studies show that as early as age 2, many children show gendered preferences for playmates and types of play (e.g., rough-and-tumble play, circle games such as hand-clapping games; Alexander & Hines, 1994; DiPietro, 1981; Fishbein & Imai, 1993; Jacklin & Maccoby, 1978; Leaper & Farkas, 2015; Maccoby and Jacklin, 1987; Martin et al., 2013; see Mathur & Parameswaran, 2015, for the early origins of gendered play in European-American society). At the same time, there appear to be moments when some children demonstrate less or little gendered play. For example, Mathur and Parameswaran (2015) found that Mexican migrant children engaged in little if any gendered play in Head Start preschools. In addition, Goble et al. (2012) observed that, whereas boys and girls preferred gender-typed activities when they were alone, girls engaged in less gender-typed play when with boys, and boys engaged in less gender-typed play when the teacher was nearby.

In addition to playmates and types of play, many children have tended to show, even into the mid 2010s, gendered preferences for play with particular toys (Dinella & Weisgram, 2018; Fagot & Patterson, 1969; Fulcher & Hayes, 2018; Liss, 1981; Vance & McCall, 1934). These preferences seem to be shaped by several factors including the person playing with the toy (Shutts et al., 2010) and the color of the toy (Weisgram et al., 2014). For example, in a clever experiment that tested children’s interests in toys and manipulated their colors, Weisgram et al. (2014) found that girls were equally interested in stereotypic girls’ toys and stereotypic boys’ toys that had been painted pink and purple. That is, girls were as likely to be interested in a tea set as they were in a monster truck that had been painted pink or purple. The boys did not show the same pattern; they were not more likely to play with a tea set in masculine colors compared to masculine toys in their typical colors or masculine toys painted pink and purple. In other words, boys preferred toys aligned with their gender no matter the color, and girls explored more widely when toys were in female-stereotyped colors of pink and purple.

Gendered Behavior in Museum Settings
Most work on gender in museums has focused on the behavior of the parents. One line of work has examined the way that parents’ talk is different with their daughters compared to their sons. At science exhibits in a children’s museum, Crowley et al. (2001) found that parents of all genders were more likely to provide explanations of the phenomena to boys than to girls. Other work has focused on the differences between mothers and fathers.
Gender Exhibit Choices | Garcia, Aleman-Teweles, and Dyer-Seymour

in their tendency to interact with their children and the exhibits or stand back and observe. Observations reveal that fathers tend to be more interactive with the exhibits and their children in museums and science centers compared to mothers who tend to observe (Diamond, 1986; Dierking & Falk, 1994; Garner, 2015; Koran et al., 1986; Nadelson, 2013; see Diamond, 1994, for a review of gendered behavior in science museums).

In addition to the data on the way parents of different genders behave in the museum, at least three studies have focused on the way children (and sometimes, parents) of different genders were attracted to different kinds of exhibits as well as how they behaved at those exhibits (Greenfield, 1995; Kremer & Mullins, 1992; Luria & Herzog, 1991). Overall, all three studies found differences by gender, but each was slightly different.

Kremer and Mullins (1992) observed children at the Center for Science and Industry in Columbus, Ohio. They focused on five exhibits in KIDSPACE: water jets, bubbles, face paints, animal lab, and build-a-house. The researchers did not categorize each exhibit a priori as gendered in any way. Instead, they counted the number of girls and boys who visited each exhibit, and they described the children's behavior at each exhibit. They found no difference in the number of girls and boys who visited each of five exhibits, yet they did find differences when they examined the children's behavior at the exhibits. For example, at the water jets exhibit, they found that more boys (28%) than girls (1%) pretended that the water jets were guns. Also, they found that more girls (35%) than boys (5%) showed nurturing behavior at the animal lab exhibit.

Unlike Kremer and Mullins (1992), the other two studies did find differences in the number of girls and boys who visited certain exhibits. Luria and Herzog (1991) observed school-aged children at three exhibits at the Boston Children's Museum during school trips: an automobile, computers, and a grocery store. The authors thought that the automobile and grocery store were potentially gender-typed, yet they stated that the museum staff thought those exhibits drew girls and boys equally. Their observations revealed that more boys (42%) than girls (29%) visited the automobile, and more boys (33%) than girls (23%) visited the computers. Conversely, more girls (48%) than boys (25%) visited the supermarket grocery store.

Greenfield (1995) observed children with their parents as well as alone at 34 exhibits separated into four categories at the special “aMAZing Science” area at the Bishop Museum in Honolulu: the human body, physical science principles, puzzles, and computer games. Greenfield did not have an a priori code the exhibits as aligning with one gender or another, but she did find differences in visits by the children’s gender. Whether children were alone or with their parents, the same patterns emerged, although the differences by gender were weaker when children were with their parents. When children were with their parents, more boys than girls visited three of the four kinds of exhibits: those focused on the human body (56% and 44%, respectively); those focused on physical science (56% and 44%, respectively); and those focused on computers (65% and 35%, respectively). One kind of exhibit saw more girls than boys visit, and that was puzzles (61% and 39%, respectively). Of note, Greenfield’s (1995) observations of adults revealed that men and women did not differ in their exhibit choices; they all tried out all the exhibits relatively equally.

Of particular interest to our investigation was that Luria and Herzog (1991) and Greenfield (1995) focused on the data that showed that the children in their studies were most likely to visit exhibits that included stereotypic activities for their own gender (i.e., automobiles, computers, and science for boys, grocery stores and puzzles for girls). Yet, their results also showed that children visited exhibits that included activities that were stereotypical for other genders. The fact that children spread their attention across multiple kinds of exhibits got little attention. In the current study, we designed our methodology so that we could follow family members around the museum and document their choices over time, enabling us to see when children chose exhibits with activities stereotypical of their own gender and when they chose exhibits with activities for other genders. This methodology allowed us to report data in the same way as previous studies (i.e., what was the percentage of boys and girls at a target exhibit), as well as report the sequence of exhibit choices made by family members during a 15-minute period. Looking at the data in both of these ways allowed for a fuller picture of the exhibit choices made by family members.

For the two different ways of looking at the data, there were a few possibilities of what could result. When looking just at specific exhibits and not following people, we could replicate the previous results and show that a greater percentage of boys and girls visited exhibits that include activities that were stereotypic for their gender. Such a finding would be consistent with gender schema theory (Martin & Halverson, 1981; Martin et al., 2002). Alternatively, results could show that things have changed in the intervening 30 years and girls and boys do not show a preference for exhibits that include stereotypic activities for their gender. For our methodology of following visitors around the museum, there were at least three possible patterns for children's and adults' choices of where to play and how much they would stay or move about the museum. The first possibility was that...
they could choose an exhibit that was specific to their gender and either stay there or move to other exhibits that were typed for their gender. Such a finding would be consistent with gender schema theory (Martin & Halverson, 1981; Martin et al., 2002). A second possibility was that they could choose an exhibit that was not typed for their gender and stay at those types of exhibits. The third possibility was that they could vary their choices and play at some exhibits that were typed for their gender, some that were not typed for their gender, and some that were gender-neutral.

In the present study we observed children with their families for approximately fifteen minutes, separated into five 3-minute snapshots. At the start of each snapshot, we noted the exhibit that each family member chose to visit. These observations allowed us to test whether our data replicated the findings from the three studies in the 1990s (Greenfield, 1995; Kremer & Mullins, 1992; Luria & Herzog, 1991), and they also allowed us to observe the movement of visitors among exhibits that aligned with their gender, those that aligned with other genders, and gender-neutral exhibits. Although the previous work on this topic categorizes activities as associated with males or females, we seek to acknowledge the ways in which gender can be nonbinary. Thus, we attempt to use language, such as “other genders” to avoid an either/or gender classification.

Method

Research Setting

Our research team collected data over the summer of 2019 at MYTown Museum, a pop-up children’s museum in Salinas, California. MYTown was designed and created by a team of community organizations led by the director of the Salinas Center for Arts and Culture and the executive director of MY Museum, the children’s museum in nearby Monterey (Herrera, 2019). The museum was located in a wide-open indoor space in the Salinas Center for Arts and Culture in downtown Salinas, which was approximately 2,950 square feet with an additional outdoor patio space. The layout was ideal for our observations because everything could be observed from one vantage point in the middle of the room. The museum was designed to represent Salinas and contained exhibits such as a library, a theatre, a grocery store, roads, and buildings. In addition, there was an outdoor area connected to the indoor area and it contained a farmer’s market, tents for camping, a Lincoln Log building area, and kayaks. There was no admission fee to enter the museum, and families could stay as long as they wanted. The museum was open in June and July of 2019 and welcomed 5800 visitors.

The setting of our research was in a largely Latinx community, and we used methods sensitive to this population. In particular, we chose not to record audio and video, and we chose not to collect any identifying information from participants as a way to reduce concerns that were present in this community in 2019 regarding immigration and deportation. Thus, we observed families without any interaction with them. One drawback of this approach is that we guessed the gender of visitors based on physical characteristics and clothes. We acknowledge the crudeness of this approach, but we did so in order to minimize any intrusions on visitors’ experience in the museum. Signage at the entrance to the museum stated that students from the psychology department at the local university were observing families to better understand child development. Gutwill (2003) found that 75% of museum visitors read such signs. In addition, we made ourselves noticeable in the museum with a white T-shirt that said CSUMB on the front and “researcher” on the back, and a few adults asked us questions about our project. We answered any questions they had. Our campus review board approved our study in May 2019 (#19-106-K38).

Participants

We observed 71 families with no contact between researchers and families. The following demographic information is based on our best estimate from our observations. There were 133 children, 50% female. One person was categorized as nonbinary. The girls’ M age = 4.03, the boys’ M age = 4.58, and both groups ranged from infancy to 10-years-old. In addition, there were 119 caregivers, 65% female. One person was categorized as nonbinary. Caregivers’ ages ranged from 20s to 60s, M = 30. There were 16 different family compositions. The most common were: one adult and one child (23%), two adults and one child (21%), two adults and two children (21%), and one adult and two children (15%). During our observations, we attempted to listen for language, and the following is the language and the percentage of families in our sample who we heard speaking each language: English (78.9%), Spanish/English (8.5%), Uncertain (7.0%), Spanish (4.2%), and other (1.4%).

Measures

We created a form built on Fasoli’s (2014) work that allowed us to record all observations by hand. It included the estimated age and apparent gender of each individual in the family units, the exhibit that each family member was in during each of five snapshots, the action that each member engaged in during each snapshot, and the language spoken by the family. For the purposes of this paper, we focus on the section of the form where we noted the exhibits that each family member visited.
Reliability
The research team established interrater reliability for the behavioral measures. First, the undergraduate coders were trained and then practiced coding in the museum for about one month before any data were collected for the current study. After that time, each of four student coders was paired with a reliability coder (i.e., the faculty member on the project). They each independently observed one family unit at the same time and filled out the observation instrument. Thus, the observations for four families were used to calculate interrater reliability. For each family there were between 30 and 48 decisions, depending on the number of people in the family, on which to agree or disagree about which exhibit each person was in. The level of agreement for each pair of observers ranged from 87% to 96%. Disagreements were discussed. The students then observed the remaining 67 families. Two of the student observers were the first and second authors on this paper. Two other student observers collected data and their data collection work was their sole contribution to the project.

Coding
Exhibits were categorized post-hoc by gender-type by the three authors on this paper. We used the categorizations of Goble et al. (2012, p. 441) from their work in Head Start preschool classrooms regarding the types of activities in which boys and girls were observed to engage. This specific study was chosen as the basis of our categorization due to the clear descriptions of each gender-typed activity based upon their observations of the children in the classrooms and play areas. Our research team categorized the exhibits by the match between the activity noted by Goble et al. (2012) and the activity afforded by each exhibit, rather than the subjective type of play that might have been observed in the museum. For example, the cornhole exhibit in the museum was coded as “masculine” because Goble et al. (2012) had an activity of “balls” in their masculine category. We deemed that the way a beanbag is meant to be thrown into the cornhole was most like a ball being thrown into a hoop or just thrown on the play yard. There was no other activity in Goble et al.’s (2012) feminine or neutral categories that fit cornhole better than “balls.” It is possible that children could have engaged in pretend play with the cornhole materials but doing so would not have meant that cornhole would have been coded as neutral. Pretend play - neutral was categorized as neutral by Goble et al. (2012). To reiterate, the coding of the exhibits was based on the activity offered by each exhibit and not by the behaviors of the visitors at the exhibits. Interrater reliability revealed that, for the 22 exhibits in the museum, all three coders agreed on the gender typing for 68% of the exhibits, and two coders agreed on the gender typing for 100% of the exhibits. Disagreements were resolved via discussion. See Table 1 for which exhibits were categorized as feminine, masculine, and neutral.

Procedure
The museum was open Thursday through Sunday. The number of family units observed on each day were Thursday = 15, Friday = 11, Saturday = 18, and Sunday = 25. Observations took place between 11 a.m. and 4 p.m. Two student researchers were stationed in a central location in the museum. The researchers wore a T-shirt with the university insignia and they held a clipboard with the data collection forms. When a group of people who appeared to be together entered the museum, one of the researchers immediately recorded demographic information for each member of the unit and started a timer for observation. The gender of each visitor was categorized by the researchers based on the visitor’s clothing and hairstyle into female, male, or nonbinary. The researchers acknowledge that this approach to coding gender is biased. Researchers used time-sampling such that they observed the family for five 3-minute snapshots for a total of 15 minutes of observation. During the first 15 seconds of each snapshot, the researcher observed the members of the unit to see the exhibits they had chosen. The researcher took about one minute or less to record the exhibit each member was in. They then waited approximately two minutes for the next snapshot to start. This was repeated for a total of five snapshots. All 71 families included in analyses stayed long enough to be observed for all five snapshots. Two other families were dropped from analyses because they left the museum before all five snapshots were recorded.

<table>
<thead>
<tr>
<th>TABLE 1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Museum Exhibits Coded Into Feminine-Stereotyped, Masculine-Stereotyped or Neutral Categories</strong></td>
</tr>
<tr>
<td>Feminine-Stereotyped Exhibits</td>
</tr>
<tr>
<td>Arts and Crafts</td>
</tr>
<tr>
<td>Grocery Store</td>
</tr>
<tr>
<td>Library</td>
</tr>
<tr>
<td>Farmer’s Market</td>
</tr>
<tr>
<td>Theater</td>
</tr>
<tr>
<td>Tree with Messages</td>
</tr>
<tr>
<td>Ribbon Weaving</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Note: “It was possible for visitors not to be in an exhibit during a snapshot. They could be in between exhibits, inside or outside. If that was the case, we identified their location with one of these locations.”
Results

Our research question asked about the extent to which museum visitors were drawn to exhibits that were stereotypically aligned with their gender. To test this question, we report our data in two ways. First, in order to directly compare our results with Herzog and Luria (1991), we report for children, during a 15-minute period, the percentage of boys and girls who visited three kinds of exhibits: feminine, masculine, and neutral. Next, we report the ways in which children and their caregivers moved about the museum. We followed them for 15 minutes and reported where they were in each of five 3-minute snapshots. Thus, for each snapshot we report the visitors’ presence at exhibits coded as feminine, masculine, and neutral, first across all visitors, and then disaggregated by age and gender.

To compare with earlier findings from the studies in the 1990s, we ran a Chi-Square test of independence comparing boys’ and girls’ visits at the wearable cars (i.e., to compare with Herzog and Luria’s automobile exhibit), the grocery (i.e., to compare to Herzog and Luria’s superette grocery exhibit), and the sand table (i.e., to compare with Herzog and Luria’s less gender-typed exhibit, the computer exhibit). MY Town did not have a computer exhibit or anything with a screen, so we chose the sand table, a neutral exhibit that involved hand manipulations similar to the hand manipulations on a computer keyboard. We found that boys and girls did not visit these exhibits equally often, $X^2 = 10.91$, $df = 2$, $p < .01$. Girls tended to visit the grocery (38% of 107 visits by girls) and the sand table (16%) more than the boys (29% and 9%, respectively, of 149 visits). The boys tended to visit the wearable cars (36%) more often than the girls (18%). These results replicated the findings of Luria and Herzog (1991).

We also examined the movement of the visitors around the museum. We wanted to know whether children and caregivers tended to stay at exhibits that were typed for their gender or whether they were likely to visit exhibits that were typed for a different gender or were neutral. Thus, we ran Chi-Square Goodness of Fit tests to see whether visitors visited the three kinds of exhibits (i.e., exhibits stereotypically for one’s gender, exhibits stereotypically for another gender, and neutral exhibits) during each snapshot. The null hypothesis for the Chi-Square Goodness of Fit test is that the frequency of visits to a particular kind of exhibit is no different from what would be expected by chance. For each snapshot, the expected frequency of visits at each of the three kinds of exhibits was 1/3 of the total number of visits. A significant Chi-Square meant that the number of visits at each kind of exhibit was different from the expected 1/3. Due to the large number of Chi-Square tests we ran, we used the Holm-Bonferroni correction to adjust the alpha level.

FIGURE 1
The Percentage of All Visitors Who Visited Each Type of Exhibit by Snapshot

Note. Due to the large number of Chi-Square tests we ran, we used the Holm-Bonferroni correction to adjust the alpha level. We report the exact alpha level and each one was below the Holm-Bonferroni correction, thus, significant at that corrected level. In each snapshot, the Chi-Square Goodness of Fit test showed that visitors did not visit the three kinds of exhibits equally often. Snapshot 1, $X^2 (2, n = 228) = 10.61, p = .005$; Snapshot 2, $X^2 (2, n = 230) = 6.11, p = .037$; Snapshot 3, $X^2 (2, N = 220) = 17.19, p < .001$; Snapshot 4, $X^2 (2, N = 215) = 8.66, p = .01$; Snapshot 5, $X^2 (2, N = 209) = 10.06, p = .005$.

FIGURE 2
The Percentage of Caregivers Who Visited Each Type of Exhibit by Snapshot

Note. The Chi-Square Goodness of Fit test showed that caregivers visited the three different kinds of exhibits equally often in snapshots 1, 2, and 4. Snapshot 1, $X^2 (2, n = 109) = 0.49, n.s.$; Snapshot 2, $X^2 (2, n = 108) = .32, n.s.$; Snapshot 4, $X^2 (2, n = 101) = 4.30, n.s.$ Caregivers did not visit the exhibits equally often in snapshots 3 and 5. Snapshot 3, $X^2 (2, n = 98) = 8.60, p = .01$; Snapshot 5, $X^2 (2, n = 99) = 8.06, p = .02$. 

Garcia, Aleman-Teweles, and Dyer-Seymour | Gender Exhibit Choices
Gender Exhibit Choices | Garcia, Aleman-Teweles, and Dyer-Seymour

In addition, we report only tests that were statistically significant per the Holm-Bonferroni correction.

Across all visitors, the results revealed that, during each of the five snapshots, family members did not visit the three kinds of exhibits equally. They were most likely to visit exhibits typed for their own gender. Snapshot 1, \( \chi^2 (2, n = 228) = 10.61, p = .005; \) Snapshot 2, \( \chi^2 (2, n = 230) = 6.11, p = .037; \) Snapshot 3, \( \chi^2 (2, n = 220) = 17.19, p \leq .001; \) Snapshot 4, \( \chi^2 (2, n = 215) = 8.66, p = .01; \) and Snapshot 5, \( \chi^2 (2, n = 209) = 10.06, p = .005 \) (see Figure 1). We then disaggregated the results by age and tested the caregivers and the children separately. The caregivers and the children showed different patterns. The caregivers visited all three kinds of exhibits equally in three snapshots: Snapshots 1, 2, and 4. In the other two snapshots, caregivers appeared to visit neutral exhibits and exhibits typed for their gender more often than exhibits typed for a different gender, Snapshot 3, \( \chi^2 (2, n = 98) = 8.60, p = .01; \) Snapshot 5, \( \chi^2 (2, n = 99) = 8.06, p = .02 \) (see Figure 2). Conversely, the children visited all the exhibits equally in just one snapshot, Snapshot 5. In the other four snapshots, children appeared to visit exhibits typed for their own gender more often than those typed for a different gender or neutral exhibits, Snapshot 1, \( \chi^2 (2, n = 119) = 17.21, p \leq .001; \) Snapshot 2, \( \chi^2 (2, n = 122) = 8.87, p = .008; \) Snapshot 3, \( \chi^2 (2, n = 122) = 14.23, p = .001; \) and Snapshot 4, \( \chi^2 (2, n = 114) = 7.00, p = .03 \) (see Figure 3).

We then disaggregated the caregivers’ and children’s results by gender. With these Chi-Square tests of Goodness of Fit, we found that the male caregivers, female caregivers, and male children more often visited the three kinds of exhibits equally, and there were just one or two snapshots where they were drawn to some exhibits more than others. Male caregivers in Snapshot 3 appeared to visit neutral exhibits most often, \( \chi^2 (2, n = 26) = 7.00, p = .03 \) (see Figure 4). Female caregivers in Snapshots 4 and 5 appeared to visit exhibits typed for their gender most often: Snapshot 4, \( \chi^2 (2, n = 68) = 11.85, p = .005; \) and Snapshot 5, \( \chi^2 (2, n = 66) = 10.09, p = .006 \) (see Figure 5). Male children in Snapshot 1 appeared to visit exhibits typed for their gender most often, \( \chi^2 (2, n = 60) = 19.20, p \leq .001 \) (see Figure 6). Conversely, the female children in Snapshot 1 visited the three exhibits equally. Thereafter, they appeared to visit the exhibits typed for their gender most often, Snapshot 2, \( \chi^2 (2, n = 61) = 12.72, p = .003; \) Snapshot 3, \( \chi^2 (2, n = 61) = 19.31, p < .001; \) and Snapshot 4, \( \chi^2 (2, n = 55) = 17.60, p < .001. \) Male children in Snapshot 1 appeared to visit exhibits typed for their gender most often, \( \chi^2 (2, n = 60) = 19.20, p \leq .001 \) (see Figure 7).

Furthermore, the family members differed from each other in terms of when during the visit they were drawn to exhibits typed for their gender and when they were drawn to exhibits typed for a different gender or neutral exhibits. Adult males visited all exhibits equally in Snapshots 1, 2, 4, and 5. It was in the middle of their...
visit, in Snapshot 3, that male caregivers were drawn to some exhibits more than others. Female caregivers visited all exhibits equally in Snapshots 1, 2, and 3, and then visited some exhibits more than others in snapshots 4 and 5. Boys visited masculine exhibits in Snapshot 1 and then visited all exhibits equally in snapshots 2–5. The opposite pattern was shown for girls where they visited all exhibits equally in Snapshot 1 and then spent time at some exhibits more than others in Snapshots 2–5.

Discussion

In this study, we observed families during their visit to a pop-up children’s museum in two different ways: we counted the number of female and male children who visited individual exhibits, and we followed family members over a 15-minute period to take note of the exhibits they visited during that time period. We counted visits at individual exhibits in order to see whether our results would replicate findings from the 1990s showing that boys and girls tended to be the most frequent visitors at exhibits that included activities that were directed toward their own gender (e.g., a grocery store for girls and cars for boys). In addition, we followed family members over a short period of time in order to test whether they tended to stay at exhibits typed for their gender or whether they visited all kinds of exhibits. The results revealed that our data in 2019 replicated the findings from the 1990s. Specifically, Luria and Herzog (1991) found that the children they observed in Boston were more likely to visit exhibits aligned with their gender, and a smaller percentage of children chose exhibits that were not aligned with their gender. Our results were the same for children in Salinas. Gender
Gender Exhibit Choices

Garcia, Aleman-Teweles, and Dyer-Seymour

schema theory (Martin & Halverson, 1981; Martin et al., 2002) can account for these results. At the same time, different results were revealed when we analyzed the visits at exhibits over time. That is, when we observed family members over time, the results revealed that adult men and women, as well as boys, most of the time visited all kinds of exhibits during a 15-minute period. Conversely, the girls we observed tended to spend most of their time throughout the 15 minutes at exhibits that only included feminine activities. In other words, most of the time the adults and the boys were just as likely to play on the make-believe road and drive the wearable cars as they were to play in the grocery or at the sand table, whereas the girls were more likely to be at certain feminine exhibits, such as the library, the grocery, and the arts and crafts table.

In addition to the consistencies and differences with previous work in museums, these results show some similarities and differences with previous work on children’s play more generally. Goble et al. (2012) found that preschool boys demonstrated less gender-typed play when their teacher was nearby, and similarly we found that the boys, at a museum with their families, demonstrated a preference to play at all kinds of exhibits rather than focus on only those exhibits typed for their gender. One of the ways in which our findings differ from previous work is that the girls, more than boys, seemed to be attracted to gender-typed exhibits. Other work on toys found that girls tended to be more open to other-gender toys, and boys were more likely to maintain gendered norms for toy selection (e.g., Weisgram et al., 2014).

It is not clear why boys and girls showed different patterns of visits to the exhibits. In more snapshots, boys visited all kinds of exhibits, whereas girls tended to focus on feminine exhibits. In other words, the boys more than the girls were more likely to explore nongender stereotyped activities. It is possible that the museum experience facilitated a certain kind of freedom for both boys and girls, and the freedom was realized differently for each group. For many of the girls, it meant that they did not have to feign interest in masculine activities. For many of the boys, it meant that they did not have to eschew feminine activities in the ways they might in school with same-gender peers. To be clear, not every boy and girl showed these patterns. The majority of the girls and boys showed these patterns, and others did not behave according to these patterns.

A second possibility for our findings is that the exhibits that were typed for females might have been the most interesting or novel exhibits, and the exhibits typed for males might not have been as interesting or novel. After the first snapshot where boys were attracted to exhibits typed for males, they might not have seen any other gender-typed exhibits that were as exciting. Further, for the girls, it is possible that the exhibits typed for males as well as the ones that were neutral were not appealing, beyond the fact that they included stereotypically male activities.

Observing children and their families in the field is a strength of this study. We were able to observe the behavior of families in a real-life setting and the participants could be observed with few demand characteristics affecting the results. Our study’s findings may generalize to other children’s museums because of the field design method, however, caution is warranted, and researchers should attend to the specific location and population served by any museum. The field design method has been employed by other researchers whose aim was to examine child development in an environment outside the home, school or laboratory, and it comes with special rewards and challenges (Callanan, 2012).

The field setting indeed made for some challenges to our observations. One challenge is that we were observing a family activity where family members most likely influenced each other’s choices about which exhibits to visit. Early on we attempted to code who appeared to be directing the action, but we could not reach a good level of agreement across observers and so we abandoned that code. Another limitation to the field observation and our desire to not interact with the visitors was that we guessed visitors’ gender. To protect visitors’ identities and reduce concerns about anonymity in a time when immigration raids were feared, we did not ask them questions about how they would characterize themselves in terms of social class, cultural background, or gender. Instead we observed from afar and made guesses about visitors’ gender from dress and hairstyle. This is admittedly crude. Furthermore, we coded the exhibits according to stereotypic, gendered preferences, albeit preferences that Goble et al. (2012) observed in children’s play in a Head Start in the early 2000s.

Additionally, we acknowledge the limitation of having two of the three authors of this paper serve as observers for data collection as well as coders for the gender of the exhibits. The observers noted which exhibit each family member was in during each snapshot. For those times when visitors appeared to be near two exhibits, it is possible that the observer could have placed them in the exhibit that aligned with the observer’s biases about where people of different genders should spend their time. At the same time, we did not code the exhibits as aligning with one gender or another until after we had collected all of our data. Thus, we did not know when we were observing which exhibit would be coded as feminine, masculine, or neutral. Categorizing the exhibits post-hoc might have reduced biases during...
Another limitation is that we did not account for the type of play that the children actually engaged in while interacting with the exhibits. Rather, we only recorded the exhibit they visited. Our choice to observe visitors from a respectful distance resulted in fewer opportunities to clearly document their play and conversation. It is possible that children's play in the exhibits was gendered, even if they were in an exhibit that was coded as not aligning with their gender. Indeed, Kremer and Mullins (1992) found that, although there was no difference in the number of boys and girls who visited the different exhibits, the researchers did observe differences in the girls' and boys' behavior at the exhibits. For example, at a water jets exhibit, 28% of boys and 1% of girls pretended the water jets were guns. At an animal lab exhibit, 35% of girls and 5% of boys showed "nurturing behavior" toward the baby chicks.

This study provided evidence that children's museums can be a space for children to play and explore in different ways. Further studies on the topic should include information about the type of play that children were engaging in within the exhibits in order to build on the work of Kremer and Mullins (1992). Such studies might clarify the differences we found in boys' tendency to explore across different gender-typed exhibits and girls' tendency to focus on own-gender-typed exhibits. In addition, it would be nice to see experimental studies that manipulate different aspects of the exhibits in the museum, similar to those done with toys (e.g., Weisgram et al., 2018) where color is manipulated to test whether it is the color or the toy's activity that entices different children to play with it. For example, what if the wearable cars exhibit had included a problem that needed to be solved in addition to the opportunity to go fast around the track? Would such an addition have enticed different children to engage with the exhibit compared to when there was no problem to solve?

In conclusion, although there have been some advances in gender equity in American society since 1990, including the first ever female vice-president of the United States elected in 2020, there is more work to be done. Children's museums have a unique position as informal learning spaces with possibly less red tape to go through to institute changes compared with formal, public schools. Studies such as ours and others (Fasoli, 2015) that followed visitors throughout their visit bring more information to exhibit designers about how visitors are engaging with the museum as a whole. Knowing that boys visited all kinds of exhibits and girls seemed to focus on those that included activities that were stereotypically associated with their own gender is a first step to thinking about how to broaden the experience for visitors. For example, Greenfield (1995) suggested that including activities appealing to "both sexes," (p. 934; we would update her words to say "appealing to people across the gender spectrum," in the same exhibit with docents to guide children's actions, as suggested by Greenfield (1995) as well as Kremer and Mullins (1992), is one way to increase the chances that everyone will see an opportunity to engage with all the exhibits. Including everyone and increasing their opportunities are both parts of the big puzzle of how we increase the diversity of teams of people focused on solving the world's big problems.

There is much work to be done to understand the preferences of children and the ways in which parents, peers, and society influence those preferences. In this study and in others, gender was considered from a binary perspective. It is our goal to move toward a less binary framework and toward a more open approach to children's preferences for and play in museum exhibits. Doing so may open opportunities for all genders to have equitable experiences, develop well-rounded skill sets, and pursue interests that spark their passions as unique individuals.

**References**


Gender Exhibit Choices

Garcia, Aleman-Teweles, and Dyer-Seymour


https://doi.org/10.1006/psob.1991.1048

Author Note
Natassia Aleman-Teweles
https://orcid.org/0000-0001-7237-1300
Jennifer Dyer-Seymour
https://orcid.org/0000-0003-4670-5667

This study began in a capstone course in psychology in June 2019. Funding was later provided by California State University, Monterey Bay’s Undergraduate Opportunities Research Center (UROC) and their grant, U.S. Department of Education Hispanic-Serving Institution Grant #P031C160221. We thank Enid Ryce and Lauren Cohen for making MY Town a reality! We thank Barry Ngu and April Regalado for their help with data collection. We thank Maureen Callanan for her expert feedback on this manuscript. The authors have no conflicts of interest to disclose.

Correspondence concerning this article should be addressed to Jennifer Dyer-Seymour, California State University, Monterey Bay, 100 Campus Center, Seaside, CA 93955. Email: jdyer-seymour@csUMB.edu
**Mental Health Symptoms Predicting American College Students’ Academic Performance: The Moderating Role of Peer Support**

Gabrielle M. Goselin and Nicolette P. Rickert*
Department of Psychology, Georgia Southern University

**ABSTRACT.** For the current study, we examined the effects of anxiety and depressive symptoms on college students’ academic performance, as moderated by peer support. An online survey was completed by 174 college students (75.3% women, 57.3% White) at a university in the southeast United States, targeting their symptoms of anxiety and depression, perceptions of peer support, and self-report GPA. Utilizing multiple regression analyses to test for moderation, it was found that, although anxiety and depression each negatively predicted students’ college GPA, these connections were moderated by perceived peer support. When students perceived having higher levels of peer support, this buffered against the negative effects of anxiety and depression on their GPA. These findings can contribute to our knowledge of how to support college students’ academic performance in the face of mental health experiences.

**Keywords:** anxiety, depression, academic performance, peer support, and college students.

The typical age range of college students (18–25) is when most individuals first experience distress due to a mental health-related concern (Antunes & Langmuir, 2021). In fact, almost 15–23% of college students with mental health problems have said their disorders have negatively impacted their academic performance (Bruffaerts et al., 2018). The purpose of this study was to determine whether students who experience symptoms of mental health concerns perform better in their academics when supported by their peers. We specifically looked at anxiety and depression given the high prevalence rates of the two diagnoses in college-aged students (Asher BlackDeer et al., 2021; Hart Abney et al., 2019; Ketchen Lipson et al., 2021). According to the American Psychological Association (n.d.), anxiety is “an emotion characterized by worried thoughts and tense feelings.” The National Alliance on Mental Illness (2022) further describes a variety of symptoms related to anxiety (i.e., restlessness, irritability, sweating, stomach clenching) that impact an individual’s day-to-day functioning. Depression is a common mental disorder typically characterized by feelings of worthlessness and a lack of interest in activities (American Psychological Association, n.d.). According to NAMI, depression is a “devastating” disorder that has the potential to only occur once in an individual’s lifetime or to recur periodically. There have been a variety of conceptualizations for peer support, usually defined either as perceived social support (i.e., having the perception that one will have support if one needs it) or received social support (i.e., objectively receiving the actual support itself when needed; Altermatt, 2019). The current study draws upon both these definitions in examining the connections between college students’ mental health, peer support, and academic performance.

**Factors That Impact Anxiety and Depression in College Students**

*College as a Period of Transition*

One of the major reasons why college students, in particular, experience high rates of anxiety and/or depression is because they go through an overall period of transition, particularly first-year students and seniors (Beiter et al., 2015; Mahmoud et al., 2015). This transition is one filled with many “firsts” and new experiences. One of those being moving out of the family home and into the “real world,” or emerging into adulthood, which is typically the onset of mental health concerns (Duffy et al., 2020).

There are even more specific transitions that take place for a college student. Attending college is a major
transitions and changes, such as moving out of the family home and learning how to take care of oneself. A combination of these factors can affect college students’ mental health.

**Academics as a Stressor**

In addition to the multiple transitions associated with entering college, the stress and pressure of college-level academics can also take a toll on the mental health of college students. Compared to their time in high school, college students experience an increase in their academic demands and workload (Saeed et al., 2018), which according to the World Health Organization is a risk factor for mental health disorders because it tends to lead to academic failure (Bahmani et al., 2018). An increase in the prevalence of anxiety and depression was also found to be due to the fact that colleges are becoming more competitive and demanding than they once were (Alsubaie et al., 2019). In fact, 33% (N = 461) of university student participants were found to have experienced depressive symptoms (Alsubaie et al., 2019). Another study was able to identify that one of the top 10 sources of stress for college students was academic performance (Beiter et al., 2015). All participants (N = 374) identified academic performance as a top-10 stressor, and 40.1% of them went on to report academic performance as an extreme stressor. It is also important to note that the Depression Anxiety Stress Scale-21, used in the study, was not given to participants during midterm or final exam periods to account for the extra stress during these times (Beiter et al., 2015).

**Lack of Coping Strategies**

One way college students can respond to these multiple transitions and stresses is to adopt appropriate coping skills (Al-Qaisy, 2011). The Transactional Model of Coping addresses one form of appropriate coping skills (Lazarus & Folkman, 1984), and Mahmoud et al. (2012) expanded on this model and described how two different forms of coping are identified in the model: adaptive and maladaptive. Adaptive coping is the most appropriate and it is when someone names the source of stress and tries to find a solution to the stressor. On the other hand, maladaptive coping is when someone tries to ignore the stressor and decides to not find a solution. Mahmoud et al. (2012) had 508 undergraduate participants complete the Brief COPE Inventory, which consist of 28-items related to adaptive and maladaptive coping strategies. It was found that participants who reported using more maladaptive coping strategies (e.g., denial and substance abuse) experienced higher levels of depression compared those who used more adaptive coping strategies (e.g., planning and positive reframing). Mahmoud et al.
found that maladaptive coping strategies also significantly predicted symptoms of anxiety. There is some evidence that college students’ use of adaptive coping strategies may increase over time (Wongtongkham, 2019), but even so, college-aged students are more likely to use maladaptive coping strategies (Mahmoud et al., 2015), putting them at greater risk.

**Peer Support and Mental Health**

As stated earlier, one of the major transitions college students experience is a shift in their social life (Al-Qaisy, 2011). Typically, college students come into their new schools with fewer friends than they once had in high school, and this lack of a social environment has been linked to higher prevalence rates of anxiety and/or depression (Eldeleklioglu, 2006). Two models take into account the effects of social support on mental health: the Main Effect Model and the Buffering Model. The Main Effect Model states that the relationship between social support and mental health is direct. Eldeleklioglu (2006) explained how the model shows that, when a person receives positive social support, this creates stability within their life and lessens the negative impact of mental health symptoms. On the other hand, the Buffering Model shows how social support actually serves as a preventive measure against detrimental health concerns, such as anxiety and/or depression. In this model, social support actually decreases the negative consequences.

The impact of different forms of social support, such as peer support, on mental health symptoms was examined in another study (Alsubaie et al., 2019). The Multidimensional Scale of Perceived Social Support was used to assess the 462 participants’ perceived levels of social support in this study. Alsubaie et al. (2019) determined that social support was a negative predictor of the onset of symptoms for depression, meaning that, as social support increased, depressive symptoms decreased. Peer support (i.e., the social support you receive from friends) was also found to be the most effective for college-aged participants out of the different forms of social support (Alsubaie et al., 2019).

More recently, studies have examined how peer support predicts college students’ mental health during the COVID-19 pandemic. In a review of the current literature, Suresh et al. (2021a) found that peer support positively bolstered mental health symptoms for college students throughout the pandemic. Further, it is important that college students form relationships with their peers because it can decrease the chance of developing a mental health problem in the future (Beiter et al., 2015). For example, Suresh et al. (2021b) found that students who utilized a student-led peer support center felt understood, better equipped to face their challenges, and that their mental health was well-supported. However, students who already experience symptoms of anxiety struggle to form relationships with their peers (Mahmoud et al., 2015). Thus, it is important for universities to encourage peer support on their campuses to help reduce students’ current and future mental health concerns (Beiter et al., 2015).

**Factors That Impact Academic Performance in College Students**

**Mental Health Concerns**

Additionally, mental health concerns have the potential to impact college students’ academic performance. Luca et al. (2016) found that college students who experience symptoms of mental health disorders are at an increased risk of performing poorly in their academics, especially first-year students and sophomores. More specifically, those students, whether upperclass or lowerclass students, with a higher prevalence of mental health symptoms had lower GPAs compared to their peers. It is important to note that Luca et al. (2016) found significant results for participants with higher severity rates. The results are consistent with what Antunes and Langmuir (2021) and Asher BlackDeer et al. (2021) noted in their studies, which is that mental health concerns are related to lower academic performance.

In another study, Svanum and Zody (2001) collected a sample of 169 participants who met the criteria for an Axis I disorder, which includes anxiety and depression, and found that 142 students withdrew from their courses entirely. Based on this observation, it is understood that college students who experience higher rates of anxiety and/or depressive symptoms perform more poorly in their academics (Svanum & Zody, 2001). For college students to minimize the impact of stress due to academics, it is important that they find a way to create a positive attitude in response to their academics (Beiter et al., 2015).

A decrease in college students’ GPA was also found in Bruffaerts et al. (2018) study. The Global Appraisal of Individual Needs Short Screener was used to determine any existing mental health concerns within the study’s 4,921 participants. The study determined that 23.7% of participants were experiencing internalizing mental health symptoms, which include those of anxiety and/or depression. Bruffaerts et al. (2018) also found that first-year college students across 38 departments had a decrease in GPA of 0.2 to 0.3 points. They concluded that symptoms of mental health can have far worse consequences for a college student than only poor academic performance; they can actually lead to an increase in risk for dropping out entirely. Although the transitional
period can take a toll on college students’ mental health, which further results in negative consequences within academic performance, peer support may reduce these negative consequences.

**Peer Support as a Solution**

Peer support can also have a major impact on academic performance (Altermatt, 2019). One of the main explanations for this is that peers can relate to the stress of academics compared to any other form of social support (Jones et al., 2018), such as parental support. Altermatt (2019) showed that the quality of the peer support is far more important than the peer support itself: For students to perform well academically, it is important that they feel supported when they are both struggling and succeeding.

Rosenkranz (2012) further showed how a mentor program that was developed at a university for first-year students had a moderate impact on improving their academic skills. These students were led through a 12-week mentoring program by senior students in their final year. The students participated in small-group workshops that ranged in topics and a brief mentoring session at the end of the semester to recap what they had learned. Overall, students felt that their academic writing skills were improved but that they required more time to improve all other academic skills (Rosenkranz, 2012). Therefore, peer support has the potential to positively support college students’ academic performance as well as their mental health.

**The Current Study**

Although extensive research has examined how anxiety and depression predict academic performance (e.g., Antunes & Langmuir, 2021; Bruffaerts et al., 2018; Luca et al., 2016) and how peer support predicts academic performance (Altermatt, 2019; Rosenkranz, 2012), limited research has explored the interaction of the two in predicting college students’ academic performance. More specifically, a dearth of research has examined how peer support may interact with students’ mental health to predict their academic performance. For the current study, we aimed to look at whether the relationship between mental health (anxiety and/or depression) and academic performance (GPA) depended on the level of college students’ perceived peer support (e.g., availability, assistance, and reliance on peers). This study evaluated the role of peer support in increasing academic performance for students with mental health symptoms through a correlational, single time-point design. Based on previous research and theory, we hypothesized that students who experience a lack of peer support would also experience greater symptoms of anxiety and/or depression, and therefore perform lower in their academics. We also hypothesized that, when students who experience symptoms of anxiety and/or depression perceive having higher levels of peer support, it would buffer their symptoms and positively impact their academic performance.

**Method**

**Participants**

A total of 174 undergraduate students at a public university in southeast Georgia participated in the current study and completed the survey in full. Of this sample, 134 participants identified as women (75.3%), 39 participants identified as men (21.9%), and 1 participant identified as nonbinary/third gender (0.6%). The participants were college-aged students ($M = 19.19$, $SD = 1.22$). More than half of the participants identified as White/European American ($N = 102, 57.3$). The remaining participants identified as follows: Black/African American ($N = 51, 28.7$%), Hispanic ($N = 9, 5.1$%), and Other ($N = 8, 4.5$%). The sample mainly consisted of first-year students ($N = 92, 51.7$%) and sophomores ($N = 56, 31.5$%) but did include some juniors ($N = 18, 10.1$%) and seniors ($N = 5, 2.8$%).

**Measures**

To measure academic performance, participants were asked to self-report their GPAs. The responses ranged from 0.21 to 4.00 ($M = 3.12$, $SD = 0.74$). Anxiety and depression were measured by using the Depression Anxiety Stress Scale-21 (Lovibond & Lovibond, 1995), which is a shortened version of the original 42-item scale. Participants were shown a prompt that asked them to indicate how much the statement applied to them over the past month. Example survey statements include, “I felt that I had nothing to look forward to” and “I felt I was close to panic.” The Depression and Anxiety subscales were the only ones analyzed in the study. Participants rated each item on a 4-point scale from 0 (Did not apply to me at all) to 3 (Applied to me very much or most of the time). Levels of depression ranged from 0 to 21 ($M = 6.14$, $SD = 5.14$) with lower scores indicating lower levels of depression. Participants’ levels of anxiety also ranged from 0 to 21 ($M = 5.19$, $SD = 5.22$) with lower scores indicating lower levels of anxiety. Both the Depression and Anxiety subscales demonstrated good internal reliability ($\alpha = .91$ for both subscales); this is consistent with previous studies which have found good reliability and validity for these scales ($\alpha = .82–.97$; Lovibond & Lovibond, 1995).

The study measured peer support by utilizing the Perceived Social Support Friend Scale (Procidano &
Heller, 1983), which consisted of 20 items. Example survey items included, “My friends are good at helping me solve problems” and “I rely on my friends for emotional support.” Participants responded by selecting either “Yes,” “No,” or “Don’t know.” Levels of perceived peer support ranged from 0 to 20 (M = 13.46, SD = 5.32) with higher scores indicating higher levels of peer support. Peer support demonstrated good internal reliability (α = .87) consistent with previous reliability and validity estimates for this measure (α = .88; Procidano & Heller, 1983). All descriptive statistics for each measure can be found in Table 1.

**Procedure**
The Georgia Southern Institutional Review Board approved the study prior to the start of data collection. The survey was made available to students from January 2022 to March 2022, during which the university was fully reopened from the COVID-19 pandemic with limited social distancing and mask guidelines. Depending on their classes and course schedules, participants might have completed the survey before, during, or after midterm exams. A computer or laptop was needed to complete the survey. Participants were able to find the survey through SONA (an online system to find and complete the survey). Participants were able to find the survey through SONA (an online system to find and sign up to participate in studies). A Qualtrics link was provided to potential participants to take the survey, and they began the survey by reading an informed consent form and some background information on the study. After consenting, the participants completed several survey questions. Finally, when the survey was completed, the participants were provided with a brief debriefing page and contact information for the university’s counseling center and a Crisis Hotline.

**Results**
Preliminary analyses were conducted to determine whether GPA, depression, anxiety, and peer support correlated with one another before running a multiple regression moderation analysis. All data was analyzed using Pearson’s correlation coefficient. Correlations between all variables were statistically significant and the pattern of interrelations was as expected (see Table 2). Anxiety and depression were both positively related to one another, but each was negatively correlated with peer support and GPA. In other words, the more depressed students were, the more anxious they also tended to be; and the more depressed or anxious students were, the lower their levels of peer support and GPA were. Peer support, however, was positively correlated with students’ GPA; as participants’ perceived levels of peer support increased, their GPA also increased.

To address the main hypotheses of the current study, a series of multiple regression analyses were conducted (see Tables 3 and 4). First, a multiple regression model was analyzed with depression, peer support, and the interaction between these two variables (Depression x Peer Support) as predictors of college students’ GPA. Peer support was re-coded as a categorical variable based on their classes and course schedules, participants might have completed the survey before, during, or after midterm exams. A computer or laptop was needed to complete the survey. Participants were able to find the survey through SONA (an online system to find and sign up to participate in studies). A Qualtrics link was provided to potential participants to take the survey, and they began the survey by reading an informed consent form and some background information on the study. After consenting, the participants completed several survey questions. Finally, when the survey was completed, the participants were provided with a brief debriefing page and contact information for the university’s counseling center and a Crisis Hotline.

**Results**
Preliminary analyses were conducted to determine whether GPA, depression, anxiety, and peer support correlated with one another before running a multiple regression moderation analysis. All data was analyzed using Pearson’s correlation coefficient. Correlations between all variables were statistically significant and the pattern of interrelations was as expected (see Table 2). Anxiety and depression were both positively related to one another, but each was negatively correlated with peer support and GPA. In other words, the more depressed students were, the more anxious they also tended to be; and the more depressed or anxious students were, the lower their levels of peer support and GPA were. Peer support, however, was positively correlated with students’ GPA; as participants’ perceived levels of peer support increased, their GPA also increased.

To address the main hypotheses of the current study, a series of multiple regression analyses were conducted (see Tables 3 and 4). First, a multiple regression model was analyzed with depression, peer support, and the interaction between these two variables (Depression x Peer Support) as predictors of college students’ GPA. Peer support was re-coded as a categorical variable based on their classes and course schedules, participants might have completed the survey before, during, or after midterm exams. A computer or laptop was needed to complete the survey. Participants were able to find the survey through SONA (an online system to find and sign up to participate in studies). A Qualtrics link was provided to potential participants to take the survey, and they began the survey by reading an informed consent form and some background information on the study. After consenting, the participants completed several survey questions. Finally, when the survey was completed, the participants were provided with a brief debriefing page and contact information for the university’s counseling center and a Crisis Hotline.

**Results**
Preliminary analyses were conducted to determine whether GPA, depression, anxiety, and peer support correlated with one another before running a multiple regression moderation analysis. All data was analyzed using Pearson’s correlation coefficient. Correlations between all variables were statistically significant and the pattern of interrelations was as expected (see Table 2). Anxiety and depression were both positively related to one another, but each was negatively correlated with peer support and GPA. In other words, the more depressed students were, the more anxious they also tended to be; and the more depressed or anxious students were, the lower their levels of peer support and GPA were. Peer support, however, was positively correlated with students’ GPA; as participants’ perceived levels of peer support increased, their GPA also increased.

To address the main hypotheses of the current study, a series of multiple regression analyses were conducted (see Tables 3 and 4). First, a multiple regression model was analyzed with depression, peer support, and the interaction between these two variables (Depression x Peer Support) as predictors of college students’ GPA. Peer support was re-coded as a categorical variable based
on tertile splits of high (peer support = 2), medium (peer support = 1), and low (peer support = 0) levels of peer support. All predictors were centered (depression was mean-centered, peer support was centered at the low group). The overall model was statistically significant, \( F(3, 162) = 5.31, p = .002 \). Depression was a statistically significant predictor of GPA (\( \beta = -.41, p = .001 \)), but peer support was not a statistically significant predictor (\( \beta = .05, p = .52 \)). However, the interaction between depression and peer support was found to be a statistically significant predictor of GPA (\( \beta = .31, p = .006 \)). As can be seen in Figure 1, peer support buffered against the negative connections between depression and GPA for college students. For highly depressed students, low levels of peer support were associated with lower levels of GPA. However, high levels of peer support corresponded with higher GPAs. This suggests that peer support, for this sample of college students, may help students with depression remain successful in school.

A multiple regression moderation analysis was also conducted with anxiety, peer support, and the interaction between these two variables (Anxiety x Peer Support) as predictors of college students’ GPA. All predictors were re-coded and centered as in the previous model. The overall model was statistically significant, \( F(3, 162) = 6.83, p < .001 \). Anxiety was a statistically significant predictor of GPA (\( \beta = -.48, p < .001 \)), but peer support was not a statistically significant predictor of GPA (\( \beta = .07, p = .393 \)). Although, the interaction between anxiety and peer support was found to be a statistically significant predictor (\( \beta = .39, p = .001 \)). Figure 2 depicts this statistically significant interaction, which was a buffering interaction like in the previous model. In other words, students who were highly anxious and had high levels of peer support had higher GPAs than their similarly anxious peers who had low levels of peer support. Again, these findings suggest the benefits of peer support in potentially offsetting the negative impacts of anxiety on students’ academic performance.

**Discussion**

**Summary of Findings**

For the current study, we sought to examine the connections between mental health, peer support, and college students’ academic performance. Based on previous research, we hypothesized that peer support would buffer against the negative associations between anxiety and/or depression and academic performance. The results of the study did indicate support for these hypotheses. In line with previous research, we found that GPA and feelings of depression were negatively related, meaning that as participants’ GPA increased in value, their feelings of depression decreased (e.g., Antunes & Langmuir, 2021; Bruffaerts et al., 2018; Luca et al., 2016). However, peer support buffered this connection. Students who perceived having higher levels of peer support, this buffered against the negative effects of depression on academic performance. Although college students’ mental health can threaten their performance in school, support from peers may help to reduce this threat. Across both depression and anxiety, students who
perceived higher levels of support from their peers had greater academic performance in school, even in the face of challenges to their mental health.

Limitations and Recommendations
In addition to adding evidence on the buffering role of peer support, the current study has several limitations that should be considered in future work examining college students’ mental health and academic performance. The current study used self-report scales for all measures, including participants’ GPA, so there is no way to know if these reports were accurate. Although the current study was unable to, we recommend that future researchers verify participants’ grades and GPA through official college records or transcripts, and consider expanding beyond GPA as the only marker of academic performance to also include other relevant constructs, such as dropout and class withdrawal rates. The study was also unable to verify if participants were actually experiencing anxiety and/or depression or a number of symptoms of these conditions. It is recommended that future researchers analyze data from participants with clinically diagnosed anxiety and/or depression.

Another limitation was that the tertile split levels of peer support re-coded as a categorical moderating variable were sample specific (upper third of sample, middle third of sample, lower third of sample) rather than scale specific (objective values of high, medium, and low levels of peer support). Sample-specific splits allowed for the examination of three distinct groups of peer support with roughly equal numbers of participants across the groups. However, these three groups may not generalize to other samples in other studies. We recommend that future researchers consider using a more objective categorization for levels of peer support that are scale specific and will more readily generalize to other samples.

With regard to generalizability, the study’s participants were also limited to only college students at a single university in southeast Georgia who had access to SONA, most of whom were mainly psychology majors. Future research should consider expanding the study across multiple universities and several different majors to include a more diverse sample. The study’s participants were also limited to traditional college students and did not explicitly sample from those students who might identify as nontraditional (e.g., enrolling several years after high school graduation, commuter). Future studies should consider examining the effects of mental health symptoms and the impact of peer support on nontraditional college students. In addition, research with more diverse samples across race, ethnicity, gender, age, generation status, socioeconomic status, commuter status, or ability status (i.e., physical and mental disabilities) could help inform colleges on the unique challenges and educational barriers students face, and the ways in which peer support/peer support programs can promote the mental health and academic success of students from varied backgrounds and experiences. For example, a more diverse sample would not only allow the results of the study to be more readily generalizable, but could also allow for the examination of how same or cross-ethnic peer relationships might buffer symptoms of mental health (e.g., Kawabata & Crick, 2015).

Because data was collected during the COVID-19 pandemic (although after the university fully reopened with limited social distancing or mask guidelines), the interconnections between college students’ mental health, peer support, and academic performance could be further exacerbated by the effects of the pandemic in this sample. This could also impact the generalizability of findings to pre- or postpandemic college students’ experiences. Initial research suggests that college students’ mental health (e.g., anxiety, burnout, stress), academic motivation (e.g., engagement, attention, self-efficacy), and academic performance have been negatively impacted by the pandemic, transitions between in-person and virtual learning, and exacerbated further by socioeconomic status and less access to resources (Garris & Fleck, 2020; Gonzalez-Ramirez et al., 2021; Hicks et al., 2021; Katz et al., 2021; Kinsky et al., 2021; Means & Neisler, 2021; Ober et al., 2021; Tasso et al., 2021; Usher et al., 2021; von Keyserlingk et al., 2021). Future studies should continue examining how mental health and peer support affect college students’ academic performance before, during, and after the effects of the pandemic have dissipated.

Finally, the current study drew upon a correlational, single time point design which did not allow us to determine the temporal ordering of predictors and outcomes or how these patterns of relations unfold over time throughout the college experience. Future research should consider completing a longitudinal study with multiple check-in periods to have a more accurate understanding of how mental health symptoms predict academic performance and the buffering role of peer support.

Implications
The current study’s results indicate that college students who perceived having support from their peers and who also experienced symptoms of anxiety and/or depression were able to maintain higher levels of academic performance than their peers who did not perceive having peer support. These results show some of the benefits that peer support can have for students that experience
mental health symptoms, especially in buffering the commonly seen negative connections with GPA and academic performance. In particular, it seems that both perceiving peer support (i.e., “My friends give me the moral support I need” and “There is a friend I could go to if I were just feeling down, without feeling funny about it later”) and actually receiving peer support (i.e., “My friends are good at helping me solve problems” and “I’ve recently gotten a good idea about how to do something from a friend”) are both valuable when it comes to buffering against the negative connections between mental health and academic performance. As mentioned earlier, research on the interaction of anxiety and/or depression symptoms with peer support in predicting college students’ academic performance is limited. The findings of this study denote the importance of college students’ perceived levels of peer support and the need for further research on these interactions.

The current study’s findings will also hopefully educate college administrators and leadership about the importance of providing students with different opportunities to form connections with their peers. Based on the findings of this study, colleges may want to consider implementing mentoring or peer support programs like the ones examined by Rosenkranz (2012) and Suresh et al. (2021b) in order to buffer against the negative connections between mental health symptoms and academic performance. Finally, the findings and limitations of the current study demonstrate that college students’ mental health, peer support, academic performance, and the interactions between these constructs should be further investigated in order to best support college students’ mental and academic well-being.

References


Mental Health, Peer Support, and Academic Performance

Gabrielle M. Goselin

12 https://orcid.org/0000-0003-1077-6446

Gabrielle M. Goselin is now at the Counseling and Psychological Services Department at Georgia State University, Atlanta, GA.

We have no conflicts of interest to disclose.

Correspondence concerning this article should be addressed to Gabrielle M. Goselin, Department of Counseling and Psychological Services at Georgia State University, 30 Pryor Street SW, Atlanta, GA, 30303, United States. Email: ggoselin1@student.gsu.edu

Author Note

Copyright 2022 by Psi Chi, the International Honor Society in Psychology (Vol. 27, No. 4, ISSN 2325-7342)
The Predictive Ability of Early Maladaptive Schemas for Aggression
Derek A. Lavoie and Elizabeth A. Harwood
Department of Psychology, Rivier University

ABSTRACT. The presence of aggression is an enduring concern that affects many domains of one's life. In those with repeated aggressive behavior, it is especially difficult to treat. Identification and treatment of underlying factors is pivotal to reduce aggression. Schema therapy offers an effective approach to treating these underlying elements, which are conceptualized as early maladaptive schemas (EMS). The confirmation of EMS’ prediction of aggression enables effective identification and treatment through this therapeutic approach. Adult participants completed the Young Schema Questionnaire – S3, Personality Inventory for the DSM-5 – Brief Form, and the Buss Perry Aggression Questionnaire. Results for a general sample of 124 participants supported that total EMS, EMS domains, and personality traits were all predictive of aggression. EMS were the most predictive in the aggression domain of hostility, $R^2 = .37, p < .001$. Personality traits were found to be more predictive than EMS for aggression, $R^2 = .76, p = .001$. However, the use of schema domains provides a more specific understanding of both the manifestation of aggression and translates directly to treatment through schema theory.

Keywords: early maladaptive schemas, personality traits, aggression, schema therapy

Individuals who exhibit aggression are likely to suffer from impaired behavioral functioning and negatively impact those around them (Gilbert & Daffern, 2011). In fact, aggressive behavior is significantly correlated with increased substance use, involvement in deviant behaviors, and risky sexual practices (Nouvion et al., 2007). Aggressive individuals are more likely to have lower self-worth with higher levels of depressive symptoms (Undheim & Sund, 2010), an increased risk of suicide (Helfritz & Stanford, 2005), and increased engagement in intimate partner violence (Próspero, 2007). Thus, it is important to understand and predict aggression.

Aggression Theory
Aggression is comprised of behaviors that pose significant risk to both the acting individual and those with whom they interact. These can involve physical or emotional harm, increased economic costs, and elevated likelihood of psychological difficulties in victims or witnesses (Flanigan & Russo, 2019). Although multiple approaches to the conceptualization of aggression exist, the most encompassing theory in contemporary research is DeWall et al.’s (2011) General Aggression Model. This model is based on multiple “mini-theories” and combines them into a single framework. It asserts that aggression arises as a result of biological factors, personality development, social processes, cognitive processes, short-term/long-term processes, and decision-making factors. The theory utilizes a three-stage cycle, first identifying the inputs of the person and the situation. This leads to one's present internal state, whether it be cognitive, arousal, or affect. The third stage is the outcome of these processes and the subsequent decision-making process (DeWall et al., 2011).

Other theories place more specific focus on the cognitive development of aggression and are useful for later considerations of intervention. Aggression scripts involve observation and repeated action as children. Through the repetition of aggression, it becomes habituated and more likely to perpetuate. When aggression scripts combine with normative beliefs, the individual...
Early Maladaptive Schemas

EMS are a useful concept to conceptualize one's underlying belief systems, as they are related to aggression. They are a central concept within schema therapy; an approach developed by Young (Young et al., 2006). EMS are patterns or fixed belief systems comprised of memories, emotions, cognitions, and sensations. They often develop as the result of negative experiences, typically during childhood or adolescence, and serve to fulfill an “unmet need” related to these experiences. The individual then interprets future situations through the lens of the schema. Maladaptive behaviors develop as the schemas are activated but are not exclusively included within the schema itself (Young et al., 2006). There are eighteen EMS categorized in the domains of disconnection/rejection, impaired autonomy/performance, impaired limits, other-directedness, and overvigilance/inhibition (Young et al., 2006).

Young et al. (2006) identified that certain aggressive behaviors can be influenced, and possibly explained, by certain schemas. Each EMS has three potential coping styles that are important for conceptualizing how aggression can arise. Overcompensation is when the individual fights the schema, doing the opposite of its influence. Surrender is when one gives up and accepts the effects of a schema. Avoidance is when one avoids any situation that would evoke a given schema.

In essence, schema theory is interested in finding the underlying drives that influence the individual’s maladaptive behavior. The theory posits that the recognition of these core belief systems and addressing their initial formations will lead the individual to healthier functioning. Aggression is a form of maladaptive behavior that would benefit from an approach following this model (Young et al., 2006). Schema theory and its use of early maladaptive schemas is a concept that demonstrates efficacy across diverse populations in multiple languages, which has been supported in several studies (Baranoff et al., 2006; Calvete et al., 2013; Corral & Calvete, 2014; Estévez et al., 2016; Masley et al., 2012; Mokhtarinejad et al., 2020; Wells, 2007).

The Relationship of EMS to Aggression

Because schema therapy asserts that maladaptive behaviors develop in response to schemas, they can be utilized to gauge the manifestation of these behaviors. Research has supported the role of EMS with maladaptive behavior in a multitude of specific populations. EMS in the disconnection and rejection domain demonstrated significant mediation effects with victims of child sexual abuse and displaced aggression (Estévez et al., 2016). In men seeking residential substance use treatment, EMS in the impaired limits domain were positively correlated with verbal aggression and aggressive attitudes. EMS in the disconnection and rejection domain were positively associated with physical aggression (Shorey et al., 2015).

In individuals diagnosed with borderline personality
disorder, disconnection/rejection EMS predicted increased suicidal ideation, physical aggression, dissociative symptoms, and eating disorders (Frias et al., 2017). Adolescents exposed to family violence were found to exhibit schemas in the domain of disconnection/rejection, which were predictive of dating violence (Calvete et al., 2018).

Research in schema therapy and anger found that EMS of abandonment/instability, mistrust/abuse, all schemas in the disconnection/rejection domain, and entitlement/grandiosity were strong predictors for the development of anger. Abandonment, entitlement, and emotional deprivation were related to behavioral manifestations of anger, and entitlement was also related to cognitive processes. Additionally, insufficient self-control was related to reactivity (Askari, 2018). In this study, entitlement and insufficient self-control were in the composite schema domain. Abandonment and emotional deprivation were in the disconnection/rejection domain.

Past research has supported the connection between EMS and aggression in specific clinical samples, it is unclear if these results extend to a general population. The inclusion of the PID-5-BF adds the ability to compare EMS to personality traits and their subsequent predictability. Confirmation of this application will allow for a much broader use of these concepts.

**Assessment of Aggression**

The prediction of aggression is a common practice in a variety of settings. In corrections, for example, the prediction of aggression is vital when considering the risk that one poses to the community prior to release, or when setting conditions of bail. In addition, the measurement of aggression occurs in treatment settings to determine the risk posed to others, and it is also commonly required by court order.

Oftentimes to predict aggression, assessments are utilized. There are multiple measures of aggression which vary widely and do not consistently adhere to any singular methodology (Yang et al., 2010). Some measures focus on the recognition of psychological disorders through diagnostic assessments (Berman et al., 1998). Others target instances of past behavior such as the Offender Group Reconviction Scale, Violence Risk Appraisal Guide, Risk Matrix 2000, and Violence Risk Scale. Some measures more heavily rely on the presence of psychopathy such as the Psychopathy Checklist – Revised (Yang et al., 2010). On many of these measures, there is no clear target from which aggression emerges, nor do they apply to a therapeutic framework.

There has been some usage of personality trait measures to predict aggression. Smith et al. (2020) studied intimate partner aggression in newlywed couples utilizing the Personality Inventory for DSM-5 – Brief Form (PID-5-BF). Their results suggested that certain elevated personality traits had an adverse effect for aggression with one’s spouse. They specifically cited that women high in the detachment domain and men high in the antagonism domain were likely to benefit from intervention. The PID-5’s domain of antagonism has been consistently conceptualized with aggression. It has had significant correlations with aggressive traits of callousness, deceitfulness, hostility, and manipulativeness (Sleep et al., 2020).

The PID-5-BF directly measures maladaptive personality traits from a diagnostic perspective of personality disorders, specifically derived from the DSM-5 Alternative Model for Personality Disorders, which emphasizes a trait-based approach. These various personality traits are grouped into relevant domains, which relate to overarching personality disorders, featuring the domains of negative affect; defined as a wide range of negative emotions, detachment; avoidance of social/emotional experiences, antagonism; behaviors that put one against other people; disinhibition; a focus on immediate gratification, and psychoticism; exhibiting odd or unusual behaviors and/or cognitions (American Psychiatric Association, 2013). The PID-5-BF is regarded as an emerging measure. Although its psychometric properties are considered strong, it is still a new measure that will benefit from further research with varying populations (Krueger et al., 2013). Due to the support for the DSM-5 Alternative Model for Personality Disorders and aggression, the PID-5 provides valuable insights of correlated aggression and traits; however, it does not directly apply to any single therapeutic treatment approach.

Schema theory asserts that maladaptive behaviors develop in response to schemas. Research has supported the role of early maladaptive schemas (EMS) with aggression in a multitude of specific acute populations such as individuals with personality disorders (Frias et al., 2017), individuals with substance use disorders (Shorey et al., 2015; Zamirinejad et al., 2017), victims of childhood sexual abuse (Estévez et al., 2016), adult male prisoners (Dunne et al., 2018b), and those who have been exposed to significant family violence (Calvete et al., 2018). The Young Schema Questionnaire (YSQ-S3; Young, 2019) is a specific tool that allows the measurement of underlying early maladaptive schemas (EMS).

Once specific EMS have been identified, appropriate interventions can be selected within the framework of schema therapy. Schema therapy is a dynamic approach with a wide array of overlap with other intervention strategies (Young et al., 2006). The use of EMS allows
more targeted and individualized treatment for those that demonstrate aggression, as they are a more specific target for intervention, due to their connection with relevant past experiences. This is apparent when EMS are compared to a diagnosis or symptoms, which are more general, especially when being applied to a treatment plan. This will improve treatment for those who have been nonresponsive to typical manualized approaches (Dunne et al., 2018a).

The comparison of EMS to personality traits from the PID-5-BF will provide additional insight into their relationship in terms of their application to aggression. EMS is a theory driven approach to potential underlying causes, whereas the PID-5-BF is diagnostic driven and needs new research to support its utility. Both concepts can be clinically useful when considering aggression in the general population.

Hypotheses

The first research hypothesis is that individuals with a higher presence of early maladaptive schemas will demonstrate higher scores on aggression. More specifically, participants with EMS of abandonment/instability, mistrust/abuse, vulnerability to harm/illness, insufficient self-control/self-discipline, subjugation, self-sacrifice, emotional inhibition, and/or punitiveness will demonstrate higher aggression scores than those with other EMS. Abandonment/instability and mistrust/abuse are found in the disconnection/rejection domain, vulnerability to harm in the impaired autonomy domain, and insufficient self-control, subjugation, self-sacrifice, emotional inhibition, and punitiveness in the composite domain. The second hypothesis is that the PID-5 domain of antagonism will demonstrate significant correlations with aggression.

Methods

Participants

Participants were recruited online from Facebook groups, Amazon Mechanical Turk, and Reddit. Focus was placed on relevant subgroups on each of these platforms, such as dissertation or research focused Facebook groups and subreddits such as r/SampleSize. Participants were recruited voluntarily through a posted description of the study. A prompt stated that they must reside in the United States of America and be able to read English proficiently. Prior to registration, potential participants were presented with an online informed consent and informed that they would have the option to enter a raffle, as a separate webpage, for a $50 Amazon gift card. In total, 203 individuals accepted informed consent and began to complete the offered surveys. Of these participants, 124 participants completed all three measures and were used in the following analyses. An a priori power analysis was conducted using G*Power version 3.1.9.7 (Faul et al., 2007) to determine the minimum sample size required for the two hypotheses. Results indicated the required sample size to achieve 80% power for detecting a medium effect, at a significance criterion of α = .05, was N = 76 for tests of linear regression for the first hypothesis and N = 91 for the second hypothesis. Therefore, the required sample size was exceeded.

Participants were asked to provide their age, gender, level of education, their range of income, and race/ethnicity. The sample included a wide range of ages from 18 to 70 years old (M = 37.74, SD = 11.10). About half of the sample identified as women (49%), half as men (49%), and a single participant identified as nonbinary. For ethnicity, the highest percentage identified as European American (70%). The second highest identified as Asian American (12%), followed by African American (8%), Hispanic (5%), and Native American or Pacific Islander (5%).

Measures

Buss Perry Aggression Questionnaire

The Buss Perry Aggression Questionnaire is comprised of 29 self-report items that measure four factors consisting of physical aggression, verbal aggression, anger, and hostility. Answers are scored on a 5-point scale. Items relate to general attitudes, hypothetical situations, and one's past (Buss & Perry, 1992). In this study, Cronbach’s alpha for the 29 BPAQ items was α = .95, indicating excellent reliability. The BPAQ has been shown to be, at minimum, moderately associated with act-based measures of aggression (Archer & Webb, 2006). It is widely used to measure aggression in a multitude of populations, particularly in studies that also examined EMS (Askari, 2018).

Young Schema Questionnaire

The most commonly utilized version of the Young Schema Questionnaire is the short-form (YSQ-S3). The YSQ-S3 features 90 items that are scored on a 6-point scale and measure all 18 early maladaptive schemas (Young, 2019). Calvete et al. (2013) studied the YSQ-S3 and found that EMS were significantly correlated with one another. They asserted that alpha coefficients were acceptable excluding the dependence schema. Calvete et al. (2013) critiqued the accuracy of the usage of domains as methods of categorization for EMS. They found consistency only with the disconnection/rejection and impaired autonomy domains. Calvete et al. (2013) suggested that the remaining three domains could be combined into a single, more accurately correlated domain. They concluded that the YSQ-S3 has adequate...
structure, consistency, stability, and concurrent validity. The YSQ has been used in multiple studies and is widely regarded as the standard for measuring early maladaptive schemas (Calvete et al., 2013; Corral et al., 2014; Zamirinejad et al., 2018). The YSQ-S3 had good internal consistency for the present study (Cronbach’s α = .73).

To obtain a participant’s total number of schemas present, binary variables were created in accordance with the participant’s score for an individual schema. If this score passed the threshold identified by the YSQ-S3, it was scored as present (Young, 2019). Schema domain variables were also calculated. Three total domains were utilized in accordance with the findings from Calvete et al. (2013) consisting of disconnection/rejection, impaired autonomy, and a composite third domain.

**Personality Inventory for DSM-5**

The Personality Inventory for DSM-5 - Brief-Form – Adult (PID-5-BF) includes 25 self-report items scored on a scale ranging from 0 (very false or often false) to 3 (very true or often true). A total item score is obtained by the sum, which represents a total personality pathology score. This total score is intended to be used as an overall screen for personality dysfunction and has been validated as such in previous studies (Combaluzier et al., 2018; Fossati et al., 2017; Gomez et al., 2020). Respective item scores are averaged and load to one of five personality trait domains. These domains include negative affect, detachment, antagonism, disinhibition, and psychoticism (Krueger et al., 2013). In this study, the PID-5-BF had a Cronbach’s alpha value of α = .95, indicating excellent reliability.

**Procedure**

After review and approval by Rivier University’s Institutional Review Board, participants completed the demographics questionnaire and all surveys online. After completion of these measures, participants were displayed a debriefing page which explained the goal of the study in more detail. At the end of this page, participants had the option to click a link to enter a raffle for a $50 Amazon gift card.

**Results**

**Descriptive Statistics**

The sample demonstrated a definite presence of early maladaptive schemas ($M = 7.28, SD = 6.49$). The mean number of early maladaptive schemas was 7.28 out of a maximum of 16 possible. All individual schemas were represented across participants. The sample’s total DSM-5 personality inventory scores were also notable, indicating a moderate presence of personality domains ($M = 2.27, SD = 0.74$). The mean presence of personality domains was just over two out of the five possible domains, with all personality domains evenly represented. Total aggression scores were similarly in the moderate range ($M = 0.55, SD = 0.17$).

**Correlation Matrix of Means, Standard Deviations, EMS, BPAQ, and PID**

<table>
<thead>
<tr>
<th>TABLE 1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Correlation Matrix of Means, Standard Deviations, EMS, BPAQ, and PID</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>M</td>
</tr>
<tr>
<td>M</td>
</tr>
<tr>
<td>SD</td>
</tr>
<tr>
<td><strong>Note:</strong> EMS = early maladaptive schema; BPAQ = Buss Perry Aggression Questionnaire; PID = Personality Inventory for the DSM-5. * Indicates significance at the .01 level.</td>
</tr>
</tbody>
</table>
Correlations
The following analyses are best understood within the full set of bivariate correlations among all the measures, available in Table 1. The data featured linearity of EMS, personality traits, and aggression scores. The sample was normally distributed, which confirmed that the assumptions of the following statistical tests were met. Pearson’s correlation coefficients were conducted with all three measures, using the domains and total scores (see Table 2). Supporting the first hypothesis, total EMS present had a strong correlation with total aggression, \( r(142) = .66, p < .001 \) according to defined ranges identified by Cohen (1988). Similarly, the schema domains of disconnection/rejection, impaired autonomy, and the composite domain all had strong correlations with total aggression.

The second correlational analysis compared total EMS with PID-5-BF domains (see Table 3). The total of EMS present had strong correlations with all five PID-5-BF domains. The PID-5-BF domains had stronger correlations with BPAQ total aggression scores than with schema domains. Supporting the second hypothesis, total PID-5-BF scores had higher correlations with BPAQ total aggression, \( r(126) = .88, p < .01 \), and was also stronger than total EMS present, \( r(135) = .66, p < .01 \).

Tests of Regression
To control for the possibility of Type I error, the Holm-Bonferroni method (Holm 1979) was utilized. The first hypothesis involved total EMS scores with total aggression and aggression domains as well as with schema domains and the same aggression variables. The results found that adjusted values of significance necessitated the elimination of a significant finding for the schema domain of impaired autonomy with verbal aggression.

The total number of EMS present significantly predicted total aggression scores, \( \beta = .02, t(115) = 9.53, p < .001 \). A significant proportion of variance was explained by total EMS, \( R^2 = .44, F(1, 115) = 90.78, p < .001 \). Concerning specific domains of aggression, total EMS present was significantly predictive of anger, hostility, verbal, and physical aggression. This was in support of the research hypothesis.

Multiple linear regressions were conducted with three domains of early maladaptive schemas. EMS domains were significantly predictive of total aggression scores and explained a higher proportion of variance than total EMS present, \( R^2 = .51, F(3, 88) = 29.96, p < .001 \). The domain of impaired autonomy had moderate partial correlations, \( r(88) = .40, p < .001 \), and was a significant predictor.

EMS domains and the aggression domain of anger were significant but less predictive than total aggression scores, \( R^2 = .42, F(3, 95) = 23.08, p < .001 \). Impaired autonomy was similarly significant. EMS domains and hostility were significant and more predictive than total aggression scores and the other aggression domains, \( R^2 = .53, F(3, 97) = 35.91, p < .001 \). All three of the EMS domains were significant. EMS domains and physical

| TABLE 2 |
| Early Maladaptive Schemas Separated by Domain |
|          | MD  | SD  |
| Disconnection & Rejection Domain |
| Emotional Deprivation            | 1.55 | 1.66 |
| Abandonment                      | 1.55 | 1.69 |
| Mistrust                         | 1.39 | 1.52 |
| Isolation                        | 1.50 | 1.65 |
| Defectiveness                    | 1.28 | 1.58 |
| Impaired Autonomy/Performance Domain |
| Failure                          | 1.39 | 1.61 |
| Incompetence                     | 1.29 | 1.53 |
| Vulnerability                    | 1.25 | 1.47 |
| Enmeshment                       | 1.20 | 1.52 |
| Composite Schema Domain          |
| Subjugation                      | 1.41 | 1.43 |
| Self-Sacrifice                   | 1.66 | 1.57 |
| Emotion Inhibition               | 1.56 | 1.52 |
| Unrelenting Standards            | 1.90 | 1.60 |
| Entitlement                      | 1.46 | 1.47 |
| Insufficient Self-Control        | 1.30 | 1.44 |
| Recognition-Seeking             | 1.45 | 1.55 |
| Pessimism                        | 1.53 | 1.54 |
| Self-Punitiveness                | 1.44 | 1.51 |
| EMS Total                        | 7.28 | 6.49 |

Note. EMS = early maladaptive schema

| TABLE 3 |
| Personality Inventory for DSM-5 |
|          | MD  | SD  |
| Negative Affect                   | 2.51 | 0.78 |
| Detachment                        | 2.35 | 0.81 |
| Antagonism                        | 2.28 | 0.90 |
| Disinhibition                     | 2.35 | 0.87 |
| Psychoticism                      | 2.28 | 0.81 |
| PID_Total                         | 2.28 | 0.74 |

Note. DSM-5 = Diagnostic and Statistical Manual of Mental Disorders – 5th Edition; PID = Personality Inventory for the DSM-5.
aggression was significant and similarly as predictive as anger, $R^2 = .42, F(3, 94) = 22.98, p < .001$. Impaired autonomy was the only significant domain. EMS domains and verbal aggression was significant but less predictive, $R^2 = .29, F(3, 96) = 13.00, p < .001$, and no domain was individually significant.

The second hypothesis, tested for Type I error, involved the total PID score with total aggression and PID domains with total aggression. The adjusted values for significance mandated the elimination of both the PID domains of psychopathic and detachment with total aggression.

Linear regression was run with the total scores on the PID-5-BF and total aggression scores. Total PID-5-BF was significantly predictive of total aggression scores, $\beta = .20, t(108) = 18.82, p < .001$. It explained the highest proportion of variance of any other total score variable $R^2 = .77, F(1, 108) = 354.12, p < .001$.

Multiple linear regression was also run with the PID-5-BF domains. The domains were significant and explained a slightly higher proportion of variance than total PID-5-BF scores, $R^2 = .77, F(5, 104) = 69.17, p < .001$. Negative affect and disinhibition were significant. This was not in support of the hypothesis, in which antagonism was cited to be a significant predictor.

**Discussion**

The initial research hypothesis stated that individuals with higher presence of EMS will demonstrate higher scores on aggression. Results supported that higher scores on both total EMS and EMS domain scores were predictive of increased aggression scores on the BPAQ. This finding is consistent with Dunne et al. (2018b), who also supported that total EMS present was significantly correlated with increased aggression, although their study was focused on a more specific sample of male prisoners. Due to most studies utilizing schema domains in their analyses, three domains were tested in the present study: impaired autonomy, disconnection/rejection, and a composite domain. All three schema domains were significant with hostility. Impaired autonomy was significant with the remaining domains of anger, physical, and total aggression scores. According to the theory, “impaired autonomy includes the schemas of dependence, vulnerability to harm, enmeshment/undeveloped self, and failure to achieve” (Young et al., 2006, p. 14–15).

It is difficult to directly compare the use of schema domains in this research to previous studies for multiple reasons. Previous studies have focused on more specific populations; this sample is intended to be a measure of a general population. Most studies have also used all five schema domains, whereas only three domains were used for this sample. This is due to following the recommendations of Calvete et al. (2013), in which adequate consistency was found only for disconnection/rejection and impaired autonomy/performance.

Considering these factors, previous research asserted that the disconnection/rejection domain was the most prevalent domain with samples of victims of child sexual abuse, men seeking residential treatment, and adolescents exposed to family violence and manifestations of aggression in the form of displaced aggression, physical aggression, verbal aggression, and dating violence respectively (Calvete et al., 2018; Estévez et al., 2016; Frias et al., 2017; Shorey et al., 2015). None of these previous studies cite the prevalence of the impaired autonomy/performance domain as significant with a form of aggression. This may support that individuals in a general sample are more likely to exhibit aggression when their ability to function is threatened, whether it is a perception of survival or to succeed in life. Impaired autonomy/performance often has its origin in an enmeshed childhood, with the child's confidence being consistently decreased. It can also arise when a child is subjected to constant overprotection or is seldom reinforced for performing well as an individual (Young et al., 2006). Regarding this domain's origins, aggression is a logical outcome for many, as frustration will likely manifest as a product of a lack of confidence in one's own abilities. As this belief worsens, it can feed into the perceptions of survival and success. It is possible that the specific samples in these previous studies had more uniform presentations of EMS due to being grouped by specific situations or experiences.

**Comparisons With the PID-5-BF**

The PID-5-BF explained higher proportions of variance over total EMS and schema domains, suggesting that the PID-5-BF is more predictive. PID-5-BF domains of negative affect and disinhibition were significant. This finding was somewhat surprising, as the personality trait of antagonism includes the description "behaviors that put an individual at odds with other people" and "callous antipathy towards others." However, the remainder of the definition includes descriptions that are more consistent with narcissistic presentations such as "exaggerated sense of self-importance and a concomitant expectation of special treatment" and "encompassing both unawareness of others' needs and feelings, and a readiness to use others in the service of self-enhancement" (American Psychiatric Association, 2013, p. 770). By these descriptions, those who score highly on antagonism may be demonstrating factors that disrupt their interactions with others, but that do not commonly rise to a level that would be considered aggressive. Sleep et al. (2021)
affirmed these factors of antagonism as not inherently aggressive, including grandiosity, suspiciousness, lack of empathy, and manipulativeness. They asserted that hostility was the most significant factor related to aggression, which is only part of this framework. Similarly, when considering the externalized constructs of antagonism, many do not involve aggression, such as distrust, entitlement, arrogance, inattentiveness, and alienation (Mullins-Sweatt et al., 2022).

Antagonism is unlikely to lead to significant impulsivity or experience activating emotion, as is seen in the related domains on the BPAQ. Negative affectivity is a logical domain for a relationship to aggression as its definition includes “frequent and intense experiences of high levels of a wide range of negative emotions,” (American Psychiatric Association, 2013, p. 770). One of these cited emotions is anger, which is inherently aggressive. It also includes other emotions that could quickly lead to aggression, such as shame and anxiety. Disinhibition is another sensible relationship, due to its impulsivity, focus on immediate gratification, and its lack of focus on future consequences (American Psychiatric Association, 2013).

Comparisons With Prior Research
The sample of this study was intended to be general American adults with no particular grouping of psychiatric disorders or histories of trauma. Other studies focused on more clinically specific populations.

For example, Frias et al. (2017) conducted a study of individuals diagnosed with borderline personality disorder. Their sample featured a significant presence of various diagnoses, including eating disorders, substance use disorder, major depression, and posttraumatic stress disorder. Their presence of EMS was far greater than in this study. This is consistent with schema theory, which holds that those with more pathological presentations will have a higher prevalence of EMS (Young et al., 2006). Interestingly, this sample’s lowest EMS was the same: enmeshment/undeveloped self ($M = 1.2, SD = 1.52$). Self-sacrifice was the second most prevalent in this study’s sample ($M = 1.66, SD = 1.57$). This supports that, although actual scores are significantly different, individual EMS prevalence remains similar across different populations.

Other studies utilizing the Buss Perry Aggression Questionnaire are also important for comparison. Lin et al. (2016) used a sample which was mostly men (72%). They excluded participants with diagnoses of psychiatric disorders. They found notably lower aggression scores than this study. These values were likely lower due to the elimination of those with known psychiatric disorders, although this study did not seek to omit these individuals.

Limitations
There are several limitations to consider in relation to these findings. This study sought to utilize a representative general sample of adults in the United States. Although all ages were represented between 18 and 70, most participants were between the ages of 30–40 with a median age of 35. As such, this sample had fewer individuals on the younger and older ends of the age range.

For race/ethnicity, the sample was generally aligned with the U.S. census for individuals identifying as European American. The sample was over-representative for those identifying as Asian American at 12%, with the U.S. Census (2020) reporting this population at 5.4%. Individuals identifying as either African-American (8%) or Hispanic (5%) were underrepresented, well below the U.S. Census (2020) findings of 12.7% African-American and 17.6% Hispanic. Future research should seek to obtain samples more representative of African-American and Hispanic populations.

Of further concern, the manner by which data was collected was entirely online in a digital survey format. This method introduced the potential for a technology barrier, meaning those less familiar with online survey platforms were less likely to find this study. The use of Amazon Mechanical Turk, Reddit, and Facebook compounds these concerns, due to the requirement for potential participants to be active users of these platforms to have a significant likelihood of recruitment. The majority of participants were recruited through Amazon Mechanical Turk. Although Amazon did not explicitly provide total demographic data on this participant pool, it is likely that this population is further limited to those more familiar with online technologies, beyond what would be expected from individuals recruited through only social media. Further research should consider these implications and ideally recruit with a traditional approach as well as offering paper surveys.

Although still significant with a variety of variables, verbal aggression was often the lowest when compared to other domains of aggression. Items loading for verbal aggression, on the Buss Perry Aggression Questionnaire, have the potential to have varying interpretations between individuals. Some items could be interpreted as more of a thought process that one expresses minimally such as the item “I often find myself disagreeing with people,” (Buss & Perry, 1992). When compared to other domains, verbal aggression appears to be the most susceptible to interpretive variation between individuals.

The selection of the PID-5-BF merits some examination. The PID-5-BF is a well-supported measure of personality domains; however, it is limited due to its brevity in terms of specificity. The Personality Inventory for DSM-5 Faceted Brief Form (PID-5-FBF)
Predictive Ability of EMS for Aggression

Lavoie and Harwood

is a more specific tool which measures a wider range of personality traits. It features scores for 25 personality trait facets measured with a total of 100 items. The Personality Inventory for DSM-5 (PID-5) measures both the personality facets and personality domains, but is significantly longer with 220 items (American Psychiatric Association, 2013). The utilization of these measures would provide a more specific and complete picture of the role of personality traits in terms of their correlation to EMS and their predictive ability with aggression. These more complex measures were not selected for this study due to concerns with their length introducing an increased amount of survey fatigue. Future research could benefit from studying these relationships separately or splitting the measures into multiple sessions.

Implications for Clinical Practice
The Personality Inventory for DSM-5 – Brief Form is an ideal measure to predict aggression from underlying traits. This can be used effectively for risk assessment, where the practitioner seeks to also develop a detailed profile of an individual's personality. This is especially valuable when compared to more traditional risk assessments, in which there is only focus on past behavior and/or more specific traits.

Although personality traits had a stronger predictive ability, the use of EMS carries a significant benefit in that they directly translate to a therapeutic modality, schema therapy. Ultimately, assessing for aggression and its origins is ideally only the first task towards minimizing the individual's aggression. The identification of specific EMS allows the practitioner a targeted approach towards empirically supported interventions that will reduce aggression at its source. Practitioners can teach the effects of individual schemas, use experiential exercises, help clients to adjust their schema responses, or even use schema modes to improve the effects of the identified EMS (Young et al., 2006).

Many previous studies focused on EMS in specific clinical populations and specific maladaptive behaviors. This study confirmed that EMS are still prevalent in a general sample of adults and that they are useful for their role in the predictability of aggression. By extension, this supports that schema therapy interventions are an effective approach for aggressive individuals.

References
Lavoie and Harwood | Predictive Ability of EMS for Aggression

https://doi.org/10.1037/a0024089


Author Note

Derek A. Lavoie @ https://orcid.org/0000-0002-2549-8533
Elizabeth A. Harwood @ https://orcid.org/0000-0001-8711-3988

This study was completed as a doctoral dissertation and was presented virtually at the 2021 New England Psychological Association Annual Meeting. We have no known conflicts of interest to disclose. Special thanks to Carol A. Langelier and Kate R. Murphy for their support and insights to this research.

Correspondence concerning this article should be addressed to Derek A. Lavoie, Rivier University, 420 South Main St, Nashua, NH 03060. dlavoie@rivier.edu, 603-440-5755

COPYRIGHT 2022 BY PSI CHI, THE INTERNATIONAL HONOR SOCIETY IN PSYCHOLOGY (VOL. 27, NO. 4/ISSN 2325-7342)
Turn Your Compassion Into a Career in Mental Health

Our online degree programs feature relationship-focused curriculum, innovative faculty, and lead to licensure.

alliant.edu/psychology
Call for Doctoral-Level Reviewers

*Psi Chi Journal* is seeking reviewers interested in providing constructive feedback on our authors’ original empirical research. To increase the journal’s scope and relevance, diverse people of varied racial, ethnic, gender identity, sexual orientation, religious, and social class backgrounds are welcomed. If you have a doctoral degree in psychology or a related field, please join us.

**TO SUPPORT OUR EXCEPTIONAL REVIEWERS**
- Our online portal allows you to submit specific subject areas that you feel comfortable with reviewing.
- At any time, you can set unavailability dates so that we will know when not to contact you.
- With each request to review, you will receive basic instructions and a template review form, which you can use in order to save you time during your review.

**TO BECOME A REVIEWER, EMAIL**

steve.rouse@psichi.org

---

An Eye on Graduate School

Guidance Through a Successful Application

Psi Chi’s digital anthology brings together our very best advice about applying to graduate school—advice accumulated from 25+ experts in over 20+ years of *Eye on Psi Chi* magazine issues.

In nineteen chapters, this eBook will help you navigate the seven primary steps that are vital to your acceptance at the graduate program of your choice.

- Preparing for and Selecting a Graduate Program
- The Application Process
- Preparing for the GRE
- Soliciting Letters of Recommendation
- Writing the Personal Statement
- Interviewing
- Choosing a Program and Succeeding in Graduate School

**Download Today at Store.PsiChi.org**
Log in as a Psi Chi Member to receive a discount!

Kindle Edition Available Now!
Publish Your Research in *Psi Chi Journal*

Undergraduate, graduate, and faculty submissions are welcome year round. Only one author (either first author or coauthor) is required to be a Psi Chi member. All submissions are free. Reasons to submit include

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

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

---

**Become a Journal Reviewer**

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

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

---

**Resources for Student Research**

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

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

---

**Learn About Psi Chi**

Psi Chi is the International Honor Society in Psychology. Membership is primarily open to undergraduates, graduate students, transfer students, full-time and part-time faculty members, and alumni.

See membership benefits and a link to apply at [www.psichi.org/page/member_benefits](http://www.psichi.org/page/member_benefits)