Published Biannually by
The Department of Counseling and Clinical Psychology
Teachers College, Columbia University

2022 – Volume 19
ISSN 1088-4661

Graduate Student Journal of Psychology

The Graduate Student Journal of Psychology is a publication of the Department of Counseling and Clinical Psychology, Teachers College, Columbia University.

AIMS AND SCOPE. Founded in 1998, the Graduate Student Journal of Psychology (GSJP) is a peer-reviewed journal consisting of empirical and theoretical articles drawing from all areas of academic psychology. The GSJP is published biannually and includes original research articles, case studies, theoretical articles, and brief reports in the areas of adult and child psychopathology, the psychotherapeutic process, community mental health, multiculturalism, assessment, child development, ethics, professional development, and other topics with a relationship to research or practice in psychology.

MANUSCRIPTS. Manuscript submission for Volume 20 will be handled on a rolling basis. Submit manuscripts according to the submission guidelines printed in this issue. Please email us your manuscript along with a cover letter to gsjp@tc.columbia.edu. For more information on submission process visit [www.tc.edu/publications/gsjp.](http://www.tc.edu/publications/gsjp)

Graduate Student Journal of Psychology

2022 – Volume 19

FORWARD

ARTICLES

TEACHERS COLLEGE
COLUMBIA UNIVERSITY

Department of Counseling and Clinical Psychology

LETTER FROM THE EDITORS:

After months of hard work and re-adjustment to the full throes of academic life, we are proud to present Volume 19 of the Graduate Student Journal of Psychology. The past few years have held a range of challenges for students and professionals in the field of psychology. Much of our work has been curtailed by the impacts of the COVID-19 pandemic. Additionally, many researchers shifted their work to examine the effects of the pandemic on mental health and well-being. These topics often dominated the research in order to find answers during a time of great global need.

As the field and the rest of the world attempt to navigate a return to our new normal, so too is the Graduate Student Journal. This volume, Volume 19, is composed of unique research studies examining a broad array of psychology-related topics. We are immensely proud to publish a journal that offers graduate students the opportunity to explore their desired research interests. This is especially true as such opportunities to demonstrate proficiency in research in a professional public forum have been limited, an issue largely compounded by the pandemic. The journal acts as a meaningful bridge between graduate students and the world of professional publication by allowing students the opportunity to share unique and innovative research.

This volume includes a variety of methodological and analytical approaches such as the a priori procedure, systematic reviews, and moderation analysis. This diversity is not only mirrored in the methodological approaches, but in the ethos of the journal itself. The journal's mission to prioritize diversity and inclusion is reflected in the content of the selected research articles. Examples include explorations of cultural identity and socioeconomic status.

This volume is a testament to the hard work of our featured authors, peer-reviewers, and editorial board. We would especially like to thank our faculty sponsor, Dr. Randall Richardson-Vejlgaard. Dr. Richardson's guidance and support has been fundamental throughout this process. Most notably, with the transition to a new Editorial Board, we want to extend our deepest gratitude to the former GSJP editors, Curreen Luongo (Editor-in-Chief, 2022), Juno Pinder (Submissions Editor, 2022), Claudia Rodriguez (Communications Editor, 2022), and Mitsu Puri (Production Editor, 2022). There are many moving parts to this process and we could not have produced this volume without their tireless efforts in ensuring a smooth transition. We are extremely grateful for everyone's dedication and contribution. We invite our readers to share their thoughts and comments with us at gsjp@tc.columbia.edu. You can also connect with us through social media (Instagram, Twitter, Linkedin) @gsjp_journal.

The Editors,

Daniella Ekstein, Cassidy Iervasi, Catherine Shorb, Ellen Somers, Evelyn Tsai, Hannah Manley, and
Lori Kim

Graduate Student Journal of Psychology
2022, Vol. 19

Merging Multiple and/or Divergent Datasets using SPSS: A Method Review and Tutorial

Lillian M. Audette, Katherine A. Johnson, Marie S. Hammond, Jenna S. Lehmann, and Michael Oyeteju,
Department of Psychological Sciences & Counseling, Tennessee State University

For many reasons, researchers place their data in multiple electronic datasets and later find that they wish to combine those datasets for a particular analysis. This article reports on an analysis of the extant literature on combining multiple and/or divergent datasets and provides both a tutorial and related syntax that combines non-matching datasets (i.e., from different sources) in such a way that all data is contained in the final combined dataset with identification of data source. This syntax represents an improvement over the existing SPSS (v23-v27) combining dataset routines in terms of 1) expanding the types of datasets that can be combined, 2) completeness of the resulting dataset, and 3) facilitating management of multiple and/or divergent data sets. An additional benefit of using this process is the incorporation of a method to test the accuracy of the merged data and thus verify the data quality. The SPSS syntax along with instructions and examples are reviewed in the article.

Keywords: SPSS, syntax, datasets, merging, longitudinal data

Many researchers face problems merging multiple datasets with divergent or mismatched cases and/or variables to form a more complete dataset. This is particularly true when conducting research with humans in which data is obtained from multiple sources, such as self-report questionnaires, institutional databases, and/or collateral contacts. Challenges that occur in utilizing data from divergent sources include different structures, different formats, or that data is incomplete when compared to the original dataset.

The present article focuses on merging multiple datasets either generated longitudinally or containing different variables. Examples of the datasets to which we refer include pre-/post-testing for intervention research, longitudinal research utilizing two or more waves of data, or instances in which data from multiple sources related to the same case. These datasets are likely to be mismatched in that they may not contain the same variables (in the instance of adding cases) or may not include the same cases (in the instance of adding variables). Other relevant datasets might be considered "complex" in that the datasets contain different variables and have differences in cases. For example, it may be that graduate students working as a part of a larger research team are tasked with managing the data and will need to understand the process and steps to efficiently combine these datasets.

It should be noted that in order to align with current terminology, rather than describing these datasets as "complex," which has a specific meaning within the social sciences (Guha et al., 2009), we will use the term "divergent" to refer to datasets that do not match variable for variable, time frame by time frame, and/or case for case. Further, the term "merge" will be used in this article rather than "combining" in that combining implies a common data structure that facilitates bringing the datasets together, while the term "merging" typically relies on a "key" or "identifier" variable that provides the common link between the datasets. This is distinct from data fusion, which occurs when multiple data sources are integrated without full preservation of all data (Haghighat et al., 2016). The goal of the process to be described is that the datasets are merged in such a way that all data is retained from its original dataset in the final product.

When it comes to merging two or more datasets that contain mismatching and/or missing cases, several complications could occur – resulting in compromised analyses. For instance, mismatched data may indicate that data were added to the dataset incorrectly. Additionally, missing data can result in decreased analysis power or, depending on the analysis required, may prevent an analysis from being completed entirely. Accurate dataset merging is vital to prevent potential errors from being introduced, prevent data loss, and allow researchers to have more reliable and valid datasets with which to conduct analyses. Thus, the accurate merging of discrete datasets allows researchers to conduct analyses on a more accurate dataset, therefore producing results that have greater reliability, are more generalizable, and are more replicable. For the purposes of this article, the focus will stay on mismatched or divergent dataset merging using SPSS software.

Literature Review

To assess the need for an informative guide regarding how to merge mismatched datasets, the authors chose two separate approaches to conducting the lit erature searches. First, a literature search using the research databases EBSCOHost (which is an umbrella dataset encompassing approximately 70 individual publisher databases), Web of Science, Sage Premier, Science Direct, and Google Scholar was conducted. The Boolean search phrase utilized for this search was as follows: SPSS AND dataset AND (merg* OR combin* OR concatenat* OR mung* OR wrangl*). For databases that do not use the Boolean system exactly (e.g., Google Scholar), these search terms were added manually to the advanced search settings of each database. The terms mung and wrangle were incorporated once we discovered that these terms are used outside of psychology to describe the data-cleaning process, which includes dataset merging (Braun et al., 2018; Endel & Piringer, 2015; Rattenbury et al., 2017). The above search phrase and its individual search terms were used to identify articles containing these terms in their abstracts. The resulting list from EBSCOHost provided 27 peer-reviewed articles, while the Web of Science search produced 24. None of these results acknowledged the problem of merging mismatched or divergent datasets. A second literature search was conducted using Google and Google Scholar. Since Google Scholar uses a limited form of Boolean, the search phrase was changed to the following: SPSS AND dataset AND (merge OR merged OR merging OR combine OR combination OR combining OR combined OR mung OR munging OR munged OR concatenate OR concatenated OR concatenating OR wrangle OR wrangled OR wrangling). Other combinations of these search terms were also used (e.g., data wrangling, data munging, merging datasets). The results of this search found examples of both peer-reviewed articles and books that address dataset merging, but neither focused on mismatched dataset merging nor used SPSS for this purpose.

Previous research literature on merging datasets for analyses of psychological data primarily used the Statistical Analysis System (SAS) software (Foley, 1998; Foley, 2005) for this process, rather than SPSS. Parenthetically, research literature reviewed the steps to obtain a merged dataset with the specific variables and cases of interest. Again, however, no mention was made of merging mismatched or divergent datasets. Scholarly works and textbooks were reviewed to identify potentially helpful instruction in this area. Both works that did focus on the use of SPSS and works that did not specify which program they recommended using covered only basic dataset merging practices. These sources either did not provide information on the difficulties likely to be encountered in merging mismatched or divergent datasets (Stehlik-Barry & Babinex, 2017) or they only emphasized the importance of considering these challenges without providing guidance or details on its technicalities (Braun et al., 2018; Endel & Pringer, 2015; Rattenbury et al., 2017). Guides on merging datasets for non-psychological research purposes have also not focused on SPSS, but rather on software programs such as R and Python (Ojeda et al., 2014).

There does, however, appear to be plenty of gray literature (articles not formally published by commercial academic publishers) surrounding this topic in SPSS (Haddaway et al., 2015, p. 1). Examples of this gray literature include 1) websites that host questions, discussions, and video tutorials related to the common practices for merging datasets (Coyer, 2013; Truong, 2016), 2) academic websites with instructions (Coleman, n.d.; Glynn, 2002), 3) basic commands given by International Business Machine (IBM; IBM, n.d.); and 4) other SPSS tutorial websites (Spss-tutorials.com, n.d.). However, most of the gray literature provides limited basic commands such as MERGE DATASETS or COMPARE DATASETS, which do not adequately address specific divergent or mismatched dataset merging issues encountered by researchers.

Dataset Merging Methods Requirements

To address the above-mentioned issues, the extant literature previously identified was reviewed to surface existing guidelines and/or requirements. This type of information was found primarily in the gray literature (e.g., Coleman, n.d.; Coyer, 2013; DeCator, 2015; IBM, n.d.; Truong, n.d.). Six requirements of effective dataset merging methods to ensure the accuracy of data were identified from this literature. The first three requirements are basic expectations of any dataset merging method (Coleman, n.d.; Coyer, 2013). The next three requirements are more complex and are specific to merging datasets that contain overlapping variables as well as overlapping cases, and whose cases sometimes contain mismatched data (DeCator, 2015; IBM, n.d.; Truong, n.d.). These six requirements have implications for the structure and variables contained in the final created database, as will be discussed below.

Requirement 1: Include All Cases
The merged database (finished product) should contain all the cases from Dataset 1 and all the cases from Dataset 2 (and any additional datasets). The case data from each respective dataset should be faithfully replicated in the merged database.

Requirement 2: Include All Variables
The merged database should hold all the variables present in all datasets, as well as all variables present in only one of the datasets. The width, number of decimal places, labels, value labels, missing specification, column width, alignment, measure specification, and role of each variable should be faithfully replicated in the merged database.

Requirement 3: Variable Settings Fidelity
The merging method integrates variables without changing the variable regardless of the type (numeric, string) and regardless of measure (scale, nominal, ordinal).

In addition to the three basic requirements listed above (Coleman, n.d.; Coyer, 2013), the merged database should provide researchers with three additional pieces of information. The following requirements are more complex, intending to assist researchers when the datasets to be merged are suspected of containing unique, overlapping, or mismatched cases, all of which should be included in the final database.

Requirement 4: Indication of Unique Cases by Dataset
A merged database should provide information as to when a case is unique to Dataset 1 or Dataset 2 (rather than being present in both original datasets), and which dataset it originates from. For example, a merged dataset should tell us if a case with ID 13 exists uniquely in Dataset 1 or Dataset 2.

Requirement 5: Indication of Overlapping Cases
A merged database should indicate when a case in Dataset 1 is also a case in Dataset 2. For example, a merged database should tell us if a case with ID 16 originated in Dataset 1 and whether it also exists in Dataset 2.

Requirement 6: Indication of Mismatched Data by Case
A merged database should tell us when a case in Dataset 1 and a case in Dataset 2 have a matching ID but contain mismatching data while retaining both instances of data in the merged database. Detecting mismatching data is a crucial requirement when merging datasets whose data, cases, or variables do not perfectly match using SPSS. For example, a merged database should tell us if a case with ID 16 in Dataset 1 has "Brief Cognitive Behavioral Therapy" entered for the "Treatment" variable, but a case with ID 16 in Dataset 2 has "Brief Object Relations Therapy" entered for the "Treatment" variable. In addition, both instances of data should be preserved within the merged database. For example, we should see both forms of cases with ID 16 in the merged database, one with "Brief Cognitive Behavioral Therapy" and one with "Brief Object Relations Therapy" entered for "Treatment." Thus, an indicator of match/mismatch and the retention of both instances of the data should appear in the merged database.

To clarify - several of the six requirements (discussed above) should be represented by one or more variables within the merged dataset. New variables within the dataset should be generated to represent at least one, if not more, of the six requirements. Each of the six requirements should be encoded into a variable within the dataset. The final merged dataset should include, in a specific variable, the dataset name from which each case was taken. This must be a unique variable, separate from the other variables. For example, for one of the authors' research projects, data was gathered from three different institutions. In this instance, in merging the data a new variable was created which identified the dataset/institution of origin. This variable, included in the merged dataset, provides information about the origin of the data and thus met this requirement. Thus, the dataset of origin variable should not rely on non-encoded methods (such as case origin based upon the cases' ordering - i.e., SPSS row number) within the dataset. If case origin is denoted by case order, and the final merged dataset is ever re-ordered, the information regarding case origin could be lost. To summarize, each of the six dataset merging requirements must be fulfilled by being encoded as separate variables within the dataset or in such a way that the information cannot be easily lost.

Conducting the Dataset Merge

Currently, SPSS does not include a built-in function that would perform a dataset merge that ensures all six requirements (above) are met. For example, the ADD FILES command does not identify cases that are unique to one of a researcher's original datasets, nor is it able to identify duplicate cases that have mismatched data, thus, not meeting requirements 4 and 6. The MERGE DATASETS command does not meet requirements 4, 5, or 6. In order to meet requirement 1 using MERGE DATASETS, additional syntax is needed. Similarly, the MATCH FILES command does not meet requirements 1, 4, 5, or 6. Finally, the COMPARE DATASETS command does not meet requirements 1, 2, 3, or 4. In determining which syntax best meets all six requirements, a literature review was conducted which yielded five (5) known possible methods for combining datasets in SPSS. A total of 12 experiments were conducted using the five methods, with each method first being tested in its simplest form and then tested again in more complex forms. A summary of the experiments conducted, and their outcomes can be found in Table 1.

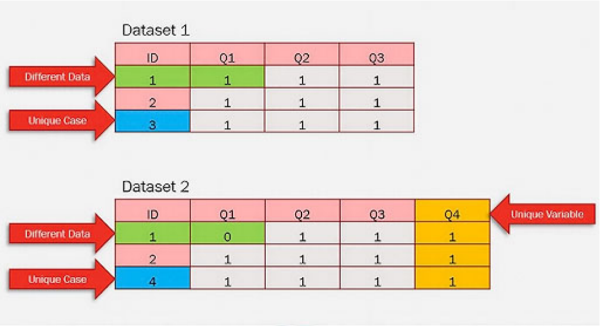
The final product (based on the outcome of experiment 12) was a set of SPSS syntax to successfully overcome the challenges associated with merging mismatched datasets using the least number of steps. It should be noted that the following steps were developed using SPSS v23 (IBM Corp., 2015) and tested using SPSS v25 and v27 (IBM Corp., 2017, 2020). It should be noted that conducting the merge using this syntax works across all versions including SPSS v28 (IBM Corp., 2021); however, changes made to v28 in the point-and-click options may create a mismatch between the instructions provided herein.

SPSS Syntax

The following instructions on how to use the syntax are applicable when researchers have two datasets that they wish to merge. It can also be used repeatedly if more than two datasets must be merged. Please refer to Appendix A for the syntax related to each step. It is also important to note that the figures provided throughout this section are simplified graphics of the analysis for demonstration only and were not conducted using the supplemental files. The supplemental files are provided for the purpose of practicing the procedure with a more realistic dataset.

Original Datasets

For variables shared by the two original datasets, the variable information should match perfectly (type, width, label, values, etc.). The following steps will erroneously identify matching variables as separate variables when they do not have identical characteristics (type, width, etc.). Before beginning the next steps, identify which dataset will be your "Dataset 1" and open both original datasets. See Figure 1 for an example of two original datasets to be merged. It is also necessary to clean the data from each of the datasets to be merged to the best of one's ability before beginning the dataset merging process.



Dataset 1

ID	Q1	Q2	Q3
1	1	1	1
2	1	1	1
3	1	1	1

Dataset 2

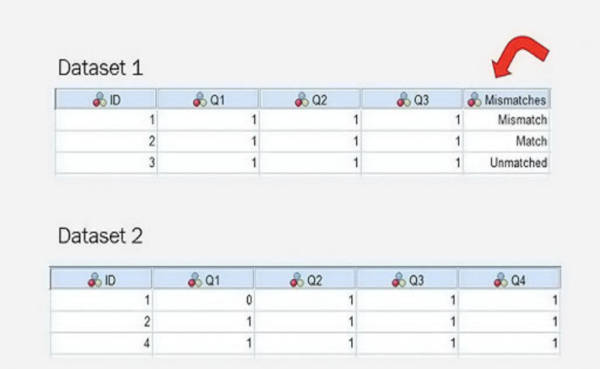
ID	Q1	Q2	Q3	Q4
1	0	1	1	1
2	1	1	1	1
4	1	1	1	1

Figure 1: Simple Datasets Pre-Combination. This is a simple example of two separate, mismatched datasets.

Syntax

Step 1: Comparing Cases Between Datasets

The purpose of Step 1 is to create a new variable that encodes the number of unique cases present in Dataset 1, the number of cases present in both original datasets that are perfectly matched, and the number of cases present in both original datasets that contain mismatching data. The output will identify the location of the mismatched data in the dataset. Please note that the datasets will remain separate. Step 1 uses point-and-click to run the COMPARE DATASETS command (which can be found under the DATA menu). Before running the command, make the following point-and-click changes. In the "Compare" tab, put the ID variable in Case IDs, and put all "Matched Fields" into "Fields to Compare." In the "Attributes" tab, select "Do not compare the data dictionaries." In the "Output" tab select "flag mismatches in a new field" and name the new variable what you wish (we used "Mismatches"). Also, in the "Output" tab, unselect "Limit the case-by-case table." Click "paste" and run the resulting syntax on Dataset 1. See Figures 2A and 2B for example Datasets 1 and 2 after Step 1 as well as an example output after Step 1.



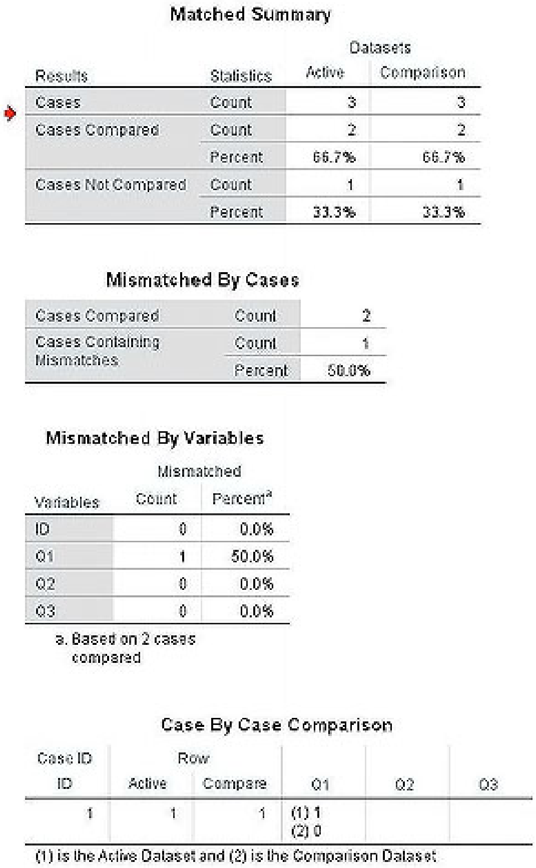
Dataset 1

ID	Q1	Q2	Q3	Mismatches
1	1	1	1	Mismatch
2	1	1	1	Match
3	1	1	1	Unmatched

Dataset 2

ID	Q1	Q2	Q3	Q4
1	0	1	1	1
2	1	1	1	1
4	1	1	1	1

Figure 2A: Simple Mismatched Datasets in SPSS after Step 1. This is an example of what the new variable created via Step 1 would look like once added.



Matched Summary

Results	Statistics	Datasets	
		Active	Comparison
Cases	Count	3	3
Cases Compared	Count	2	2
	Percent	66.7%	66.7%
Cases Not Compared	Count	1	1
	Percent	33.3%	33.3%

Mismatched By Cases

Cases Compared	Count	2
Cases Containing Mismatches	Count	1
	Percent	50.0%

Mismatched By Variables

Variables	Mismatched	
	Count	Percent[a]
ID	0	0.0%
Q1	1	50.0%
Q2	0	0.0%
Q3	0	0.0%

a. Based on 2 cases compared

Case By Case Comparison

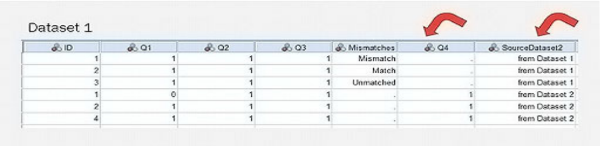
Case ID	Row				
ID	Active	Compare	Q1	Q2	Q3
1	1	1	(1) 1 (2) 0		

(1) is the Active Dataset and (2) is the Comparison Dataset

Figure 2B: Example Step 1 Output. This is an example of the output that would be generated after completing Step 1 which provides a summary of case comparisons between the two datasets.

Step 2: Merging Variables and Cases into One Dataset and Identifying Cases by Dataset

Step 2 merges all variables and cases into Dataset 1. From Dataset 1, select MERGE FILES then ADD CASES command. This is also found in the DATA menu. Before running the command, make the following point-and-click changes. Put all "Unpaired Variables" in "Variables in New Active Dataset." As needed, manually pair variables from the two original datasets. Select "Indicate case source as variable" and name the new variable that you wish to use (we used "SourceDataset2"). Click "paste" and run the resulting syntax on Dataset 1. This step will add all variables and all cases to Dataset 1. It will also identify which cases come from Dataset 1 or Dataset 2 in the newly created variable. See Figure 3 for the example Dataset 1 after Step 2. After completing this step, it may be helpful to label the values of this new variable, the syntax for which can be found in Appendix A. A value of 1 indicates that a case originates from Dataset2 while a value of 0 indicates that a case originates from Dataset 1.



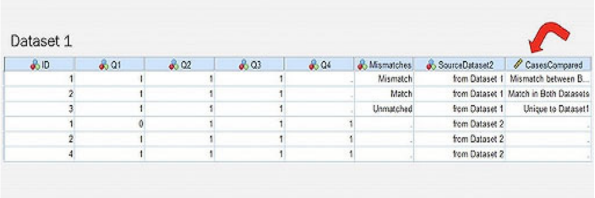
Dataset 1

ID	Q1	Q2	Q3	Mismatches	Q4	SourceDataset2
1	1	1	1	Mismatch	.	from Dataset 1
2	1	1	1	Match	.	from Dataset 1
3	1	1	1	Unmatched	.	from Dataset 1
1	0	1	1	.	1	from Dataset 2
2	1	1	1	.	1	from Dataset 2
4	1	1	1	.	1	from Dataset 2

Figure 3: SPSS Data View After Step 2. This is an example of the variables added after step 2 which includes all variables existing in Dataset 2 that do not exist in Dataset 1 and a new variable that identifies the source of each case.

Step 3: Identifying Mismatches, Matches, and Unique Cases Within Dataset 1

Step 3 duplicates the variable created in Step 1, which identifies mismatching, matching, and unique cases present in Dataset 1. The duplicated variable is the one that will be manipulated in Steps 4 and 5. Step 3 can be done with the syntax (provided in Appendix A) using your preferred variable names inserted (we used "CasesCompared"). The resulting duplicate variable only encodes information about matches, mismatches, and unique Dataset 1 cases for those cases from Dataset 1. The duplicated variable contains no information for cases added from Dataset 2, a problem that will be addressed in Step 4 and Step 5. See Figure 4 for the example Dataset 1 after Step 3. Once again, we suggest labeling the values of this new variable. A value of 1 in dicates a mismatch between both datasets, 0 represents a match in both datasets, and -1 indicates that a case is unique to Dataset1. The syntax for creating these labels can be found in Appendix A.



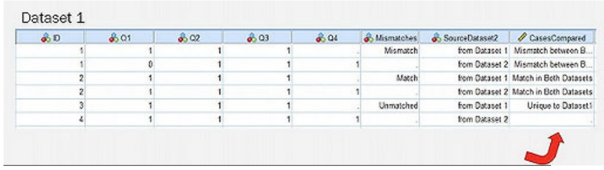
Dataset 1

ID	Q1	Q2	Q3	Q4	Mismatches	SourceDataset2	CasesCompared
1	1	1	1	.	Mismatch	from Dataset 1	Mismatch between B...
2	1	1	1	.	Match	from Dataset 1	Match in Both Datasets
3	1	1	1	.	Unmatched	from Dataset 1	Unique to Dataset1
1	0	1	1	1	.	from Dataset 2	.
2	1	1	1	1	.	from Dataset 2	.
4	1	1	1	1	.	from Dataset 2	.

Figure 4: SPSS Data View After the Creation of a Variable that Identifies Mismatches, Matches, and Unique Cases Regarding Cases Originating From Dataset 1 (Step 3).

Step 4: Labeling Matches and Mismatches from Dataset 2

Step 4 uses the syntax we developed (utilizing the LAG function), and the variables created in Steps 2 and 3. Utilizing the LAG function and the “SourceDataset2” variable, Step 4 encodes for all cases within the newly created “CasesCompared” variable (from Step 3). Step 4 encodes information as to whether the case is a match between datasets, a mismatch between datasets, or unique to Dataset 1. This step is important because without it only cases from Dataset 2 would have information encoded in the “CasesCompared” variable. Again, the syntax is found in Appendix A. When running the syntax, be careful to change the variable names to your chosen variable names, and to run the sorting syntax first (we included examples to facilitate understanding). Your ID variable must be in numerical descending order (e.g. 1, 1, 2, 3, 4, 4), with cases from Dataset 1 listed or appearing before cases from Dataset 2. This syntax solves the problem from Step 3, however, the variable “CasesCompared” still does not encode when a case is unique to Dataset 2. See Figure 5 for the example Dataset 1 after Step 4.



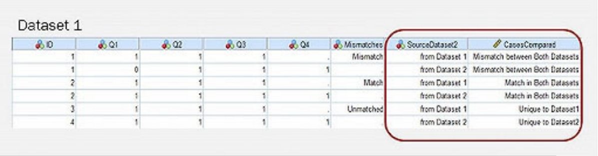
Dataset 1

ID	Q1	Q2	Q3	Q4	Mismatches	SourceDataset2	CasesCompared
1	1	1	1	.	Mismatch	from Dataset 1	Mismatch between B...
1	0	1	1	1	.	from Dataset 2	Mismatch between B...
2	1	1	1	.	Match	from Dataset 1	Match in Both Datasets
2	1	1	1	1	.	from Dataset 2	Match in Both Datasets
3	1	1	1	.	Unmatched	from Dataset 1	Unique to Dataset1
4	1	1	1	1	.	from Dataset 2	.

Figure 5: SPSS Data View After Step 4 which Identifies Mismatches and Matches in Cases Originating from Dataset 2.

Step 5: Identifying Cases Unique to Dataset 2

Step 5 encodes in the “CasesCompared” variable when cases are unique to Dataset 2. The syntax can be found in Appendix A. Since the only cases with no data in “CasesCompared” are those which are unique to Dataset 2, this syntax identifies empty data in “CasesCompared” and encodes them as unique to Dataset 2. The syntax also includes a method of labeling this new value. See Figure 6 for the example Dataset 1 after Step 5.



Dataset 1

ID	Q1	Q2	Q3	Q4	Mismatches	SourceDataset2	CasesCompared
1	1	1	1	.	Mismatch	from Dataset 1	Mismatch between Both Datasets
1	0	1	1	1	.	from Dataset 2	Mismatch between Both Datasets
2	1	1	1	.	Match	from Dataset 1	Match in Both Datasets
2	1	1	1	1	.	from Dataset 2	Match in Both Datasets
3	1	1	1	.	Unmatched	from Dataset 1	Unique to Dataset1
4	1	1	1	1	.	from Dataset 2	Unique to Dataset2

Figure 6: SPSS Data View After Step 5 which Identifies Cases Unique to Dataset 2.

Discussion

There is significant literature that provides tutorials and instructions on the basic merging of files with identical variables and/or cases. Significant literature discussing these basic merging techniques can be found for programs such as R, SPSS, and STATA. What has not been discussed in the literature but may be of use to students and researchers using small- to medium-sized datasets, is a procedure that reliably merges datasets with missing or mismatched cases and/or variables within SPSS. While numerous software programs are used to conduct research, one of the more frequently used software programs to teach statistical analysis, particularly within the social sciences, is SPSS. Oftentimes students are introduced to data analysis using the drop-down menus within SPSS rather than syntax. This article has provided information on both the use of drop-down menus and syntax to conduct data merging. Depending on the individual’s comfort level and the version of SPSS being used, either of these procedures may be more effective. In the case of an individual using version 27 or earlier, the drop-down menus provide point-and-click ease of conducting the analysis. While for those with greater comfort and/or facility using syntax, merging using the syntax is likely to be more comfortable regardless of the version used.

There are several advantages to using this procedure to merge divergent datasets. The first advantage is that it allows the researcher to store all data within a single dataset. This facilitates data analysis for most procedures. The second advantage is that by having all data in one dataset, analyses accounting for missing data are more easily conducted from within one dataset rather than across multiple datasets. The third advantage is that this procedure allows one to conduct analyses that can only be performed within a single dataset. The fourth advantage is that embedded within the syntax is a data quality check which increases the likelihood that the merged data is accurate and represents the population of interest.

An additional consideration is related to the speed or efficiency of the procedure. It should be noted that the speed, efficiency, and amount of storage required to contain the dataset will be affected by the sizes of the datasets involved. Large datasets (such as those found in the National Center for Education Statistics, etc.), will require greater computing power and storage capacity than smaller datasets.

Conclusion

Merging datasets accurately is vital to ensuring that no data is altered or lost and that researchers can easily understand the sources of all their data. Presently, there is no consensus on best practices for merging SPSS datasets with overlapping and potentially mismatched cases. The present article introduces one method to handle such a situation. The recommended method results in a merged dataset that includes all cases and variables, in their original form, stemming from two or more original datasets. It creates an output file that identifies each instance of case mismatch and the location of each mismatch within the dataset. It also results in a single variable that encoded whether a case was 1) a perfect match between datasets 2) a mismatch between datasets with non-matching data 3) unique to one dataset and which original dataset contains that unique case. Alternative strategies are recommended when the datasets to be merged contain overlapping cases but no or few shared variables.

References

Braun, M. T., Kuljanin, G., & DeShon, R. P. (2018). Special considerations for the acquisition and wrangling of big data. *Organizational Research Methods, 21(3)*, 633-659. https://doi.org/10.1177/1094428117690235

Coleman, M. (n.d.). *Merging datasets in SPSS*. Retrieved December 12, 2017, from http://www.d.umn.edu/~mcoleman/tutorials/spss/merge.html

Coyer, L. (2013, April 20). *How can you merge two files in SPSS when the cases are not perfectly the same?* [Online forum post]. ResearchGate. Retrieved December 12, 2017, from https://www.researchgate.net/post/How_can_you_merge_two_files_in_SPSS_when_the_cases_are_not_perfectly_the_same.

DeCator, D. D. (2015, February 8). *SPSS Syntax Part 4: Double Entry Comparison*. Retrieved October 12, 2017, from https://www.ddecator.com/blog/2015/2/spss-syntax-part-4-double-entry-comparison

Endel, F., & Piringer, H. (2015). Data wrangling: Making data useful again. *IFAC-PapersOnLine, 48(1),* 111-112. https://doi.org/10.1016/j.ifacol.2015.05.197

Foley, M. J. (1998). Match-merging: 20 some traps and how to avoid them. In the P*roceedings of the 23rd Annual SAS Users Group International Conference* (pp. 277-286). Retrieved from https://support.sas.com/resources/papers/proceedings/proceedings/sugi23/Advtutor/P47.pdf

Foley, M. J. (2005). Merging vs. joining: Comparing the DATA step with SQL. In the *Proceedings of the 30th Annual SAS Users Group International Conference* (pp. 184–200). Retrieved from http://www.scsug.org/SCSUGProceedings/2005/Foley_Merging vs Joining - 184.pdf

Glynn, P. (2002). SPSS for Social Science Research - Using Syntax. Retrieved December 12, 2017 from http://staff.washington.edu/glynn/spssclas/

Guha, S., Kidwell, P., Hafen, R.P., & Cleveland W.S. (2009). Visualization databases for the analysis of large complex datasets. Proceedings of the 12th International Conference on Artificial Intelligence and Statistics (AISTATS), USA, 193-200. Retrieved October 20, 2022 from https://proceedings.mlr.press/v5/guillory09a.html.

Haddaway, N. R., Collins, A. M., Coughlin, D., & Kirk, S. (2015). The role of Google Scholar in evidence reviews and its applicability to grey literature searching. *PLOS ONE, 10(9)*. https://doi.org/10.1371/journal.pone.0138237

Haghighat, M., Abdel-Mottaleb, M., & Alhalabi, W. (2016). Discriminant correlation analysis: Real-time feature level fusion for multimodal biometric recognition. *IEEE Transactions on*

Information Forensics and Security, 11(9), 1984–1996. https://doi.org/10.1109/TIFS.2016.2569061

IBM Corp. (2015). IBM SPSS Statistics for Windows (Version 23.0) [Computer software]. IBM Corp.

IBM Corp. (2017). IBM SPSS Statistics for Windows (Version 25.0) [Computer software]. IBM Corp.

IBM Corp. (2020). IBM SPSS Statistics for Windows (Version 27.0) [Computer software]. IBM Corp.

IBM Corp. (2021). IBM SPSS Statistics for Windows (Version 28.0) [Computer software]. IBM Corp.

IBM Corp. (n.d.). Merging Data Files. Retrieved December 12, 2017 from https://www.ibm.com/support/knowledgecenter/en/SSLVMB_23.0.0/spss/base/idh_idd_add _gating.html

Ojeda, T., Dasgupta, A., Bengfort, B., & Murphy, S. P. (2014). *Practical data science cookbook.* Packt Publishing.

Rattenbury, T., Hellerstein, J. M., Heer, J., Kandel, S., & Carreras, C. (2017). *Principles of data wrangling: Practical techniques for data preparation.* O'Reilly Media.

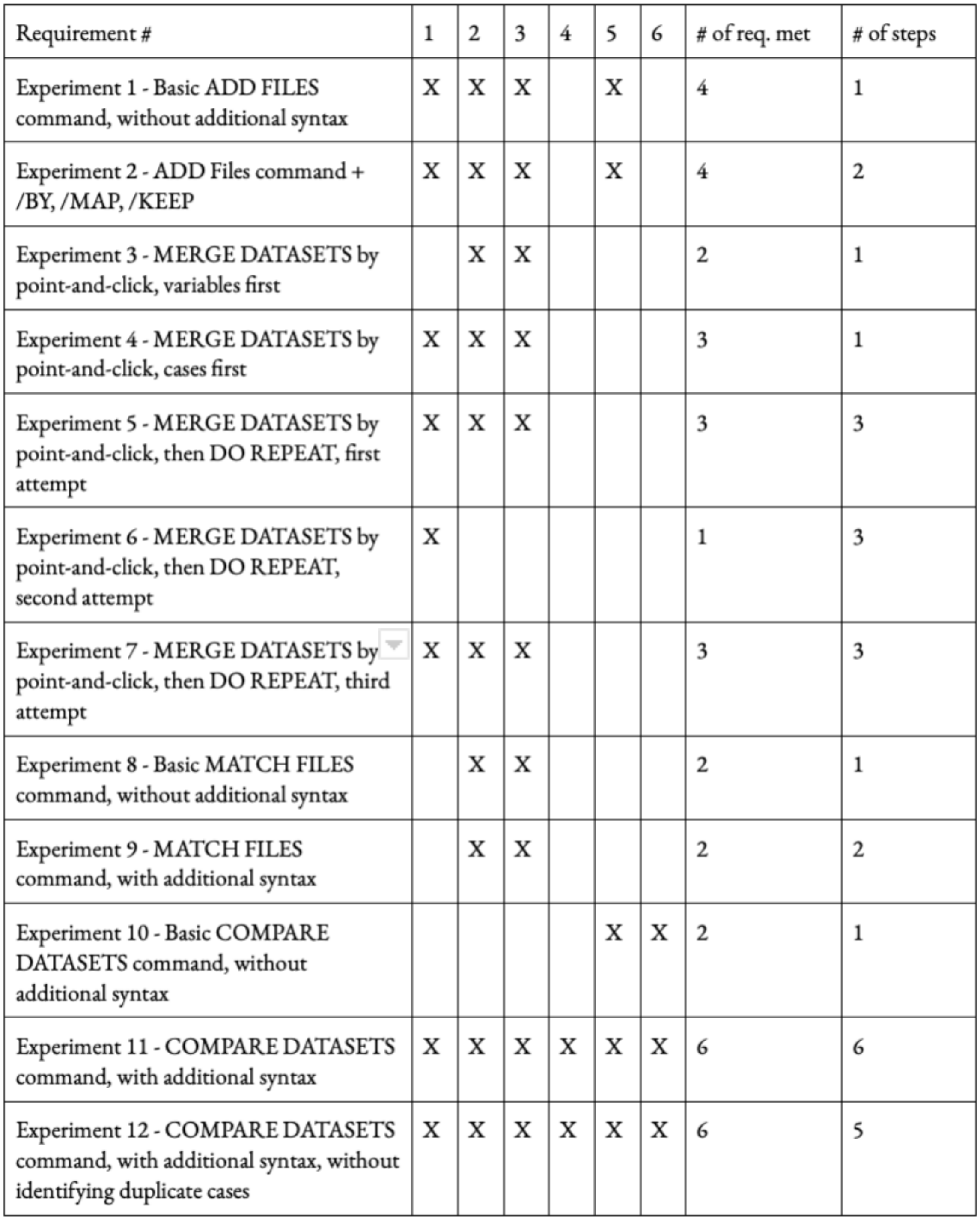
Spss-tutorials.com. (n.d.). *Merging Data Files in SPSS.* Retrieved December 12, 2017 from https://www.spss-tutorials.com/merging-data-files/

Stehlik-Barry, K. & Babinec, A. J. (2017). *Data analytics with IBM SPSS statistics.* Packt Publishing.

Truong, D. [Dothang Truong]. (2016, July 21). *SPSS missing values.* [Video]. YouTube. https://www.youtube.com/watch?v=DNThjjqLz9Q

Table 1

Experiments Conducted to Determine Best Syntax for Merging Divergent Datasets



Requirement #	1	2	3	4	5	6	# of req. met	# of steps
Experiment 1 - Basic ADD FILES command, without additional syntax	X	X	X		X		4	1
Experiment 2 - ADD Files command + /BY, /MAP, /KEEP	X	X	X		X		4	2
Experiment 3 - MERGE DATASETS by point-and-click, variables first		X	X				2	1
Experiment 4 - MERGE DATASETS by point-and-click, cases first	X	X	X				3	1
Experiment 5 - MERGE DATASETS by point-and-click, then DO REPEAT, first attempt	X	X	X				3	3
Experiment 6 - MERGE DATASETS by point-and-click, then DO REPEAT, second attempt	X						1	3
Experiment 7 - MERGE DATASETS by point-and-click, then DO REPEAT, third attempt	X	X	X				3	3
Experiment 8 - Basic MATCH FILES command, without additional syntax		X	X				2	1
Experiment 9 - MATCH FILES command, with additional syntax		X	X				2	2
Experiment 10 - Basic COMPARE DATASETS command, without additional syntax					X	X	2	1
Experiment 11 - COMPARE DATASETS command, with additional syntax	X	X	X	X	X	X	6	6
Experiment 12 - COMPARE DATASETS command, with additional syntax, without identifying duplicate cases	X	X	X	X	X	X	6	5

Appendix A

Syntax for Steps One Through Five

Step 1:

```
***Step 1***
*compare datasets
DATASET ACTIVATE DataSet1.
SORT CASES BY ID.
COMPARE DATASETS
   /COMPDATASET = DataSet2
   /VARIABLES ALL
   /CASEID ID
   /SAVE FLAGMISMATCHES=YES VARNAME=Mis-
matches MATCHDATASET=NO
MISMATCHDATASET=NO
   /OUTPUT VARPROPERTIES=NONE CASE-
TABLE=YES TABLELIMIT=600.
```

Step 2:

```
***Step 2***
*merging all variables and cases into DataSet1
ADD FILES /FILE=*
   /FILE='DataSet2'
   /IN=SourceDataset2.
VARIABLE LABELS SourceDataset2
   'Case source is DataSet2'.
EXECUTE.

ADD Value Labels
SourceDataset2
1 'from Dataset2'
0 'from Dataset1'.
Execute.
```

Step 3:

```
***Step 3***
*creating new mismatches variable called CasesCompared
compute CasesCompared = Mismatches.
Execute.
ADD Value Labels
CasesCompared
1 'Mismatch between Both Datasets'
0 'Match in Both Datasets'
-1 'Unique to Dataset1'.
Execute.
```

Step 4:

```
***Step 4***
*duplicating CasesCompared encoded information for
cases from Dataset2
SORT CASES BY ID(A) SourceDataset2(A) Mismatches-
(A).
Execute.
IF ((ID = lag(ID)) AND (sysmis(CasesCompared)))
CasesCompared=lag(CasesCompared).
Execute.
```

Step 5:

```
***Step 5***
*encoding in CasesCompared those cases which are unique
to Dataset2
IF (sysmis(CasesCompared))
CasesCompared=-2.
Execute.
ADD Value Labels
CasesCompared
-2 'Unique to Dataset2'.
Execute.
```

Graduate Student Journal of Psychology
2022, Vol. 19

Serving Whom? An Exploration of Cultural Taxation Themes in Latinx Graduate Students

Madasen Briggs, M.S., Jonathan Cajas, B.A., Audon Archibald, Ph.D., and Yolanda Flores Niemann, Ph.D.,
Department of Psychology, University of North Texas

Cultural taxation refers to the extra responsibilities (e.g., increased service expectations, serving as unofficial diversity consultants) placed on members of marginalized racial groups within academia. However, the extant literature on cultural taxation does not clearly indicate the extent to which that research applies to graduate students who often fulfill similar tasks as members of academia and faculty. Furthermore, the academic context of a Hispanic Serving Institution (HSI) is particularly relevant to this line of work as this designation, while projecting an image that an institution is inclusive, is not directly representative of the institution's commitment to equitable practices. The goal of the present research is to identify how Latinx graduate students are potentially taxed and whether their experiences of cultural taxation are analogous to the taxation that faculty experience. Transcripts from 20 interviews with Latinx students at an HSI were analyzed via qualitative thematic analysis to identify potential themes of this sample's experiences with cultural taxation. Findings suggest that Latinx graduate students' cultural taxation converge in some ways with faculty patterns, but with key exceptions. Major themes from students' responses include a sense of increased distance and lack of support from one's home department, as well as difficulties integrating diversity and inclusion work cleanly into one's graduate career. These findings supplement existing cultural taxation literature by highlighting the experiences of an underrecognized population within academia and presenting initial findings for taxations imposed on this group.

Keywords: cultural taxation, graduate students, Latinx students, identity taxation, thematic analysis

Cultural taxation refers to the elevated workload expected of academics from marginalized communities (Padilla, 1994). This elevated, but rarely compensated, workload often includes participation in diversity and equity work due to a faculty member's presumed knowledge of diversity, regardless of its relevance to their research expertise (Padilla, 1994; Joseph & Hirshfield, 2011; Cleveland et al., 2018). Among the most identifiable forms of taxation include: being expected to serve as an expert on all matters of diversity even when they are not relevant to one's expertise, being called on to educate the majority group on diversity issues outside of one's job description, serving on an affirmative action committee that only results in recommendations that have already been unmet, serving as a spokesperson from an ethnic community to the university, sacrificing time to serve as a multicultural problem solver, and being asked to serve as a translator or interpreter (Padilla, 1994; Gutiérrez y Muhs et al., 2012; Boyd et al., 2017; Njie-Carr et al., 2020).

In practice, cultural taxation could also be inclusive of discriminatory practices such as the tokenization of a department's marginalized members (Amie, 2020). Tokens are numerically or culturally underrepresented groups within an organization who are viewed as a symbolic representation of that group as a whole (Niemann, 2016). An example of this phenomenon would be a faculty member of color being pointed to by administration as a symbol of a department's commitment to diversity. While the perspectives and contributions of marginalized community members are vital to an organization's success (Page, 2007), involvement in such work that draws attention to their marginalized status can heighten the burden already in place by a standard academic teaching and research load. Faculty members expected to fulfill these extra demands have experienced stress, longer hours, and depression (Joseph & Hirshfield, 2011) as well as a difficult relationship with their professional identity (Amie, 2020).

The expectation that racialized[1] faculty will fulfill these additional responsibilities becomes more insidious when considering that White academics continue to occupy about 75% of all faculty positions (NCES, 2020; Pew Research, 2019). Even with a gradual increase in the diversity of faculty, student bodies continue to be more diverse by comparison (Pew Research, 2019). Already underrepresened, racialized faculty are also less likely to secure tenure or promotion in their institution, even after controlling for human capital, cultural taxation, and discipline (Wijesingha & Ramos, 2017).

---

[1] For the purposes of this project, "racialized" refers to those who are marginalized according to their racial group. This term was selected for consistency throughout the paper, but we recognize that "racialized" is an imperfect term that could falsely imply that Whiteness is not factored into the treatment of someone according to their racial group.

Increased expectations of racialized faculty impose barriers by taking time away from more personally important tasks, such as the research that is necessary to advance their careers (Domingo et al., 2022). Unfortunately, the combination of stagnant performance and increased responsibilities can lead to decreased work satisfaction, longer promotion times, and increased position resignation (Domingo et al., 2022). This translates to decreased representation among racialized faculty, which can negatively impact racialized student success and retention (Tram et al., 2020). Outcomes for students can include decreased program satisfaction due to unmet expectations, lack of financial support, and an inability to connect with faculty mentors who may not understand cultural needs (Tram et al., 2020). This domino effect speaks volumes to the importance of racialized faculty's presence in academia and students' vulnerability, particularly graduate students, who rely on a close relationship with faculty members.

While faculty have received much research attention, the extant literature is limited in directly addressing the unique positioning and circumstances of graduate students as a potentially culturally taxed population. Nearly thirty years ago, Padilla (1994) called for investigations on the effects of cultural taxation on marginalized graduate students. Padilla notes that graduate students face unique challenges in the pursuit of ethnic scholarship, finding mentorship that would accompany that work, and the inability to access all of an institution's resources (Padilla, 1994). However, clarification is needed on the ways in which graduate students with marginalized racial and ethnic identities are affected by and cope with cultural taxation, specifically in how they relate to the faculty model proposed by Padilla. Some culturally informed models identify common graduate student experiences, such as imposter phenomenon amongst Black graduate students at a predominately White institution (Stone et al., 2018). However, these models have typically failed to incorporate the perspective of Latinx[2] graduate students, a group sorely in need of targeted work given their status as the largest ethnic minority group in the United States. Testimonies from graduate students and early career academics reveal the immediacy with which we must address the interplay of graduate student status, identity, and cultural taxation (Templeton et al., 2021).

Lerma and colleagues (2020) elaborated upon forms of cultural taxation in student populations in the form of "racialized equity labor." This labor refers to the actions taken by marginalized students within an organization to address racial inequity that is then appropriated by those with more organizational power. The stress and demands from this cycle can force already marginalized individuals to relinquish additional resources, such as time, which may reflect poorer performance in professional and academic settings (Lerma et al., 2020). However, Lerma's innovative work 1) did not distinguish between the work of undergraduate and graduate students and 2) sought out participants specifically for their involvement in university change that benefits marginalized students. It is unclear how racialized equity labor, and by extension, the labor appropriation cycle outlined by Lerma (2020), occurs amongst graduate students who may objectively have more power but also may have more complex tethers to their universities than undergraduates (Grady et al., 2014). Like marginalized faculty, graduate students are often expected to put their own needs or goals aside for those of the department (Grady et al., 2014). Furthermore, members of Latinx communities already face unique challenges integrating into higher education, such as increased psychological and acculturative stress (Wang et al., 2016), and the highest likelihood of holding first-generation student status compared to their Black and White peers (PNPI, 2021).

Beyond Student Status: The Importance of Gender and Culture

Beyond student status, members of marginalized communities experience cultural taxation based on race and gender intersections (Hirshfield & Joseph, 2012; Diaz & Bui, 2016). Making space for these identity intersections within research is critical to appropriately framing people's experiences with nuance and avoiding the homogenization of an entire group's experiences. At the intersection of race and gender, women of color are often looked to as supporting pillars of their communities through the perpetual association with women, inherent communality, and nurturing expectations (Gutiérrez y

[2] For the purposes of this project, Latinx is used during each research stage as a gender-neutral term to refer to those of Latin American descent (e.g., Latino, Hispanic, Mexican American, Chicano, Puerto Rican, Cuban American, etc.). The authors want to recognize that, while the term has gained popularity within university student populations and higher education research, Latinx is an imperfect term that may contribute to the homogenization of individual cultures (Salinas & Lozano, 2017).

Muhs et al., 2012; Velásquez et al., 2004). Culturally taxing experiences for women of color in the professorate are further compounded by gendered ascriptions and ongoing professional disparities in academia. Women faculty are canonically more involved in departmental or university service work than their male colleagues (Niemann, et al., 2020; Njie-Carr, et al., 2020; O'Meara et al., 2017), which partially explains why women, on average, publish less and receive fewer research grants than men (Wijesingha & Ramos, 2017). Adding to their challenging experiences is the relatively small number of women of color in upper-level professor and administrative ranks (Sánchez et al., 2021).

Culturally specific circumstances may also shape experiences. Mexican Americans face marginalization within the United States that may lead to steep social consequences, such as feelings of alienation and 'othering' perpetrated through negative stereotypes (Hester et al., 2020; Olguin-Aguirre et al., 2022). Further, discrimination can be perpetrated through public policy that imposes limits on equitable access to quality education and health-related resources (Brenes, 2019; Shi et al., 2018). Arbona and colleagues (2010) theorize that Latinx individuals are especially vulnerable because their identity is uniquely affected by the simultaneous experience of generational cultural practices and of American society outside of the home. Although some Latinx academics may feel comfortable navigating a bicultural perspective, others, particularly those who have recently immigrated to the United States, could face increased adversity that may lead to acculturative stress (Berry, 1992; Meca et al., 2017).

The research outlined in this paper explores Latinx graduate students' experiences of cultural taxation at a Hispanic Serving Institution (HSI). While an HSI designation requires that a minimum of 25% of students enrolled be Hispanic (White House, n.d.), that does not guarantee representation at the graduate level. For example, Latinx graduate enrollment at this southwestern public university (SWPU) equals about half that of undergraduate enrollment, with even fewer faculty in place to support those graduate students (SWPU Factbook, 2020). Although an HSI designation presents opportunities to apply for federal funding to serve Latinx students, assessments indicate that universities often funnel these extra funds into colorblind programs that do not prioritize Latinx students (Vargas & Villa-Palomino, 2018). Investigations of faculty's cultural taxation within an HSI also indicate that this institutional designation does not protect against racism, marginalization, and taxation for faculty of color (Martinez et al., 2017). Beyond numerical enrollment, it is not clear what the term Hispanic Serving Institution means to the experiences of Latinx students.

The goal of the present study is to 1) understand how Latinx graduate students at an HSI identify with cultural taxation and related labor on campus and 2) identify how these experiences compare to faculty taxation. Specifically, we aim to identify the ways in which existing understandings of cultural taxation align with Latinx graduate students' experiences of cultural taxation unique to this population. Furthermore, we seek to understand how the intersections of race and gender play into Latinx women graduate students' experiences with cultural taxation. Findings from this work will help identify where this sample of Latinx graduate students falls in relation to existing cultural taxation literature. It is our hope that such knowledge will highlight pathways of action for programs and mentors in specific support of marginalized graduate students. Additionally, the current work seeks to bolster the understanding of graduate students' experiences as a population already subjected to exploitative practices surrounding their learning and labor (Cohen & Baruch, 2021).

Methods

Participants

A total of 20 Latinx graduate students participated, including 12 who identified as women, and eight who identified as men. Students represented over 10 different departments on campus seeking a variety of degrees, including both master's and Ph.Ds. Students came from a blend of business, STEM (Science, Technology, Engineering, Math), and social science programs with career tracks inclusive of both industry and academia. Participants' ages ranged from 22 to 48. Participants came from a range of Latinx backgrounds: 25% of the participants identified as having mixed heritage (e.g., Mexican and Salvadorian, mixed indigenous and Mexican American), 20% identified as Hispanic, 20% identified as Latinx, and the remaining participants identified as Mexican American or Argentinian.

Data collection took place during the Spring 2021 semester when many students had still not yet returned to face-to-face classes due to the COVID-19 pandemic. While questions related to participant experiences did not specify a particular timeframe, student responses are a mix of both their pre-pandemic and pandemic experiences on campus.

Research Design

A series of one-on-one, open-ended interviews were conducted and analyzed qualitatively. Open-ended questions were based on Padilla's (1994) cultural taxation framework, which specifies an increased burden of service and mentorship to students, colleagues, and the university faced by racialized faculty (Jacobs et al., 2002). Consideration was also given to other aspects of this phenomenon, such as its intersectional nature (Josephs & Hirshfield, 2012), to ensure that interviews were capturing the full breadth of graduate students' experiences with cultural taxation (Lerma et al., 2020; Blake, 2018). All questions were iterated over a series of meetings amongst the research team until the final set of 25 questions was reached (see Appendix B for list of interview questions). An example item from this list is "In what ways, if any, was your service rewarded, validated, or noticed?" To not lead participants to specific answers, questions asked participants about 'their identities' as opposed to specifying a particular racial or ethnic group. Participants who identified as women were given an additional set of four questions to discuss their experiences within the intersection of race and gender in graduate school. An example item from this set is "Do you feel as if expectations in your department are identical for men and women?"

It is important to note that the research team who conducted interviews, subsequent analyses, and authored this manuscript identify as follows: White nonbinary person, cisgender Latino/Hispanic man, cisgender biracial Mexican man, and cisgender Mexican American woman.

Procedure

The present project received institutional review board approval at a large southwestern public university (SWPU) during the Spring of 2021. Study invitations were sent to all graduate students informing them of an opportunity for Latinx graduate students to self-select into a study to discuss their experiences with race on campus. A request for Latinx students was used in the mass mailer, but participants were asked to self-report their racial/ethnic identity if their identity was not fully encapsulated by the term Latinx. Students were also given the opportunity to specifically request a Latinx interviewer due to the sensitive nature of the questions. Following recruitment, all participants took part in individual interviews with a member of the research team. All 20 interviews were conducted via Zoom, recorded, and transcribed live using Panopto. After the interview stage concluded, the authors reviewed the transcriptions produced by Panopto for any potential errors.

Data analyses followed Braun and Clarke's (2006) six phases for Thematic Analysis, a qualitative data analysis technique that identifies overarching themes and patterns of meaning in a particular dataset. The phases for thematic analysis are: 1) familiarizing yourself with your data, 2) generating initial codes, 3) searching for themes, 4) reviewing themes, 5) defining and naming themes, and 6) producing the report. Overall, due to the novelty of the work, we took a primarily inductive approach, meaning our codes and themes are derived from the semantic content of the data rather than coding the data according to pre-existing theoretical concepts (Braun & Clarke, 2012). While the research team was deeply familiar with pre-existing frameworks of cultural taxation, "our analytic lens does not completely override their [participants'] stories" (Braun & Clarke, 2012, p. 60). This analytic technique was preferred because of its theoretical flexibility compared to other techniques such as discourse analysis which requires theoretical framing for the role and meaning of language (Braun & Clarke, 2012).

With the final transcriptions, each author independently read and generated individual lists of codes for the data using Nvivo (version 12). For this technique, codes identify a specific feature of the dataset, and these codes are grouped into overarching themes. Coding involved the individual labeling of data segments that were potentially relevant to the research question. These independent code lists were reduced collaboratively based on codes that all four research team members collectively agreed were sufficiently present and meaningful in the data, resulting in a final list of over 40 codes. After code generation concluded, the research team sorted these codes into themes through collaborative discussion (see Table 1 for overview). Themes for this project were organized groups of codes that represented patterns and shared meanings across the dataset. Coder reliability was not as sessed quantitatively; these numeric itatively; these numeric measures of intercoder reliability are more common in qualitative work guided by post-positivistic values. We aligned with a paradigm that emphasizes an organic approach to the data and each coder's natural, reflexive response (Braun & Clarke, 2021).

Results

Five themes were identified using the above coding process. These themes are Distance/Isolation from Support; Power, Labor, and 'Something's Gotta Give;' Culturally Taxing Levies Against Latinx Student Identities; Unique Pressures on Women; and Reluctance to Assign Accountability and Responsibility. In general, participants reported that they were not entirely supported by their departments and largely abstained from diversity labor. Women participants reported undertaking more social and emotional labor than their male counterparts. Finally, participants provided evidence that they experienced cultural taxation (e.g., feeling that they are made into a representative for all Latinx peoples), but were hesitant to assign accountability to the parties contributing to these taxations.

An in-depth description of each theme and some supporting data points are provided below. We provide a brief overview of how each theme relates or diverges from existing literature; these points are further explored in the discussion. Throughout the following section, we refer to participants according to the identities they provided on their demographics questionnaire while still maintaining their confidentiality.

Theme One: Distance/Isolation from Support

Latinx interviewees acknowledged acute differences between the graduate and undergraduate experiences on campus. Many of these differences manifested as a physical and psychological distance from campus life and culture, as well as distance from potential campus support networks within and outside their department. This was compounded in no small amount by the realities of the COVID-19 pandemic in the United States, with many students lamenting that working from home and the stressors of the pandemic exacerbated the pre-existing lack of support for Latinx graduate students. One Hispanic graduate student speaks up about how departmental support for students of color, in general, does not necessarily fill in the needs that Latinx graduates display based on their cultural backgrounds: "...being Hispanic, my views are a little bit different. I don't think [departmental support] is Hispanic serving, I don't think it fulfills that. I think for minorities as a whole, [SWPU] does a great job with lots of inclusion, lots of opportunities for graduate students to work with faculty members... But for Hispanics, they love to preach how they're going to help and that they're going to support you... but in the ways they actually do it's very, very limited... [For example] when my father was dying of liver disease... and my mother, she had breast cancer at the same time, it's like, hey, I can't live on campus because I need to live close to my parents due to that whole cultural aspect of Hispanic children taking care of their parents. So, I said "Hey, I'm not going to be able to relocate," and they said, "Ok well if you can't fulfill this, you're out."

Faculty and administrators' seeming inability to understand the responsibilities many students felt to their identity and community was not unique to this student. Other students spoke more to how physical distance from campus life inhibited the formation of more support networks. For many of the Latinx graduate student interviewees, the realities of being disconnected from the physical campus began long before the pandemic (Erichsen & Bolliger, 2011). One nontraditional male graduate student who identifies as indigenous Mexican responded "I mean, I'm so disconnected because I'm online. I really have no idea how SWPU serves students of color. I don't know honestly... There's very little interaction, at least for me..."

These feelings of isolation were further characterized by students often being in a position where their existing support networks were challenged by unique graduate student circumstances. This lack of face-to-face support, physical distance from campus, and the lack of fellow Latinx academics in one's home department played a key role in why an interviewee believes many of their Latinx peers drop out. In addition, even for students who mentioned that they had a support network in place, it was often not the case that the support network was housed within the university. These students often sought support outside of campus, such as in the form of an international Latinx graduate support group, having been driven to this by the university's failure to provide them the support they needed.

These experiences with isolation speak to a key element of cultural taxation - the feeling of a lack of belonging in one's department due to the to kenizing nature of the lack of support for your racial and ethnic background (Joseph & Hirschfield, 2011). The common mentor/mentee model did not alleviate the isolation felt by many graduate students, especially when they sought guidance. For instance, a woman graduate student of mixed Latinx ancestry speaks about how she does not have someone in her department she can approach: "Unfortunately, no, I don't...even with my research mentor, I have asked for a lower workload in the past because I feel very overwhelmed and basically got told no. I learned very early on that I can't trust to reach out to faculty for support ... I have to be very direct to be able to get the support that I need and when I'm not, it's not as helpful. So, it's hard. When I don't know the questions to ask, then I end up not getting the help I need... And the growth that I've had, the progress I make, I always have to approach them for it. It's always about self-advocacy, rather than them bringing it to you in a potential teachable moment."

Some students were not willing to ascribe this lack of mentorship as a problem (i.e., students saying they hated to bother faculty or that they believed faculty were inherently too busy to be available for graduate students). However, students still indicated that they felt their mentors often failed to discuss information crucial to their degree advancement or failed to take notice of their major milestones. For instance, another half-Hispanic, half-White woman graduate student contributes a sentiment shared by most of the graduate student participants under the mentor/mentee model: "Looking at my department, my advisor talks to me once a semester... a lot of students in my department especially feel that we have not been supported through the [graduate student] process."

This lack of mentorship is often present for junior full-time faculty members, but the additional power differential felt by graduate students, specifically in relation to the completion of their degrees, suggests a potential gap between student and faculty models of cultural taxation.

Theme Two: Power, Labor, and 'Something's Got to Give'

Perhaps the most striking difference in how cultural taxation manifests in graduate students, compared to other members of academia, is in labor expectations. While faculty often undergo heightened expectations of racialized equity labor that their non-White peers do not experience (Padilla, 1994), respondents did not spontaneously express the same expectation. When asked specifically about how they participated in diversity labor and advocacy, surprisingly, most interviewees indicated that they did not commonly participate in racialized equity labor. Their reasons included: fear of retaliation that threatened their positions in the department, lack of knowledge of where to begin advocacy work, and just wanting to 'keep their heads down' during the turbulence of graduate school. This also bares an interesting difference with how other students' experiences of cultural taxation often manifest. As outlined by Lerma (2020), students whose personal and professional lives suffer for their involvement in racialized equity labor are often forced to let their academic performance drop due to the effort it takes to support campus change.

In contrast, respondents did not believe that a drop in academic or professional performance in exchange for engaging in racialized equity labor was a valid option. One female student with a mixed indigenous and Mexican American identity states regarding proposed bans of Critical Race Theory in U.S. schools, a politically charged discussion of which students were aware: "What makes me want to pause is being aware of legislature with what they want to do about race theory, so that does make me consider that there will be retaliation linked to [diversity] work... if this happens, it is going to come down to my professional or educational experiences being impacted... and I just can't let that happen."

When they did engage in racialized equity labor, more often than not interviewed graduate students preferred to conduct it 'behind the scenes' or even off-campus entirely. This came from, in part, the disconnection they felt from the campus community (see theme 1) as well as not wanting to jeopardize their standing in their programs. The same woman graduate student who commented on the lack of adequate mentorship in Theme 1 elaborates, "It's hard to get involved... I think I've just kind of been working behind the scenes, here for support for people who have negative experiences on campus... I've been approached to be recruited for some leadership positions [in organizations that support students of color], but with my professional life and mental health, I just didn't want to do it... I just want to graduate and move on."

Respondents often had concerns about engag ing in racialized equity labor at the forefront of their experiences. That said, unlike undergraduates of color who are perhaps more likely to 'give up' their academic work to support diversity labor on campus, or racialized faculty who are compelled to participate in diversity labor as an aspect of their employment, our study respondents have trouble being as active in campus equity efforts as they would like to be.

It is here that we see a unique manifestation of how cultural taxation may have been realized for this specific sample of Latinx graduate students. As much as they would like to participate in more diversity labor, many fear 'rocking the boat' and jeopardizing their position on campus. Simultaneously, graduate students may not be afforded as many chances to engage in racialized equity labor as faculty (even if their participation is tokenizing and directly conducive to culturally taxing experiences). This lack of engagement can turn into a threat to one's cultural identity and affect their overall graduate school trajectory as one Indigenous Mexican student puts it: "No, personally, I don't engage in work on behalf of campus diversity efforts... I learned that keeping people at ease gets you far, you've got to know how to keep people from calling the cops and, you know, not give them a reason to feel threatened, but when you start realizing that doing so definitely changes you... I've lost so much. I've had to sacrifice my own identity to, you know, stay alive... worrying about everything that's happened distracts me so much from grad school. You'd hope that this is something that only comes up every so often, but no, it comes up every day."

Theme Three: Culturally Taxing Levies Against Latinx Student Identity

A recurring element of responses was difficulty navigating their professional workload due to levies, or unique impositions, against Latinx identities. Students spoke about how their Latinx identity, rather than their general identity as students of color, was a specific stressor when managing their graduate careers. Navigating their program responsibilities was also difficult due to the mixed-race background of many participants, who had their diversity-centric research interests undermined due to 'not being Hispanic enough' for other scholars to take their work on Latinx populations seriously. One student relayed an anecdote whereby a professor requested he change his research topic to something more 'suitable' to him due to 'passing' as a person of color. This highlights how many non-Latinx academics may be incorrectly assuming what Latinx students 'should' present as and how that should inform their academic interests. Additionally, many interviewees highlighted how 'Latinx' was not a term with a lot of explanatory power of their cultural backgrounds in the first place. Thus, it was often difficult to navigate feelings surrounding their identity (a key feature of the faculty model of cultural taxation) when treated as though they are 'generic.' Specifically, interviewees expressed not wanting to be forced to be a representative of all Latinx people. For instance, a graduate student from South America comments, "They confuse our different cultures and there's this really racist notion that everything south of the U.S. is Mexico... Often times they would ask me about Mexican stuff. I always have to clarify that I'm not Mexican. It's like constantly trying to tell them you're not Mexican and having to prove that we [people from my country] have different cultures and things, and all I do is get asked 'do you eat spicy food?'"

Interviewees also expressed concerns about the lack of support for first-generation and immigrant students, especially in the face of severe underrepresentation of these persons amongst faculty and staff in their graduate programs. These feelings of underrepresentation and lack of support were especially apparent in those that come from communities that were majority-Latinx. For these students, campus life brought the additional stress of being separated from a community driven by commonalities to which they were accustomed, with accompanying feelings of isolation vis a vis a culturally taxing experience. Students also often did not have access to specifically Latinx faculty and staff who understood their experience. As a result, students found themselves facing additional levels of solitude (see theme 1) not necessarily felt by those who were more used to majority-White settings. For instance, a Hispanic woman states, "I mean, I would not call myself a representative where I live in [city with Latinx majority population] because it's a minority majority...But as far as in [City of University] or as far as campus goes... my name doesn't fit in, I don't know where to go, and suddenly I'm looked upon differently."

This sentiment was echoed by other stud

dents who agreed that going from an environment where they are not looked upon as a representative of a race or culture into an environment where it is very easy to be perceived as one was jarring and disruptive to their work. Here we see a relevant, more Latinx-specific sentiment that may modulate how culturally taxing experiences are understood by Latinx graduate students. Many do not necessarily come from backgrounds where they built an understanding of their identity as coming from a 'minority group,' often traveling to academia from environments that were majority Latinx. It is specifically within academia that these students are exposed to, even at HSIs, White-dominated space.

Theme Four: Unique Pressures on Women

Women-identifying interviewees were aware of the link between their identity as women, their Latinx identity, and the social perceptions of these intersecting identities. Pressures on these interviewees were further compounded by the awareness of professionalism or leadership expectations that seem to conflict with womanhood and femininity. For instance, a woman respondent said, "But once you get into the ranks of leadership, that was always historically men. And so, you have to prove that while retaining your femininity and your womanhood, you can be an effective leader. And that may look a little bit different than it would on a male."

However, the way women responded to these unique pressures varied. Some women felt the need to stifle certain forms of expression to adhere to a more traditional archetype of professionalism. The mixed-race student who previously commented on isolation and keeping a low profile in Themes 1 and 2 elaborates, "Like facial expressions and like expressing emotion, like not wanting to come across too harsh, like continuing to be like smiley and happy...as women we have to like teeter the line of being like happy and smiley, but also not being, you know, too assertive... I think adjusting, knowing when to elevate their femininity to help in like graduate school or the workplace and knowing when to kind of dial it down to where it's like, well, she's just a woman... female students have been called too emotional by faculty... finding that balance has been the difficult part."

Further, a Latina student offers, "I have made it a point to, like in my emails, to be more definite in the way I talk and less kind. Like. I don't want to say submissive...but less submissive...I do feel the need to be more like of a dominant person or like, I don't want to say manly, but like more just serious."

Conversely, other women, even though they were aware that their womanhood and femininity might be hypervisible, refused to modify their behavior or expressions of femininity. "And like I said, I don't and I wouldn't want to suppress it because I'm Mexican, I just can't help myself." One interviewee also describes the ways in which her identity as a woman is pressured externally both from family and the university: "Women especially have this responsibility [caring for family]... And especially being from a from a Hispanic background, that's definitely the case for myself...And so it's just really unfortunate because being a Hispanic female, trying to go through graduate school, it's almost like a double hit...Not only am I hearing it from my traditional family members, 'are you not helping take care of your parents or sick elderly parents?' But I'm also hearing from the graduate school saying, 'oh, no...this is our role. You have to stick to them.'"

While some participants felt that their departments' standards were equal for both male and female graduate students, other interviewees noted tangibly different standards, both social and procedural. When discussing faculty responses to female students' work, one Mexican woman interviewee said, "I think there has to be more behind the scenes work to present it and be taken the same way as if like a male student did." Regarding other departmental standards, another Hispanic interviewee said, "I think men are given a little bit more academic freedom to express themselves, to pursue more things, to take on more and to take on more diverse research, whereas women are kind of put into a box."

These patterns primarily align with previously discussed findings for women of color in academia, either as students, staff, or faculty members (Gutiérrez y Muhs et al., 2012; Hirshfield & Joseph, 2012; Anantachai & Chelsey, 2018; Niemann et al., 2020). Responses indicate that Latinx women engage in labor intended to balance their personal identities with others' expectations and face unique identity challenges as graduate students. This is evidenced in the frequency that interviewees discuss personal and cultural obligations to engage in these forms of labor in the context of not being adequately represented in one's department amongst students and faculty. One Latina interviewee stated, "I'm the only one of me. I never saw anybody that I can really think of, I saw one other Hispanic girl in all my classes this semester, but that was it."

That same interviewee also alluded to Joseph and Hirshfield's (2012) discussion of the ways that negative perceptions of affirmative action programs can cast a shadow on women when she says, "I did have once a guy tell me he thinks that women have more opportunities...which kind of bothered me or definitely bothered me because that's not true." Multiple women also discussed how they felt their research or intellectual contributions were devalued because they involved discussions of race or diversity. One woman interviewee who identifies as White Latino and Mexican states, "That [area of expertise] gets downplayed...And then also being like a female and also being like Latina, like not knowing what the issues actually are. Even though they disproportionately affect Latinos ...And so I'm not being taken seriously on all those accounts."

Theme Five: Reluctance to Assign Responsibility and Accountability

Another theme identified by the research team was the apparent reluctance of Latinx graduate students to label their experiences and the actions of others as potential instances of racial prejudice or systemic problems. When asked if they had experienced prejudice or discrimination on campus, some participants qualified their answers as being 'not serious,' being unsure if it was prejudice, or otherwise not qualifying events as instances of racial prejudice, despite then proceeding to acknowledge that these events may have been racially motivated or inducive to harm for people of a Latinx background. For instance, a student who identified as half Mexican and half White said, "No, I really haven't [experienced discrimination], for my race or ethnic background... but you know, there are aspects of your identity that causes others to discriminate or be prejudiced towards you... I guess like, with my teachers, they say things that are offensive in a classroom [to Latinx students] ... I don't know if it counts, but there's definitely been like things that have been said or done that are kind of crappy, but actually not really offensive to me specifically."

Students also seemed hesitant to identify behavior as prejudiced when it was more systemic or covert in nature. Interviewees often went to great lengths to provide the benefit of the doubt, or outright deny the possibility their experiences may have been based on systemic issues. For instance, one Mexican woman states, "I couldn't really say [if I had experienced prejudice on campus] ... but it's not like they're going to say, 'we're going to discriminate against you.' But you see the inequalities, you see what's happening with the person next to you and yourself... I see it, but I can't prove it."

One other Hispanic man describes that, "I've never experienced any racism or discrimination within campus... I was just given a bad set of cards in my hand and well, tough luck right? No one ever said life was fair. So sometimes you've got to put in the extra work, you got to put in the extra mile, but it's not always easy, right? I think I remember sometimes where I was like man, I wish I was born a different way, I wish my family would have prepared for me to come to a college where I didn't have to worry about work..."

When asked to elaborate on their experiences, several students cited examples that included, but are not limited to: feeling pressured to be more prepared for classroom discussion than White male colleagues who 'got away' with not doing their homework or the reading, being forcibly labeled with terms they actively did not associate with (such as Chicanx or Latinx), having to defend their interest in diversity research, having to defend their expertise in work unrelated to diversity, and feeling like prejudice they experienced was more a fact of reality for them rather than something with a solution that they or the university could pursue. One Mexican woman student describes, "Regardless of what I do... People are going to see the color of my skin, and they're going to be prejudiced... but I know I can't advocate for me or my people... this doesn't feel like a place where the perception of you, or the perception of your classmates is ever going to change."

This very same student in an earlier response, however, professed that they believed the university to be doing very well in supporting students of color, claiming, "The university is basically doing all the things they can to really increase the visibility of the [student of color] population, and something I pride myself in as a student of the university."

In applying this absolution to cultural taxation, a key distinction between the Latinx graduate, faculty, and undergraduate experiences may be unveiled here: graduate students may be more likely to undergo the mental and emotional labor associated with witnessing inequities without finding culpable players in one's environment. This is not to say tgraduate students deny that action is important (see theme 2). Rather, in light of a situation in which participating in diversity labor is something that is not made accessible, it may be safer for students to draw their focus away from areas in need of improvement.

Discussion

The present study explores whether current understandings of cultural taxation, often modeled from faculty's experiences, are applicable to Latinx graduate students. Findings suggest that there are aspects of Padilla's model of cultural taxation that apply, but that universal application across faculty and graduate students may not be appropriate. Themes identified in this sample suggest that there are power differentials at play that distinguish cultural taxation in graduate students from that of other academic populations. Specifically, like faculty, present Latinx graduate students outlined feelings of isolation that were often intensified by their origination from majority-minority settings and the lack of Latinx colleagues. Students also cited fatigue due to perceived slights against their identities that stemmed from being asked to represent and/or work towards diversity efforts that may or may not have been of interest to them professionally. However, diverging slightly from faculty's experiences, we see power differentials play a more direct role due to students' reliance on a direct supervisor with stark levels of control over their graduate experience (and, to an extent, livelihood). These power differentials are further exacerbated by strong ties to university administration, and all these tethers could inhibit the ease with which graduate students were able to participate in equity labor.

Isolation in one's department reflects patterns in racialized faculty outlined by Padilla (1994). Both graduate students and faculty must overcome the isolation that accompanies statistical underrepresentation throughout the department. Tokenism may also be an underlying commonality between faculty and graduate student experiences. Among the consequences of tokenism are isolation, representativeness, lack of regard for race-based research areas, attributional ambiguity (which refers to not knowing how to interpret feedback), and loneliness (Niemann, 2016), all of which were reflected in students' responses in this study.

Our sample's difficulty in incorporating diversity into their work while navigating departmental dynamics is reflective of other graduate students' testimonies. Carbajal (2021) described the ways in which multiple members of his predominately White department referred to diversity and identity characteristics as nuisance variables. Discussions surrounding diversity have a steep impact on departmental and program culture, particularly for graduate students (Grady et al., 2014). Further, when faculty members are dismissive of related work or efforts, it can discourage the graduate student and add to the isolation they feel. Unfortunately, this can lead graduate students to feel that they do not have a trusted resource who can identify with their experiences to help their navigation through academia and professional skill development (Moore et al., 2020). A graduate student's livelihood is also dependent on maintaining a positive relationship with the department and, by extension, the administration. Racialized graduate students not only hold this tenuous relationship but also must endure the additional costs of being a racialized person within a predominately White space (Vargas & Villo-Palomino, 2018).

Levies or taxations against our sample of Latinx students seem to stem from multiple sources. First, the homogenization of all identities that are supposed to be encompassed by the 'Latinx' label places students in a precarious situation where they are motivated to defend themselves and their individual identity while also maintaining social harmony. This is reflective of impositions on racialized faculty who are presumed to serve as a 'cure-all' for diversity needs, recycling the racist notion that one person of color can serve as a single representative of a department's diversity values (Lopez, 1997; Niemann, 2016). However, our sample engaged in an extra layer of mental and emotional labor that sought to eclipse the university's role in inequitable practices. In application, this theme likely occurs simultaneously with both the theme of isolation and the theme of power differentials: graduate students do not feel the necessary sense of community nor the motivational autonomy to go head-to-head with powerful players in the department or administration. Therefore, to help calm their senses of personal or cultural obligation, they 1) do not seek out participation in equity labor that could put their security at risk and 2) do not place blame on those who could instigate their educational and professional destruction.

Concerning the taxation of Latinx graduate women, nearly all women participants cited unique pressures to their combined gender and cultural identities during their time as graduate students. Pressures associated with identity as it relates to womanhood and professionalism are a common challenge due to the antithetical associations between femininity and what is considered professional (Lewis, 2011). Women in working environments often feel pressured to endorse just enough womanly behaviors to avoid disconfirming feminine stereotypes while simultaneously conforming to rigid standards of professionalism (Rudman & Phelan, 2008). While women hold more than half of all graduate degrees and outnumber men in assistant professorships, academia still perpetuates these disparate standards through the overrepresentation of men in both full professorships and administrative positions (Lee & Won, 2014). Identity pressures are further exacerbated for women of color as they balance expectations based on both race and gender in addition to navigating often restrictive systems of professionalism (Rudman & Glick, 2001; Eagly & Carli, 2003; Brescoll & Uhlmann, 2008). These pressures may be exacerbated by Latinx cultural values about the primary role of women as wives and mothers.

Future research should explore other marginalized racial and ethnic groups within the graduate student population to identify unique or overlapping needs and experiences as they relate to cultural taxation. While we did not find explicit evidence of Lerma's labor appropriation cycle in our sample, this framework should still be explored in samples of graduate students (i.e., those that are actively involved in leading university diversity efforts). More work is also needed to understand the unique position of women graduate students of color, particularly regarding academic and labor expectations. While there can be similarities amongst racialized groups' experiences, researchers should take care to disaggregate their methods and analyses regarding race, gender, and cultural experiences. A simple step in this direction could be allowing participants the opportunity to provide their racial, ethnic, and gender identities in a free-response format rather than relying on normalized checkbox questionnaires that could limit participants' expression.

Limitations

There are limitations to this project that must be addressed. There may be other experiences within this population that were not presented by our limited sample. Additionally, this work relied entirely on subjective qualitative analyses. While qualitative work allows for the nuanced depiction of participants' lived experiences, these topics could be supplemented with quantitative work (e.g., a cross-university survey that could illustrate these experiences proportional to enrollment). Similarly, while quantitative measures of intercoder reliability are often considered inappropriate for judging qualitative work (Braun & Clarke, 2013), our choice not to implement such a measure may influence others' perceptions of our interpretations. Finally, data only reflect the experiences of Latinx graduate students at one university in the southwestern United States. It is possible that manifestations of cultural taxation for Latinx graduate students could fluctuate according to the region and racial and ethnic makeup of the surrounding areas.

Conclusion

Cultural taxation and related models must more inclusively account for the graduate student experience, including the psychological and physical toll that may result from tokenization, lack of cultural knowledge, and seeming practical irrelevance of the word "serving" for HSI designations. Within these contexts, students are driven to not only advocate for themselves as they navigate their degree but to simultaneously avoid rocking the boat in the hopes that they can graduate and make a difference in their families and communities with their careers. Furthermore, students may be driven to expend mental and emotional labor absolving their institutions of blame in the face of inequity due to their unique position as not-just-students and not-quite-faculty. We invite scholars to incorporate mental and emotional burdens, as well as related phenomena like tokenism, into their conceptualization of cultural taxation. The cognitive work associated with processing a discriminatory or culturally exclusive environment (e.g., questioning one's experiences and coping with isolation and discrimination) undoubtedly contributes to the invisible work conducted by marginalized members of the department.

The dearth of culturally sensitive faculty mentors exacerbates the challenges of Latinx students. For instance, despite being in an area of the country where nearly 40% of citizens self-identify as Latinx, graduate student enrollment at this project's SWPU lags significantly behind this number at 15%, which also falls be-

Factbook, 2020). Starker still is how these numbers compare to the presence of Latinx faculty. Inclusive of lecturers, tenure-track, and tenured professors, only 6% of over 1000 faculty identified as Latinx, an underrepresentation that directly impacts the mentor-mentee model that many graduate programs employ. Without knowledgeable mentors who can help Latinx students navigate and interpret feedback and experiences, students may live in a state of attributional ambiguity. That is, they may not know whether their experiences are based on their accomplishments, successes, writing, skills, etc., or whether they are a function of racism, sexism, homophobia, and other biases.

Finally, universities and institutions should not only take care to ensure that graduate students have access to the resources and mentorship that they need, but also be aware that, even within the context of an HSI, students could be challenged by culturally taxing experiences due to tokenism, isolation, heightened need for self-advocating labor, and sacrificing participation in university change. Institutions must move beyond upholding diversity for accreditation and honor, or funding, as is the case with an HSI designation. It is time for university administrators to pay particular attention to the word "Serving" in the label Hispanic Serving Institution.

References

Amie, K. N. (2020). *The experiences of underrepresented faculty and administrators: A study of cultural taxation* [Doctoral Dissertation, Florida International University]. FIU Electronic Theses and Dissertations. https://digitalcommons.fiu.edu/etd/4407

Anantachai, T., & Chesley, C. (2018). The burden of care: Cultural taxation of women of color librarians on the tenure-track. In Chou, R. & Pho, A. (Eds.), *Pushing the margins: Women of color and intersectionality in LIS* (pp. 301-327). Sacramento, CA: Library Juice Press.

Arbona, C., Olvera, N., Rodriguez, N., Hagan, J., Linares, A., & Wiesner, M. (2010). Acculturative stress among documented and undocumented Latino immigrants in the United States. *Hispanic Journal of Behavioral Sciences, 32(3),* 362–384. https://doi.org/10.1177/0739986310373210

Berry, J.W. (1992). Acculturation and adaptation in a new society. *International Migration, 30,* 69-85.https://doi.org/10.1111/j.1468-2435.1992.tb00776.x

Blake, D. (2018). Motivations and paths to becoming faculty at minority serving institutions. *Education Sciences,8(1).* https://doi.org/10.3390/educsci8010030

Boyd, B., Caraway, S. J., & Niemann, Y.F. (Eds.). (2017).*Surviving and Thriving in Academia: A Guide for Members of Marginalized Groups.* Washington, DC: American Psychological Association.

Braun,V.,&Clarke,V.(2006).Usingthematicanalysisinpsychology. *Qualitative Research in Psychology, 3(2),* 77–101. https://doi.org/10.1191/1478088706qp063oa

Braun, V., & Clarke, V. (2012). Thematic analysis. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), APA handbook of research methods in psychology, Vol. 2. Research designs: *Quantitative, qualitative, neuropsychological, and biological* (pp. 57–71). American Psychological Association. https://doi.org/10.1037/13620-004

Braun, V., & Clarke, V. (2013). *Successful qualitative research: A practical guide for beginners.* London: Sage.

Braun, V., & Clarke, V. (2021). Can I use TA? Should I use TA? Should I not use TA? Comparing reflexive thematic analysis and other pattern-based qualitative analytic approaches. *Counselling & Psychotherapy Research, 21(1)*, 37–47. https://doi.org/10.1002/capr.12360

Brenes, F. (2019). Hispanics, mental health, and discriminating policies: Brief report. *Hispanic Health Care International, 17(4)*, 178–180. https://doi.org/10.1177/1540415319875103

Brescoll, V. L., & Uhlmann, E. L. (2008). Can an angry woman get ahead? *Psychological Science, 19(3),* 268-275. doi:10.1111/j.1467-9280.2008.02079.x

Carbajal, I. (2021). "They called diversity a nuisance variable". In Templeton, E., Love, B. H., & Johnson, O. (Eds.), *Elevating marginalized voices in academe: Lessons for a new generation of scholars.* Routledge.

Cleveland, R., Sailes, J. D., Gilliam, E., & Watts, J. (2018). A theoretical focus on cultural taxation: Who pays for it in higher education. *Advances in Social Sciences Research Journal, 5(10)*, 95–98. https://doi.org/10.14738/assrj.510.5293

Cohen, A., & Baruch, Y. (2021). Abuse and Exploitation of Doctoral Students: A Conceptual Model for Traversing a Long and Winding *Road to Academia. Journal of Business Ethics, 180,* 505-522. https://doi.org/10.1007/s10551-021-04905-1

Diaz, T., & Bui, N. H. (2016). Subjective well-being in Mexican and Mexican American women: The role of acculturation, ethnic identity, gender roles, and perceived social support. *Journal of Happiness Studies, 18,* 607-624. https://doi.org/10.1007/s10902-016-9741-1

Domingo, C. R., Gerber, N. C., Harris, D., Mamo, L., Pasion, S. G., Rebanal, R. D., & Rosser, S. V. (2022). More service or more advancement: Institutional barriers to academic success for women and women of color faculty at a large public comprehensive minority-serving state university. *Journal of Diversity in Higher Education, 15(3),* 365-379. https://doi.org/10.1037/dhe0000292

Eagly, A. H., & Carli, L. L. (2003). The female leadership advantage: An evaluation of the evidence. The Leadership Quarterly, 14, 807-834. doi:10.1016/j.leaqua.2003.09.004Erichsen, E. A., & Bolliger, D. U. (2011). Towards understanding international graduate student isolation in traditional and online environments. *Educational Technology Research and Development, 59(3),* 309–326. https://doi.org/10.1007/s11423-010-9161-6

Grady, R. K., La Touche, R., Oslawski-Lopez, J., Powers, A., & Simacek, K. (2014). Betwixt and between: The social position and stress experiences of graduate students. *Teaching Sociology, 42(1),* 5–16. https://doi.org/10.1177/0092055X13502182

Gutiérrez y Muhs, G., Niemann, Y.F., González, C., & Harris, A. (Eds.). (2012). Presumed incompetent: *The intersections of race and class for women in academia*, University Press of Colorado and Utah State University Press (an imprint of University Press of Colorado). ISBN: 978-0-87421-922-7.

Hester, N., Payne, K., Brown-Iannuzzi, J., & Gray, K. (2020). On intersectionality: How complex patterns of discrimination can emerge from simple stereotypes. *Psychological Science, 31(8),* 1013–1024. https://doi-org.libproxy.library.unt.edu/10.1177/0956797620929979

Hirshfield, L. E., & Joseph, T. D. (2012). 'We need a woman, we need a black woman': Gender, race, and identity taxation in the academy. *Gender and Education, 24(2), 213–227.* https://doi.org/10.1080/09540253.2011.606208

Jacobs, L., Cintrón, J., & Canton, C. E. (2002). *The politics of survival in academia: Narratives of inequity, resilience, and success.* Lanham, MD: Rowman & Littlefield.

Joseph, T. D., & Hirshfield, L. E. (2011). "Why don't you get somebody new to do it?" Race and cultural taxation in the academy. *Ethnic and Racial Studies, 34(1), 121–141.* https://doi.org/10.1080/01419870.2010.496489

Lee, Y.-Joo, & Won, D. (2014). Trailblazing women in academia: Representation of women in senior faculty and the gender gap in junior faculty's salaries in higher educational institutions. *The Social Science Journal, 51(3),* 331–340. https://doi.org/10.1016/j.soscij.2014.05.002

Lerma, V., Hamilton, L. T., & Nielsen, K. (2020). Racialized equity labor, university appropriation and student resistance. *Social Problems, 67(2),* 286–303. https://doi.org/10.1093/socpro/spz011

Lewis, P. (2011). The search for an authentic entrepreneurial identity: *Difference and professionalism among women business owners. Gender, Work & Organization, 20(3), 252–266.* https://doi.org/10.1111/j.1468-0432.2011.00568.x

Martinez, M., Chang A., & Welton A. (2017). Assistant professors of color confront the inequitable terrain of academia: a community cultural wealth perspective, *Race Ethnicity and Education, 20(5),* 696-710, doi: 10.1080/13613324.2016.1150826

Meca, A., Cobb, C., Xie, D., Schwartz, S. J., Allen, C., & Hunter, R. (2017). Exploring adaptive acculturation approaches among undocumented Latinos: A test of Berry's Model. *The Counseling Psychologist, 45(8),* 1115–1140. https://doi.org/10.1177/0011000017745978

Moore, S. E., Hines-Martin, V. P., & Gattis, M. N. (2020). Paying it forward: The role of senior Black faculty in preparing junior faculty and Black doctoral students for career succes*s. Journal of Negro Education, 89(2)*, 146–157.

NCES (National Center for Education Statistics). (2020). "Characteristics of Postsecondary Faculty." Washington, DC: U.S. Department of Education.

Njie-Carr, V. P. S. Niemann, Y.F., & Sharps, P. W.

(Eds.). (2020). Disparities in the academy: *Accounting for the elephant.* Allegheny County, PA: Dorrance Publishing. ISBN: 978-1-6461-0641-7.

Niemann, Y. F., Gutiérrez y Muhs, G., & González, C. (Eds.). (2020) Presumed incompetent II: *Race, class, power, and resistance of women in academia*. Logan, UT: University Press of Colorado and Utah State University Press (an imprint of University Press of Colorado).

Niemann, Y. F. (2016). The social ecology of tokenism in higher education institutions. *Peace Review: A Journal of Social Justice, 28 (4)*, pp 451-458.

O'Meara, K., Kuvaeva, A., & Nyunt, G. (2017). Constrained choices: A view of campus service inequality from annual faculty reports. *The Journal of Higher Education, 88(5),* 672-700. http://dx.doi.org/10.1080/00221546.2016.1257312

Olguin-Aguirre, R., Boals, A., Niemann, Y.F., & Wang, D.C. (2022). Impact of potential changes in U.S. immigration policies on distress and mental health of Latinx college students. *Journal of Latinos and Education.* https://doi.org/10.1080/15348431.2022.2117700

Padilla, A. M. (1994). Ethnic minority scholars, research, and mentoring: Current and future issues. *Educational Researcher, 23(4)*, 24–27. https://doi.org/10.2307/1176259

Page, S.E. (2007). Making the difference: applying the logic of diversity. *Academy of Management Perspectives, 21(4),* 6-20. https://doi.org/10.5465/amp.2007.27895335

Pew Research Center, Washington, D.C. (July 31, 2019). *College faculty have become more racially and ethnically diverse, but remain far less so than student*s.https://www.pewresearch.org/fact-tank/2019/07/31/us-college-faculty-student-diversity/.

PNPI. (2021, February 1). *First Generation Students.* https://pnpi.org/first-generation-students/.

Rudman, L. A., & Glick, P. (2001). Prescriptive gender stereotypes and backlash toward agentic women. *Journal of Social Issues, 57(4),* 743-762. doi:10.1111/0022-4537.00239

Rudman, L. A., & Phelan, J. E. (2008). Backlash effects for disconfirming gender stereotypes in organizations. *Research in Organizational Behavior, 28,* 61-79. doi: 10.1016/j.riob.2008.04.003

Salinas, C., & Lozano, A. (2017) Mapping and recontextualizing the evolution of the term Latinx: An environmental scanning in higher education. *Journal of Latinos and Education, 18(4), 302-315.* https://doi.org/10.1080/15348431.2017.1390464

Sánchez, B., Salazar, C., & Guerra, J. (2021). "I feel like I have to be the whitest version of myself": Experiences of early career Latina higher education administrators. *Journal of Diversity in Higher Education, 14(4),* 592–602. https://doi.org/10.1037/dhe0000267

Stone, S., Saucer, C., Bailey, M., Garba, R., Hurst, A., Jackson, S. M., Krueger, N., & Cokley, K. (2018). Learning while Black: A culturally informed model of the impostor phenomenon for Black graduate students. *Journal of Black Psychology, 44(6),* 491–531. https://doi.org/10.1177/0095798418786648

SWPU Factbook (2020) [Redacted for Blind Review].

Templeton, E., Love, B. H., & Johnson, O. (2021). Elevating marginalized voices in academe: *Lessons for a new generation of scholars.* Routledge.

Tram, J. M., Nwankwo, N., Khan, A. N., & Sabado, J. A. (2020). Impact of faculty mentoring on ethnic and racial minority student program satisfaction. *Scholarship of Teaching and Learning in Psychology*, https://doi.org/10.1037/stl0000231

Vargas, N., & Villa-Palomino, J. (2018). Racing to serve or race-ing for money? Hispanic-serving institutions and the colorblind allocation of racialized federal funding. *Sociology of Race and Ethnicity, 5(3),* 401-415. https://doi.org/10.1177/2332649218769409.

Velásquez, R., Arellano, L. M., & McNeill, B. (2004). *The handbook of Chicana/o psychology and mental health*. Lawrence Erlbaum.

Wang, C. D., Scalise, D. A., Barajas, M. I., Julio, K., & Gomez, A. (2016). Attachment, acculturation, and psychosomatic complaints among Hispanic American university students. *Journal of College Counseling, 19(1),*45-60. https://doi.org/10.1002/jocc.12030.

White House Hispanic Prosperity Initiative. (n.d.). *Hispanic Serving Institutions (HSIs).* https://sites.ed.gov/hispanic-initiative/hispanic-serving-institutions-hsis/.

Wijesingha, R., & Ramos, H. (2017). Human capital or cultural taxation: What accounts for differences in tenure and promotion of racialized and female faculty? *Canadian Journal of Higher Education, 47(3),* 54–75. https://doi.org/10.47678/cjhe.v47i3.187902

Appendix A

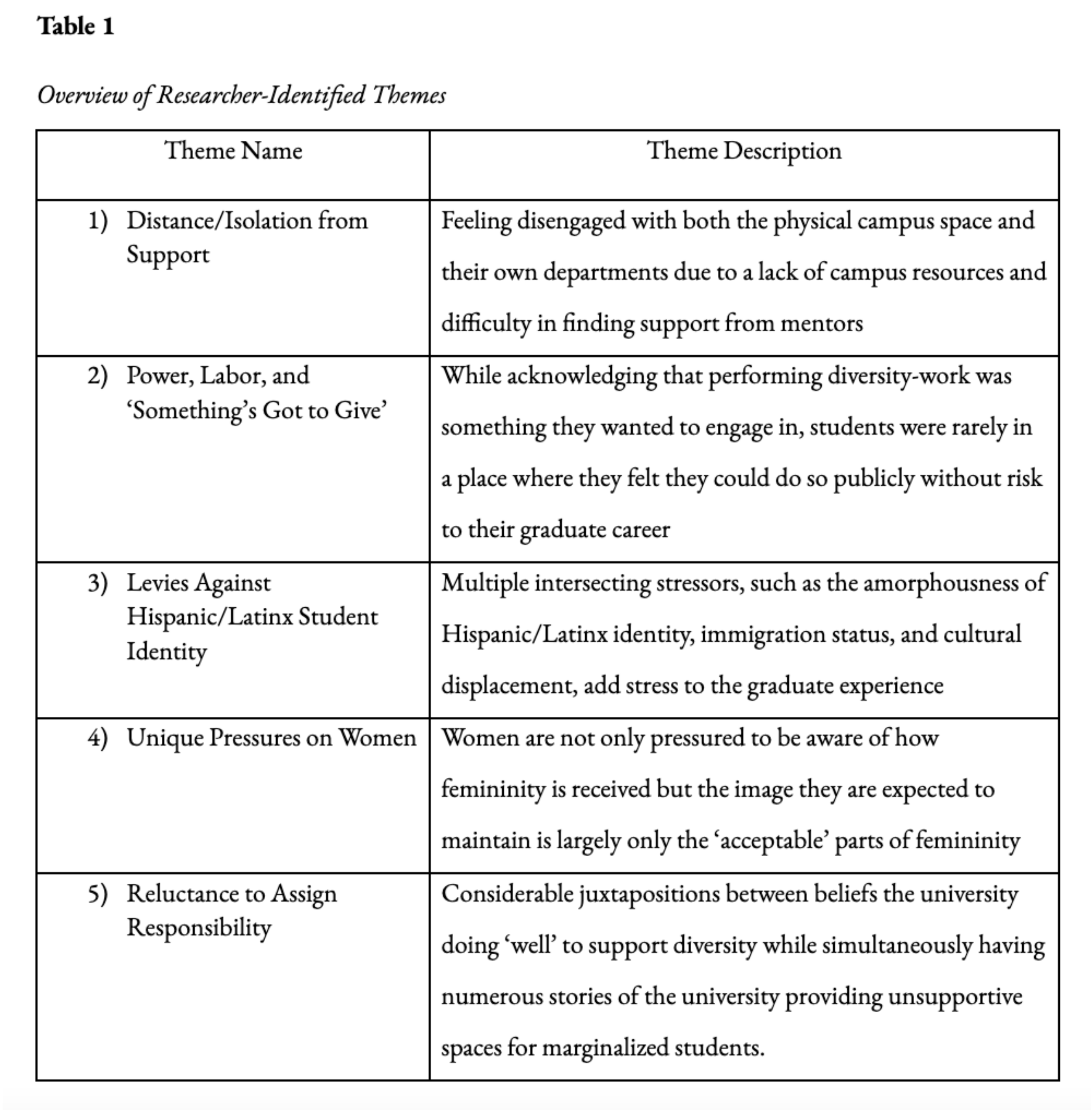


Table 1

Overview of Researcher-Identified Themes

Theme Name	Theme Description
1) Distance/Isolation from Support	Feeling disengaged with both the physical campus space and their own departments due to a lack of campus resources and difficulty in finding support from mentors
2) Power, Labor, and 'Something's Got to Give'	While acknowledging that performing diversity-work was something they wanted to engage in, students were rarely in a place where they felt they could do so publicly without risk to their graduate career
3) Levies Against Hispanic/Latinx Student Identity	Multiple intersecting stressors, such as the amorphousness of Hispanic/Latinx identity, immigration status, and cultural displacement, add stress to the graduate experience
4) Unique Pressures on Women	Women are not only pressured to be aware of how femininity is received but the image they are expected to maintain is largely only the 'acceptable' parts of femininity
5) Reluctance to Assign Responsibility	Considerable juxtapositions between beliefs the university doing 'well' to support diversity while simultaneously having numerous stories of the university providing unsupportive spaces for marginalized students.

Appendix B

Interview Questions for All Participants

1. What is your ideal version of campus diversity?
2. How does that ideal compare with current campus stances and efforts on diversity?
3. [university] is recognized as both an HSI and an MSI. In what ways, if any, does [university] serve students of color?
4. To your knowledge, has there been any pressure from students for [university] to support campus diversity efforts?
5. Have you ever done any work on behalf of campus diversity efforts? (If so) Can you tell me about that?
6. How did you come by that service? Did you volunteer, were you assigned, did someone request the service from you, etc.? (If they did not volunteer) Who requested or assigned that service from you?
7. In what ways, if any, was your service rewarded, validated, or noticed?
8. Were you satisfied with that validation? Why or why not?
9. Do you believe your White colleagues were met with the same validation?
10. Do you think anything was changed as a result of your labor?
11. In what ways is your service connected to your racial/ethnic identity?
12. What, if anything, has been the costs to you for your service? Have they been material, social, professional, time-based, etc.?
13. Would you say these costs in any way were specifically related to your status as a Latinx student and not more broadly a student of color?
14. In what ways, if any, do you feel like a representative of your ethnic/racial group on campus?
15. In what ways, if any, are you treated as an expert on issues of diversity?
16. In what ways, if any, do you believe your power compares with that of your White peers?
17. In what ways, if any, has your own area of expertise ever been downplayed by those with more power than you in favor of work on race or diversity?
18. Have you experienced discrimination or prejudice on campus as a result of your identity?
19. (If yes to previous question) Were you ever made to feel responsible for these experiences with discrimination?
20. Do you fear retaliation if/when you speak out against issues of racism?
21. Have you adapted any coping strategies to deal with anything we've discussed so far (such as experiences of discrimination or diversity/equity work)?
22. Do you feel the department and faculty keep up with your progress and milestones?
23. Do you feel like you receive adequate mentorship/advising that improves your work?
24. Do you feel like your work is ever undermined because of diversity-centric responsibilities?
25. Do you have a faculty member you feel like you could go to in order to discuss/address your elevated workload or experience of discrimination?

Additional Questions for Woman-Identifying Participants

1. Have you ever felt like you have to take on additional work because of your identity as a woman?
2. Do you feel as if expectations in your department are identical for men and women?
3. Do you feel like your male colleagues are ever threatened by your accomplishments?
4. Do you ever feel the need to adjust your expressions of womanhood or femininity within the context of your department?

Graduate Student Journal of Psychology
2022, Vol. 19

Parents' Psychological Adaptation after Receiving a Fetal Diagnosis: A Systematic Review

Catherine R. Gaspar,
Teachers College, Columbia University

A fetal diagnosis places an unexpected psychological burden on parents and triggers a complex pregnancy experience. Parents who choose to continue the pregnancy have unique perspectives as they prepare for birth. It is crucial to understand these families' experiences to inform their care and support. This qualitative systematic review explored the psychological adjustment of parents who continued gestation after they received a fetal diagnosis. A systematic database search was conducted with subsequent thematic analysis of fourteen included studies. Parents experienced intense initial emotional reactions to the diagnosis including shock and grief, followed by a complex processing period influenced by personal and social factors. The findings demonstrate a need for improved multidisciplinary parental support for families who receive a fetal diagnosis and add rationale for the addition of psychological services to the care teams of prospective parents.

Keywords: prenatal, fetal diagnosis, pregnancy, parents, emotions, process

Pregnancy is a time of major life adjustment for any parent. Parents who undergo typical pregnancies experience changes in lifestyles, emotions, and identities (Edvardsson et al., 2011). Pregnancies that receive a fetal diagnosis place an additional psychological burden on prospective parents, especially at the time of initial identification (van der Steen et al., 2016). These stressors continue after birth as the family shifts into living daily life with the infant and their specific developmental needs (Woolf-King et al., 2017).

Due to advances in technologies, such as noninvasive methods, prenatal screenings have become standard in much of the developed world (Pös et al., 2019; WHO, 2012). These screenings are generally accepted and perceived as necessary by parents (Aune & Möller, 2010, Ekelin et al., 2016), and receiving positive, on-track information about their unborn child's development can contribute to a positive pregnancy experience (Richter et al., 2020; Wittman et al., 2016). However, the widespread use of modern screening technologies also means greater detection of prenatal conditions such as birth defects or genetic disorders (Carlson & Vora, 2017). Reports show that about one in 33 births is complicated by a birth defect (CDC, 2008) which can often result in physical or mental disabilities (Boyle & Cordero, 2005).

Due to advances in technologies, such as noninvasive methods, prenatal screenings have become standard in much of the developed world (Pös et al., 2019; WHO, 2012). These screenings are generally accepted and perceived as necessary by parents (Aune & Möller, 2010, Ekelin et al., 2016), and receiving positive, on-track information about their unborn child's development can contribute to a positive pregnancy experience (Richter et al., 2020; Wittman et al., 2016). However, the widespread use of modern screening technologies also means greater detection of prenatal conditions such as birth defects or genetic disorders (Carlson & Vora, 2017). Reports show that about one in 33 births is complicated by a birth defect (CDC, 2008) which can often result in physical or mental disabilities (Boyle & Cordero, 2005).

A fetal diagnosis is difficult news for families to receive. While many families decide to terminate these pregnancies (Hawkins et al., 2012), some families decide to continue. In recent years, rates of continued pregnancy after receiving a fetal diagnosis have increased (Madeuf et al., 2016). The decision to continue a pregnancy is multifaceted. For many parents it is an ethical dilemma, with worries over playing God and wanting the pregnancy to occur naturally, or they feel a sentimental attachment to the fetus (Winn et al., 2018). The timing of diagnosis also matters. Parents farther along in gestation have a greater likelihood of choosing to continue the pregnancy (Madeuf et al., 2016; Michalik & Preis, 2013; Winn et al., 2018). Additionally, diagnosis severity and other variables play a role, where less severe fetal diagnoses and conditions with a history of greater postnatal success have a higher likelihood of pregnancy continuation (Hawkins et al., 2012; Madeuf et al., 2016; Winn et al., 2018). Other socio-contextual factors such as parental education, race, geographic location, and finances also impact the decision (Hawkins et al., 2012; Michalik & Preis, 2013).

For parents, the decision to continue pregnancy rather than terminate may improve their psychological outcomes (Cope et al., 2015), a crucial buffer as prospective parents already face vulnerabilities to their mental health during pregnancy (Cindy-Lee et al., 2017; Condon et al., 2004; Davalos et al., 2012). In typical pregnancies, factors such as lower socioeconomic status, financial hardships, younger maternal age, and histories of prior mental health struggles, put women at an increased risk for mental health conditions such as depression (Rich-Edwards et al., 2006). Lower social support is also a risk factor; parents with low social support may lack social networks which can provide advice, information, and reduce negative emotions that may result from stressors (Bedaso et al., 2021).

However, many factors can act protectively for prospective parents' psychological health, such as high social support which may promote psychological well-being (Corno et al., 2022). A qualitative study on public health clinics found that pregnant mothers utilized friendships to manage stress and relieve tension through venting and laughing together (Abdou et al., 2010). Similarly, relationship quality and partner support act as protective factors. Involved partners and strong relationships have been found to benefit maternal well-being for mothers undergoing typical pregnancies (Rini & Dunkel Schetter, 2010). Other literature on typical pregnancies has demonstrated that social support, family functioning, and relationship satisfaction buffered links between stressful events and depression (Divney et al., 2012). In recent work on specific stressful life events, such as the COVID-19 pandemic, support from partners, social networks, and healthcare providers was identified as playing protective roles in the mental health of pregnant women (Khoury et al., 2021; Vacaru et al., 2021). Further support can stem from parents' mindsets. Pregnant women who engaged in emotion-focused coping in early and middle pregnancy had lower distress (Huizink et al., 2002), and those with more positive cognitive appraisal in stressful life events like COVID-19 had less mental health problems (Khoury et al., 2021). Utilization of belief systems like religion or spirituality as a source of guidance or comfort has also been suggested as helpful factors (Abdou et al., 2010).

Further support of mental health vulnerabilities in pregnancy may be the receipt of psychological services (Kinser et al., 2021; Urizar et al., 2019), yet currently, many prospective parents undergoing typical pregnancies struggle to access adequate psychological services (Schwartz et al., 2021). Pilot intervention programs have been implemented to facilitate access to services (Slade et al., 2021), yet less is known about universal or broader efforts. Supporting the spread of information, the internet has helped aid parents to access informational resources (Fleming et al., 2014) Recent years have seen the development and utilization of e-mental health tools, where parents use web-based strategies for the delivery or enhancement of mental health information and services (Fonseca et al., 2016). For high-risk pregnancies, including those with fetal diagnoses, formal social supports (e.g., targeted support networks within peers or practitioner-facilitated groups) are also documented as effective resources (Coffman & Ray, 2002; Kugler & Farmer, 2015).

It is important to note that culture is an important factor when considering the emotional well-being, coping, and resources of parents (Cindy-Lee et al., 2017; Dunkel Schetter, 2011). The availability of resources and reasons for seeking support varies across cultures and geographic regions (Baron et al., 2015; Dunkel Schetter, 2011; Tol et al., 2018), yet further knowledge is needed on parent emotional processing and use of supports across cultures. Similarly, much of the work on the emotional well-being, emotional processing, and subsequent resources for prospective parents has been performed with typical pregnancies, however, less is known on these topics for those with high-risk pregnancies such as fetal diagnoses. These parents may have unique needs and experiences (van der Steen et al., 2016), thus further exploration of their emotional process and the factors which influence this process is needed.

A prenatal diagnosis vastly shifts parents' perspectives of pregnancy (Horsch et al., 2013). Other reviews have explored this phenomenon, but none have focused solely on the emotional processing and acclimation of parents who decided to continue their pregnancy. Lou and colleagues (2017) completed a thorough review of parent responses to prenatal diagnosis that included studies with both continued and terminated pregnancies. Johnson and colleagues (2020) performed a comprehensive review of prospective parents' views when a fetal abnormality was identified but focused solely on anomalies detected via ultrasound and included insights from healthcare professionals.

Prior reviews offer important insights on parents and prenatal diagnosis, but additional work is needed on the psychological state and external influences for parents who continue these pregnancies to better understand parents of infants with atypical development. As such, the present systematic review aimed to synthesize the prenatal emotional adjustment of parents who continued pregnancy after receiving a fetal diagnosis and to identify factors influencing their emotional responses to the diagnosis.

Methods

Search Procedures

This review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. The search strategy utilized SPIDER, a search tool for qualitative research (Cooke et al., 2012). SPIDER identifies the (S) sample studied ("parents, prospective parents"), (PI) phenomenon of interest ("prenatal diagnosis, prospective pregnancy"), (D) targeted study design ("interview, survey"), (E) evaluation of the phenomenon ("experiences, perceptions"), and (R) type of research ("qualitative").

Inclusion Criteria

Included studies were peer-reviewed original empirical works from any country that focused on the experiences of current or prospective parents (e.g., mother, father, familial caregivers) who continued their pregnancy after receiving a fetal diagnosis. Included studies focused on the prenatal period during and after the diagnosis, but prior to birth. If studies included both pre- and post-natal diagnoses, only prenatal data was used. Fetal diagnosis was conceptualized as an abnormality which would impact the child's post-birth functioning, development, or health. Parents were conceptualized as any primary caregiver of the child in the perinatal period. Studies needed to utilize parent-reported qualitative data (e.g., solely qualitative articles or qualitative sections of articles using mixed methods) and be written in English.

Exclusion Criteria

Studies were excluded if they used solely quantitative methods or were not published in a peer-reviewed journal. Studies with quantitative data were excluded from the present review as the focus was to provide an in-depth synthesis of parent experiences and emotions, a level of depth often better captured by qualitative data as it highlights lived experiences of participants while quantitative approaches aim to quantify and order participant data (Ponterotto, 2002). Studies with samples that included only stakeholders and no primary caregivers were excluded. Studies that included parent perspectives of terminated or miscarried pregnancies, or postnatally administered diagnoses, were also excluded. Studies were excluded that only focused on parents' postnatal experiences, or only detailed parents' prenatal experiences prior to an official diagnosis.

Study Selection

Study selection is summarized in Table 1. The selection process consisted of two stages. First, databases were searched using search terms and screened based on titles and abstracts. Next, full texts of eligible studies were read and eligibility criteria were applied, resulting in the inclusion of ten publications. A manual search of reference lists and library resources identified four more eligible papers, resulting in the inclusion of 14 total publications.

Quality Appraisal

Methodological quality of included studies was assessed by the author using the Critical Appraisal Skills Programme (CASP) qualitative appraisal tool (Critical Appraisal Skills Programme, 2018). The CASP tool evaluates based on criteria from three sections consisting of a total of ten items: A) Are the results of the study valid? (e.g., items 1) Was there a clear statement of the aims of the research? 2) Is a qualitative methodology appropriate? 3) Was the research design appropriate to address the aims of the research? 4) Was the recruitment strategy appropriate to the aims of the research? 5) Was the data collected in a way that addressed the research issue? 6) Has the relationship between the researcher and participants been adequately considered?), B) What are the results? (e.g., items 7) Have ethical issues been taken into consideration? 8) Was the data analysis sufficiently rigorous? 9) Is there a clear statement of findings?), and C) Will the results help locally? (e.g., item 10) How valuable is the research?). Each item was rated Yes, Can't Tell, or No. After rating, each item was assigned a point value (i.e., Yes = 2, Can't Tell = 1, No = 0) and items were totaled with 20 as the maximum possible score. Studies receiving a score of 17 or higher were classified as high methodological quality, scores between 16 and 14 as moderate methodological quality, and 13 or below as lesser methodological quality. No stud ies were excluded because of the appraisal (Table 1).

Data Synthesis

This review utilized thematic analysis (Thomas & Harden, 2008) which allowed results of the included publications to be synthesized. Specifically, the analysis procedures employed a thematic synthesis approach outlined by Thomas and Harden (2008) which allows for an effective translation of concepts and connection of qualitative research findings. The thematic synthesis guidelines were used to generate and identify themes and subthemes. First, included studies were read multiple times and notes were taken of initial ideas for coding. Next, the qualitative results sections of each study were reviewed and coded into a set of initial, broad codes. Data relevant to each code was extracted and sorted under the respective code. Codes were inductively developed and added to as needed, resulting in a set of overarching themes (e.g., initial reactions, processing period, social factors, and coping strategies) which were checked for fit with the data then defined and named. From here, data under each theme was reviewed and sorted further into narrower subthemes under each broader theme. Subthemes were reviewed and refined for fit within the broader themes.

Results

The 14 included studies were conducted in a range of countries: five from the United States, three from Australia, two from Sweden, one from Denmark, one from Ireland, one from South Korea, and one from the United Kingdom. The studies included a total of 251 primary caregivers, including 173 mothers, 67 fathers, nine grandparents, and two undisclosed sexes. While included studies varied in aims, qualitative design, and distinct focus, all examined parents' insight and experiences after receiving a prenatal diagnosis.

Based on the thematic analysis (Thomas & Harden, 2008), the synthesized results show that emotional adjustment to a fetal diagnosis was an ongoing process characterized by two main timepoints: initial diagnosis and processing period. Table 2 details a summary of themes and corresponding studies. Parents experienced different perspectives depending on the timepoint. They reported early emotions at the time of diagnosis, then shifted emotions as they underwent the processing period. While families consistently reacted strongly to the initial diagnosis, these emotions were not homogenous and shifted as parents adjusted to the news. Individual parent experiences of the overall adjustment process were influenced by two main factors: interactions with others and coping strategies. Figure 2 models parents' emotional process.

Initial Reactions. Parents reported a range of emotions when they first received the fetal diagnosis. Most parents described intense shock when they received the news (Carlsson et al., 2017; Clark et al., 2020; Côté-Arsenault & Denney-Koelsch, 2011; Goff et al., 2013; Hickerton et al., 2011; How et al., 2019; Im et al., 2018; Johnson et al., 2018; Lokmic et al., 2017; McKechnie & Pridham, 2012; O'Connell et al., 2019) and described how this shock made it difficult to grasp the diagnosis (Carlsson et al., 2017). Shock was especially salient in first-time mothers (O'Connell et al., 2019). Parents also reported grief, sadness, and mourning (Carlsson et al., 2015; Carlsson et al., 2017; Clark et al., 2020; Côté-Arsenault & Denney-Koelsch, 2011; Hickerton et al., 2011; How et al., 2019; Im et al., 2018; Johnson et al., 2018; Lou et al., 2020; McKechnie & Pridham, 2012; McKechnie et al., 2015; O'Connell et al., 2019). They described intense emotional suffering and devastation (Carlsson et al., 2015; Clark et al., 2020; Im et al., 2018; McKechnie & Pridham, 2012; O'Connell et al., 2019), and felt a deep sense of loss surrounding their original expectations for the child and pregnancy (Côté-Arsenault & Denney-Koelsch, 2011; Hickerton et al., 2011; Johnson et al., 2018; Lou et al., 2020; McKechnie et al., 2015).

A portion of parents felt anger when they first heard the diagnosis (Carlsson et al., 2017; Goff et al., 2013) while others felt confused (Carlsson et al., 2015; Hickerton et al., 2011; Im et al., 2018; McKechnie & Pridham, 2015). They questioned why (Goff et al., 2013; McKechnie & Pridham, 2012) and felt the diagnosis did not make sense (Carlsson et al., 2015; Im et al., 2018). Many parents also reported fear and anxiety (Carlsson et al., 2015; Im et al., 2018; Johnson et al., 2018; Lokmic et al., 2017; Lou et al., 2020; McKechnie et al., 2015). Some were fearful of fetal loss or worsening of the condition (Carlsson et al., 2015; McKechnie et al., 2015), while others experienced panic about the future (Im et al., 2018; Lou et al., 2020). Parents also reported feelings of guilt (Carlsson et al., 2017; Clark et al., 2020; Côté-Arsenault & Denney-Koelsch, 2011; Hickerton et al., 2011; Im et al., 2018; Lokmic et al., 2017). Many felt guilt for potentially causing the anomaly (Carlsson et al., 2017; Côté-Arsenault & Denney-Koelsch, 2011; Lokmic et al., 2017), while others felt moral guilt in considering whether to terminate the pregnancy (Im et al., 2018; Hickerton et al., 2011). For some parents, guilt was compounded with other emotions, like guilt about their sadness or hope for a false-positive diagnosis (Carlsson et al., 2017; Clark et al., 2020; Côté-Arsenault & Denney-Koelsch, 2011).

Overall Trends during the Processing Period. Adjusting to the diagnosis was a process for parents. They overwhelmingly reported gratitude in receiving the diagnosis prenatally versus postnatally (Carlsson et al., 2015; Carlsson et al., 2017; Clark et al., 2020; Côté-Arsenault & Denney-Koelsch, 2011; Goff et al., 2013; Hickerton et al., 2011; How et al., 2019; Im et al., 2018; Johnson et al., 2018; Lokmic et al., 2017; Lou et al., 2020; McKechnie & Pridham, 2012; McKechnie et al., 2015; O'Connell et al., 2019) as this gave them time to prepare. However, after receiving the diagnosis, parents felt they needed time to digest the news (Clark et al., 2020; Goff et al., 2013; How et al., 2019; McKechnie & Pridham, 2012). In this processing period, they shifted their thoughts and feelings toward the future (Carlsson et al., 2015; Carlsson et al., 2017; Clark et al., 2020; Goff et al., 2013; Hickerton et al., 2011; How et al., 2019; Im et al., 2018; Johnson et al., 2018; Lou et al., 2020; McKechnie & Pridham, 2012; McKechnie et al., 2015). As they looked ahead, most described an acceptance of the diagnosis (Carlsson et al., 2015; Carlsson et al., 2017; Clark et al., 2020; Goff et al., 2013; Hickerton et al., 2011; How et al., 2019; Im et al., 2018; Lou et al., 2020; McKechnie & Pridham, 2012; McKechnie et al., 2015). Many felt having time to process emotions helped their acceptance (Clark et al., 2020; How et al., 2019; Goff et al., 2013; McKechnie & Pridham, 2012), and used this time to reframe their original expectations of the child, future parenting, and life milestones (Hickerton et al., 2011; McKechnie et al., 2015). Parents began to celebrate and see their unborn baby as any other child, with individuality, personality, and hardships that all parents face (Goff et al., 2013; How et al., 2019; Im et al., 2018; Lou et al., 2020; McKechnie et al., 2015).

Despite acceptance of their future child and circumstances, most parents reported ongoing anxieties during the processing period (Carlsson et al., 2015; Carlsson et al., 2017; Clark et al., 2020; Goff et al., 2013; Hickerton et al., 2011; How et al., 2019; Im et al., 2018; Johnson et al., 2018; Lou et al., 2020; McKechnie & Pridham, 2012; McKechnie et al., 2015; O'Connell et al., 2019). Worries focused on the current pregnancy and the remaining gestational development of their child (Carlsson et al., 2017; How et al., 2019; Lou et al., 2020; McKechnie & Pridham, 2012; McKechnie et al., 2015). Other fears revolved around the future. Parents were uncertain about the outlook for themselves and their child (Carlsson et al., 2015; Carlsson et al., 2017; Clark et al., 2020; Goff et al., Hickerton et al., 2011; 2013; Johnson et al., 2018; Lou et al., 2020; McKechnie & Pridham, 2012; McKechnie et al., 2015). They worried about navigating the new influx of information and meeting their child's future needs, such as medical and other support services (Carlsson et al., 2017; Clark et al., 2020; McKechnie & Pridham, 2012; McKechnie et al., 2015). Many felt societal pressures such as the expectation to produce a healthy child (Im et al., 2018), the impact of a child with a disability on family functioning (Carlsson et al., 2015; Carlsson et al., 2017; Goff et al., 2013; Lou et al., 2020), and their child's future social experiences including stigma and social milestones (Clark et al., 2020; Johnson et al., 2018; Lou et al., 2020; McKechnie et al., 2015). Parents also reported uncertainty in their social functioning, especially when interacting with parents of typically developed children (Goff et al., 2013; Hickerton et al., 2011; Johnson et al., 2018).

Though anxious, parents reported a newfound bond with their unborn child (Carlsson et al., 2017; Côté-Arsenault & Denney-Koelsch, 2011; Im et al., 2018; Lou et al., 2020; McKechnie et al., 2015; O'Connell et al., 2019). As they accepted their child, they described a strong sense of love and attachment and felt an increasing connection that strengthened their parental duty and commitment (Carlsson et al., 2017; Im et al., 2018; Lou et al., 2020; McKechnie et al., 2015; O'Connell et al., 2019). They saw their baby as an individual, and desired for others to view their child the same way (Côté-Arsenault & Denney-Koelsch, 2011; How et al., 2019; Im et al., 2018; O'Connell et al., 2018). Filled with acceptance and love, parents reported feelings of positivity and hope as they progressed through the pregnancy (Clark et al., 2020; Im et al., 2018; McKechnie & Pridham, 2012; McKechnie et al., 2015).

Influencing factors. Despite consistent overall trends, parents did not all follow the same rate of adjustment during the processing period. Some had a slower adjustment and others rebounded from the diagnosis quickly. Some had a positive, fulfilling pregnancy, while others felt more cynical and distanced. Emergent themes evidenced that experiences were shaped by outside factors. Throughout included studies, parents mentioned multifaceted details in social interactions and coping strategies which contributed to their pregnancy experience and acclimation process.

***Social Interactions with Other*s**

Medical Professionals. Medical professionals ranged from doctors, nurses, doulas, and other pregnancy support staff. Some parents developed close relationships with their healthcare professionals and communicated with them frequently after receiving the diagnosis (McKechnie & Pridham, 2012; McKechnie et al., 2015). These parents had a select few which they trusted and looked to for support (Carlsson et al., 2015; Goff et al., 2013; Hickerton et al., 2011; Lou et al., 2020). They appreciated their knowledge and advice relating to the pregnancy and diagnosis (Carlsson et al., 2015; Côté-Arsenault & Denney-Koelsch, 2011, Hickerton et al., 2011; How et al., 2019; Im et al., 2018; McKechnie et al., 2015; O'Connell et al., 2019). Parents were especially satisfied when medical professionals listened and supported their decisions and felt most at ease getting continuous care from their trusted team (Carlsson et al., 2015; Hickerton et al., Lou et al., 2020).

However, many parents in included studies overwhelmingly reported negative experiences with medical professionals after the diagnosis. They reported that healthcare professionals had pessimistic attitudes and delivered diagnostic news poorly (Carlsson et al., 2015; Clark et al., 2020; Côté-Arsenault & Denney-Koelsch, 2011; Goff et al., 2013; Hickerton et al., 2011; Im et al., 2018; Johnson et al., 2018; Lokmic et al., 2017; Lou et al., 2020; McKechnie et al., 2015; O'Connell et al., 2019). Parents felt providers held negative stereotypes about disabilities and routinely pushed for a termination of pregnancy, often making repeated comments about termination (Goff et al., 2013; Hickerton et al., 2011; How et al., 2018; Johnson et al., 2018; Lokmic et al., 2017; Lou et al., 2020). Many parents described medical professionals as grim, unemotional, and uncompassionate in their care (Clark et al., 2020; Goff et al., 2013; Lou et al., 2020), and recounted hurtful, inappropriate remarks from providers about their babies and choices (Johnson et al., 2018; McKechnie et al., 2015; O'Connell et al., 2019). Parents felt the professionals lacked crucial knowledge and resources about the diagnoses and were inconsistent in the advice and information they shared (Carlsson et al., 2015; Carlsson et al., 2017; Côté-Arsenault & Denney-Koelsch, 2011; Goff et al., 2013; Johnson et al., 2018; Lokmic et al., 2017; Lou et al., 2020). Often, unannounced specialists attended appointments, which made parents increasingly anxious and distrustful (Johnson et al., 2018; Lokmic et al., 2017).

Friends and Family. Family and friends ranged from immediate and extended family to circles of friends. Parents' social networks were important sources of support during the pregnancy (Carlsson et al., 2017; Clark et al., 2020; Côté-Arsenault & Denney-Koelsch, 2011; Goff et al., 2013; Hickerton et al., 2011; McKechnie & Pridham, 2012; McKechnie et al., 2015; O'Connell et al., 2019). Parents retreated to their trusted, inner social circles after the diagnosis (Clark et al., 2020; McKechnie & Pridham, 2012; O'Connell et al., 2019), and relied on them to listen and help with practical needs (Carlsson et al., 2017; McKechnie & Pridham, 2012).

However, not all of the parents' social groups were accepting or helpful (Carlsson et al., 2017; Clark et al., 2020; Côté-Arsenault & Denney-Koelsch, 2011; Hickerton et al., 2011; How et al., 2019; McKechnie et al., 2015; O'Connell et al., 2019). Parents felt family and friends were not always supportive (Côté-Arsenault & Denney-Koelsch, 2011; Hickerton et al., 2011; McKechnie et al., 2015). Parents received insensitive remarks and felt pressured by others to terminate the pregnancy (Hickerton et al., 2011; How et al., 2019; O'Connell et al., 2019). Many friends and family members were unsure how to respond to the diagnosis, while others refused to accept the pregnancy or unborn baby (Clark et al., 2020; Côté-Arsenault & Denney-Koelsch, 2011; Hickerton et al., 2011). Parents felt lonely, isolated, and stigmatized, and felt others could no longer relate to them (Côté-Arsenault & Denney-Koelsch, 2011; McKechnie et al., 2015, O'Connell et al., 2019). They underwent a great divide in their personal and social experiences and reported the distancing and changing of friendships (Carlsson et al., 2017; Côté-Arsenault & Denney-Koelsch, 2011; McKechnie et al., 2015).

Parents' Coping Strategies

Research and Information-Gathering. Most parents felt a need to find information on the diagnosis, which they did through personal research. They tried to educate themselves and build their knowledge as much as possible so they could be prepared for the future (Carlsson et al., 2015; Carlsson et al., 2017; Clark et al., 2020; Im et al., 2018; Johnson et al., 2018; Lokmic et al., 2017; Lou et al., 2020; McKechnie & Pridham, 2012; McKechnie et al., 2015; O'Connell et al., 2019). Some found this strategy helpful and attained valuable information on their baby's condition (Carlsson et al., 2015; Im et al., 2018; Lou et al., 2020; McKechnie & Pridham, 2012), while others found it overwhelming and confusing due to the large quantity of resources, much of it negative or outdated (Carlsson et al., 2015; Clark et al., 2020; Lokmic et al., 2017; McKechnie & Pridham, 2012; McKechnie et al., 2015).

Experiential Knowledge. Parents also relied on the experiences of other families with the same diagnoses. Some parents connected with these families through their research efforts, while others reached out to families they already knew. For some, the experiences of others caused fear and distress, especially in cases with poor outcomes (Carlsson et al., 2015; Carlsson et al., 2017; McKechnie et al., 2015). But for many, the experiences of other parents helped them feel positive and reassured (Clark et al., 2020; Johnson et al., 2018; Lokmic et al., 2017; McKechnie & Pridham, 2012; McKechnie et al., 2015; O'Connell et al., 2019). Collaborating with others lessened their anxiety and reshaped expectations, especially when they saw children with similar conditions living happy lives (How et al., 2019; Johnson et al., 2018). Some parents also joined parent groups and disability organizations to extend their social support (Johnson et al., 2018; McKechnie & Pridham, 2012; McKechnie et al., 2015).

Healthcare Planning. Parents also took comfort in planning for the future. They focused on the logistical health consequences of the diagnosis, joined healthcare waitlists, and planned postnatal care so they could be actively involved in upcoming healthcare decisions (Carlsson et al., 2015; Johnson et al., 2018; McKechnie & Pridham, 2012; McKechnie et al., 2015). Others reflected on their personal health and began eating better and exercising (Im et al., 2018).

Perspective-Taking. As another strategy, parents reexamined their perspectives of the diagnosis. Some utilized religion and began to view their baby as a blessing from a higher power (Goff et al., 2013; Im et al., 2013), while others actively worked to change their idealized future to better align with the diagnosis (Clark et al., 2020; How et al., 2019; McKechnie et al., 2015). These parents emphasized gaining rather than losing and focused on the essential responsibility of raising the child, identifying themselves as parents and embracing their parental role (How et al., 2019; Lou et al., 2020; McKechnie et al., 2015).

Discussion

This systematic review of 14 qualitative studies explored the psychological processes of prospective parents after receiving a fetal diagnosis. The review found that receiving a prenatal diagnosis marks a multiplex adjustment period for parents and adds an unexpected psychological burden as they absorb the news. At the initial diagnosis, parents often experienced mourning and shock. Past work suggests that these reactions are common for parents, and they are not alone in these feelings (Staham et al., 2000) as the initial diagnosis is the most emotionally challenging time for parents (Chaplin et al., 2005). Beyond the first diagnosis, parents in the present review progressed through a multifaceted adjustment process filled with dynamic, complicated emotions. Consistent with prior work, parents in the present review differed in processing time as they accepted the child as an individual (Chaplin et al., 2005; Lou et al., 2017) and oriented themselves to how the diagnosis shaped their present and future (Johnson et al., 2020; Statham et al., 2000). Outside factors including social support and coping strategies influenced parents' individual adjustment trajectories.

A mental shift is common for prospective parents. Past work has shown that the transition to parenthood lowers new parents' psychological well-being, even for those with typically develop pregnancies and high self-esteem prior to becoming a parent (Chen et al., 2020). However, parents who receive a fetal diagnosis are at an especially increased mental health risk (Cole et al., 2016) as they shift their mindsets and prepare for the future.

Given the emotional adjustment occurring in these pregnancies and the influence of external factors documented in the present review, parents may benefit from appropriate care and support during this period to boost positive influencing factors the present review. Strengthening protective factors in this population is especially imperative as parents may be at risk for long-term psychological distress as children born with a fetal abnormality have an increased risk of altered developmental outcomes and disabilities throughout their lifespan (Liu et al., 2016; Love et al., 2011) and parents of children with disabilities often show decreased psychological health (Olsson & Hwang, 2008).

One possibility for a reinforced protective support is through healthcare providers. The present review and past literature show that parents with fetal diagnoses often experience negative, insensitive, and uninformed medical care (Chaplin et al., 2005; Stock et al., 2019), but literature suggests that healthcare providers may be ill-prepared and receive little training in parent practices post-diagnosis (Johnson et al., 2020; Luz et al., 2017). Despite this, parents strongly rely on and desire genuine, trusting relationships with their healthcare team (Oulton et al., 2020), and parents in the present review who received supportive care from their healthcare team felt more comfortable and prepared. To address gaps in provider quality, health organizations may want to consider updating and expanding parent resources and encourage healthcare professionals to partake in additional professional development and training on parent support practices and fetal abnormalities to better support families in the perinatal period.

As evidenced in the present review and prior work (Coffman & Ray, 2002; Kugler & Farmer, 2015), targeted social networks also offer a promising route for a bolstered protective role. In addition to parents' immediate social circles, findings suggest advantages to participation in formal social outlets such as support groups, especially those with similar perinatal experiences. Parents in past literature who participated in parent groups described these relationships as imperative for their adjustment to parenthood as it provided a space to discuss their shared experiences and challenges with new parenting (Glavin et al., 2017). Parents who receive fetal diagnoses may benefit from such social opportunities to connect with other families with shared experiences.

A third, innovative approach to strengthening protective factors is the addition of psychological professionals to prospective parents' support teams. Literature suggests that parents are best supported by a multidisciplinary team of healthcare and psychological professionals (Catlin et al., 2008; Statham et al., 2000), and as such parents may benefit from psychological support as they process the diagnosis. Recent work on mindfulness interventions (Reid et al., 2016), group prenatal care (Ickovics et al., 2019), grief support (Navidian et al., 2017), and other psychological counseling (Rohde et al., 2008) indicate promising results for perinatal parent populations. Pretest counseling has also been demonstrated as beneficial to help emotionally prepare parents for prenatal screening results (Dorner et al., 2020).

Limitations and Conclusion

There are some limitations in this review. First, while efforts were made to conduct a thorough, comprehensive search and selection of literature, additional studies may have been published since the initial searches were performed and were not included in the present selection. Similarly, the analysis process may differ between researchers. Although procedural steps are detailed, it is possible that others may utilize different coding arrangements and judgements. Additionally, although qualitative studies were included in the present review, quantitative data may also provide valuable insights. Future work in this area may want to perform an analysis of quantitative findings or synthesize a combination of both qualitative and quantitative literature. Studies in this review covered a wide range of countries. It is important to note the influence of culture in each target population, which may shape study results including individual responses, coping, and perspectives. Future reviews may benefit from narrowing the focus to a more homogenous selection of countries, while further empirical work is needed to explore fetal diagnoses and available resources across cultures and geographic regions.

Receiving a fetal diagnosis is life-altering news for families. The present review suggests that prospective parents have similar initial reactions to prenatal diagnosis, but the subsequent adjustment process is shaped by various influencing factors and external characteristics. Parents who have strong protective factors, such as supportive social networks and sensitive, informed healthcare support may have improved experiences and adjustment. Findings suggest the benefit of providing improved medical, psychological, and social resources to parents after a fetal diagnosis. Taken together, these findings provide deeper understanding of parent experiences and related factors and indicate directions for future parent support.

References

Abdou, C. M., Schetter, C. D., Jones, F., Roubinov, D., Tsai, S., Jones, L., Lu, M., & Hobel, C. (2010). Community perspectives: mixed-methods investigation of culture, stress, resilience, and health. *Ethnicity & disease, 20(1 Suppl 2)*, S2–48.

Aune, I. & Möller A. (2010). "I want a choice, but I don't want to decide" - A qualitative study of pregnant women's experiences regarding early ultrasound risk assessment for chromosomal anomalies. *Midwifery, 28(1),* 14– 23. doi:10.1016/j.midw.2010.10.015

Baron, E., Field, S., Kafaar, Z., & Honikman, S. (2015). Patterns of use of a maternal mental health service in a low-resource antenatal setting in South Africa. *Health & Social Care in the Community, 23(5)*, 502–512. doi:10.1111/hsc.12167

Bedaso, A., Adams, J., Peng, W., & Sibbritt, D. (2021). The relationship between social support and mental health problems during pregnancy: A systematic review and meta-analysis. *Reproductive Health, 18,* 1-23. doi:10.1186/s12978-021-01209-5

Boyle, C. A., & Cordero, J. F. (2005). Birth defects and disabilities: a public health issue for the 21st century. *American Journal of Public Health, 95(11)*, 1884–1886. doi:10.2105/AJPH.2005.067181

Carlson, L. M., & Vora, N. L. (2017). Prenatal diagnosis: Screening and diagnostic tools. *Obstetrics and Gynecology Clinics of North America, 44(2)*, 245–256. doi:10.1016/j.ogc.2017.02.004

*Carlsson, T., Bergman, G., Marttala, U.M., Wadenstein, B., & Mattsson, E. (2015). Information following a diagnosis of congenital heart defect: Experiences among parents to prenatally diagnosed children. *PLos ONE, 10(2)*. doi:10.1371/journal.pone.0117995

*Carlsson, T., Starke, V., & Mattsson, E. (2017). The emotional process from diagnosis to birth following a prenatal diagnosis of fetal anomaly: A qualitative study of messages in online discussion boards. *Midwifery, 48,* 53-59. doi:10.1016/j.midw.2017.02.010

Catlin, A., Askeldottir, B., Conroy, S., & Rempel, G. (2008). From diagnosis to birth: Parents' experience when expecting a child with congenital anomaly. *Advances in Neonatal Care, 8(6)*, 348-354. doi:10.1097/01.ANC.0000342768.94734.23

Centers for Disease Control and Prevention. (2008). Update on overall prevalence of major birth defects—Atlanta, Georgia, 1978–2005. *MMWR Weekly Report, 57(1)*, 1–5. Retrieved October 19, 2020, from http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5701a2.htm

Chaplin, J., Schweitzer, R., & Perkoulidis, S. (2005). Experiences of prenatal diagnosis of spina bifida or hydrocephalus in parents who decide to continue with their pregnancy. *Journal of Genetic Counseling, 14(2),* 151-162. doi:10.1007/s10897-005-0488-9

Chen, E.Y.-J., Tung, E.Y.-L., & Enright, R.D. (2021). Pre-parenthood sense of self and the adjustment to the transition to parenthood. *Journal of Marriage and Family, 83,* 428-445. doi:10.1111/jomf.12709

Cindy-Lee, D., Falah-Hassani Kobra, & Rahman, S. (2017). Prevalence of antenatal and postnatal anxiety: Systematic review and meta-analysis. T*he British Journal of Psychiatry, 210(5)*, 315-323. doi:10.1192/bjp.bp.116.187179

*Clark, L., Canary, H. E., McDougle, K., Perkins, R., Tadesse, R., & Holton, A. E. (2020). Family sense-making after a Down Syndrome diagnosis. *Qualitative Health Research, 30(12),* 1783–1797. doi:10.1177/1049732320935836

Coffman, S., & Ray, M. A. (2002). African American women describe support processes during high-risk pregnancy and postpartum. *Journal of Obstetric, Gynecologic, and Neonatal Nursing, 31(5),* 536–544. doi:10.1111/j.1552-6909.2002.tb00078.x

Cole, J.C.M., Moldenhauer, J.S., Berger, K., Cary, M.S., Smith, H., Martino, V., Rendon, N., & Howell, L.J. (2016). Identifying expectant parents at risk for psychological distress in response to a confirmed fetal abnormality. *Archives of Women's Mental Health, 19(3),* 443–453. doi: 10.1007/s00737-015-0580-6

Condon, J. T., Boyce, P., & Corkindale, C. J. (2004). The First-Time Fathers Study: a prospective study of the mental health and wellbeing of men during the transition to parenthood. *The Australian and New Zealand Journal of Psychiatry, 38(1-2)*, 56–64. doi:10.1177/000486740403800102

Cooke, A., Smith, D., & Booth, A. (2012). Beyond

1PICO: The SPIDER tool for qualitative evidence synthesis. *Qualitative Health Research, 22(10)*,435-1443. doi: 10.1177/1049732312452938

Cope, H., Garrett, M. E., Gregory, S., & Ashley-Koch, A. (2015). Pregnancy continuation and organizational religious activity following prenatal diagnosis of a lethal fetal defect are associated with improved psychological outcome. *Prenatal Diagnosis, 35,* 761– 768. doi: 10.1002/pd.4603

Corno, G., Villani, D., de Montigny, F., Pierce, T., Bouchard, S., & Molgora, S. (2022). The role of perceived social support on pregnant women's mental health during the COVID-19 pandemic. *Journal of Reproductive and Infant Psychology, 1–15.* doi:10.1080/02646838.2022.2042799

*Côté-Arsenault, D., & Denney-Koelsch, E. (2011). "My baby is a person": Parents' experiences with life-threatening fetal diagnosis. *Journal of Palliative Medicine, 14(12).* 1302-1308. doi:10.1089/jpm.2011.0165

Critical Appraisal Skills Programme. (2018). *CASP Qualitative Checklist.* https://casp-uk.net/wp-content/uploads/2018/01/CASP-Qualitative-Checklist-2018.pdf

Davalos, D. B., Yadon, C. A., & Tregellas, H. C. (2012). Untreated prenatal maternal depression and the potential risks to offspring: a review. *Archives of Women's Mental Health, 15(1)*, 1–14. doi:10.1007/s00737-011-0251-1

Denney-Koelsch, E. M., Côté-Arsenault, D., & Jenkins Hall, W. (2018). Feeling cared for versus experiencing added burden: Parents' interactions with health-care providers in pregnancy with a lethal fetal diagnosis. Illness, *Crisis & Loss, 26(4),* 293–315. doi:10.1177/1054137316665817

Divney, A. A., Sipsma, H., Gordon, D., Niccolai, L., Magriples, U., & Kershaw, T. (2012). Depression during pregnancy among young couples: the effect of personal and partner experiences of stressors and the buffering effects of social relationships. *Journal of Pediatric and Adolescent Gynecology, 25(3)*, 201–207. doi:10.1016/j.jpag.2012.02.003

Dorner, R.A., Boss, R.D., Burton, V.J., Raja, K., & Lemmon, M.E. (2020). Parent preferences for neurodevelopmental screening in the neonatal intensive care unit. *Developmental Medicine and Child Neurology, 62,* 500-505. doi:10.1111/dmcn.14457

Dunkel Schetter, C. (2011). Psychological science on pregnancy: stress processes, biopsychosocial models, and emerging research issues. *Annual Review of Psychology, 62,* 531–558. doi:10.1146/annurev.psych.031809.130727

Edvardsson, K., Ivarsson, A., Eurenius, E., Garvare, R., Nyström, M.E., Small., & Mogren, I. (2011). Giving offspring a healthy start: Parents' experiences of health promotion and lifestyle change during pregnancy and early parenthood. *BMC Public Health, 11,* 936. doi:10.1186/1471-2458-11-936

Ekelin, M., Persson, L., Välimäki, A., & Crang Svalenius, E. (2016). To know or not to know – Parents' attitudes to and preferences for prenatal diagnosis. *Journal of Reproductive & Infant Psychology, 34(4),* 356–369. doi: 10.1080/02646838.2016.1200019

Fleming, S.E., Vandermause, R., & Shaw, M.R. (2014). First-time mothers preparing for birthing in an electronic world: Internet and mobile phone technology. *Journal of Reproductive and Infant Psychology, 32,* 240 - 253. doi: 10.1080/02646838.2014.886104

Fonseca, A., Gorayeb, R., & Canavarro, M. C. (2016). Women's use of online resources and acceptance of e-mental health tools during the perinatal period. International *Journal of Medical Informatics, 94,* 228–236. doi:10.1016/j.ijmedinf.2016.07.016

Glavin, K., Tveiten, S., Økland, T., & Hjälmhult, E. (2017). Maternity groups in the postpartum period at well child clinics - mothers' experiences. *Journal of Clinical Nursing, 26(19-20)*, 3079–3087. doi:10.1111/jocn.13654

*Goff, B.S.N., Springer, N., Foote, L., Frantz, C., Peak., M., Tracy, C., Veh, T., Bentley, G. E., & Cross, K.A. (2013). Receiving the initial Down Syndrome diagnosis: A comparison of prenatal and postnatal parent group experiences. J*ournal of Intellectual and Developmental Disabilities, 51(6),* 446-457. doi: 10.1352/1934-9556-51.6.446

Hawkins A., Stenzel A., Taylor J., Chock V., & Hudgins L. (2012). Variables influencing pregnancy termination following prenatal diagnosis of fetal chromosome abnormalities*. Journal of Genetic Counseling, 22(2)*, 238–48. doi:10.1007/s10897-012-9539-1

*Hickerton, C.L., Aitken, M., Hodgson, J., & Delatycki, M.B. (2011). "Did you find out in time?": New life trajectories of parents who choose to continue a pregnancy where a genetic disorder is diagnosed or likely. *American Journal of Medical Genetics, 158A(2),* 373-383. doi: 10.1002/ajmg.a.3439

Horsch, A., Brooks, C. and Fletcher, H. (2013). Maternal coping, appraisals and adjustment following diagnosis of fetal anomaly. P*renatal Diagnosis, 33(12),* 1137-1145. doi:10.1002/pd.4207

*How, B., Smidt, A., Wilson, N. J., Barton, R., & Valentin, C. (2019). 'We would have missed out so much had we terminated': What fathers of a child with Down syndrome think about current non-invasive prenatal testing for Down syndrome. *Journal of Intellectual Disabilities, 23(3),* 290–309. doi:10.1177/1744629518787606

Huizink, A. C., de Medina, P. G., Mulder, E. J., Visser, G. H., & Buitelaar, J. K. (2002). *Coping in normal pregnancy. Annals of behavioral Medicine, 24(2),* 132–140. doi:10.1207/S15324796ABM2402_10

Ickovics, J. R., Lewis, J. B., Cunningham, S. D., Thomas, J., & Magriples, U. (2019). Transforming prenatal care: Multidisciplinary team science improves a broad range of maternal–child outcomes. *American Psychologist, 74(3),* 343-355. doi:10.1037/amp0000435

*Im, Y.-M., Yun, T.-J., Yoo, I.-Y., Kim, S., Jin, J., & Kim, S. (2018). The pregnancy experience of Korean mothers with a prenatal fetal diagnosis of congenital heart disease. B*MC Pregnancy and Childbirth, 18(1),* 467. doi: 10.1186/s12884-018-2117-2

Johnson, J, Dunning, A, Sattar, R, Arezina, J., Karkowsky, E.C., Thomas, S., & Panagioti, M. (2020). Delivering unexpected news via obstetric ultrasound: A systematic review and meta ethnographic synthesis of expectant parent and staff experiences. *Sonography, 7(2),* 61– 77. doi:10.1002/sono.12213

*Johnson, J., Johnson, O., Heyhoe, J., Fielder, C., & Dunning, A. (2018). Parent experiences and preferences when dysmelia is identified during the prenatal and perinatal periods: A qualitative study into family nursing care for rare diseases. *Journal of Family Nursing, 24(2),* 271–293. doi:10.1177/1074840718772808

Khoury, J. E., Atkinson, L., Bennett, T., Jack, S. M., & Gonzalez, A. (2021). COVID-19 and mental health during pregnancy: The importance of cognitive appraisal and social support. *Journal of Affective Disorders, 282,* 1161–1169. doi:10.1016/j.jad.2021.01.027

Kinser, P., Thacker, L., Rider, A., Moyer, S., Amstadter, A., Mazzeo, S., Bodnar-Deren, S. & Starkweather, A. (2021). Feasibility, acceptability, and preliminary effects of "Mindful Moms". *Nursing Research, 70 (2),* 95-105. doi: 10.1097/NNR.0000000000000485.

Kugler, L., & Farmer, M-C. (2015). Initiation of group prenatal care in the center for fetal diagnosis and treatment. *Journal of Obstetric, Gynecologic & Neonatal Nursing, 44(1),* S20-S21. doi: 10.1111/1552-6909.12689

Liu, M. Y., Zielonka, B., Snarr, B. S., Zhang, X., Gaynor, J. W., & Rychik, J. (2018). Longitudinal assessment of outcome from prenatal diagnosis through fontan operation for over 500 fetuses with single ventricle-type congenital heart disease: The Philadelphia fetus-to-fontan cohort study. *Journal of the American Heart Association, 7(19).* doi:10.1161/JAHA.118.009145

Love, K., Huddleston, L., Olney, P., Wrubel, D., & Visootsak, J. (2011). Developmental outcomes of Down Syndrome and Dandy-Walker malformation. *Journal of Pediatric Neurology, 9(3),* 405-408. doi: 10.1159/000363179

*Lokmic, Z., Hallenstein, L., Penington, A.J. (2017). Parental experience of prenatal diagnosis of lymphatic malformation. *Lymphology, 50(1),* 16-26. pmid: 30231304

Lou, S., Jensen, L. G., Petersen, O. B., Vogel, I., Hvidman, L., Møller, A., and Nielsen, C. P. (2017). Parental response to severe or lethal prenatal diagnosis: A systematic review of qualitative studies. *Prenatal Diagnosis, 37(8),* 731– 743. doi: 10.1002/pd.5093.

*Lou, S., Panther, M.R, Hagestjerne, N., Peterson, O.B., & Vogel, I. (2020). "This is the child we were given": A qualitative study of Danish parents' experiences of a prenatal Down syndrome diagnosis and their decision to continue the pregnancy. S*exual and Reproductive Healthcare, 23.* doi:10.1016/j.srhc.2019.100480

Luz, R., George, A., Spitz, E., & Vieux, R. (2017). Breaking bad news in prenatal medicine: A literature review. *Journal of Reproductive and Infant Psychology, 35(1),* 14-31, doi:

Madeuf, A, Roman, H, Verspyck, E. (2016). Continuation of pregnancy despite a diagnosis of severe fetal anomaly: A retrospective French study. *Acta Obstetricia Gynecologica Scandanavica, 95,* 934–940. doi: 10.1111/aogs.12906

*McKechnie, A. C., & Pridham, K. (2012). Preparing heart and mind following prenatal diagnosis of complex congenital heart defect. Qualitative *Health Research, 22(12),* 1694–1706. doi:10.1177/1049732312458371

*McKechnie, A. C., Pridham, K., & Tluczek, A. (2015). Preparing heart and mind for becoming a parent following a diagnosis of fetal anomaly. *Qualitative Health Research, 25(9),* 1182–1198. doi:10.1177/1049732314553852

Michalik, A., & Preis, K. (2014). Demographic factors determining termination of pregnancy following the detection of lethal fetal malignancy. The *Journal of Maternal-Fetal & Neonatal Medicine, 27(13),* 1301–1304. doi:10.3109/14767058.2013.856411

Navidian, A., Saravani, Z., & Shakiba, M. (2017). Impact of psychological grief counseling on the severity of post-traumatic stress symptoms in mothers after stillbirths. *Issues in Mental Health Nursing, 38(8),* 650–654. doi:10.1080/01612840.2017.1315623

*O'Connell, O., Meaney, S, & O'Donoghue, K. (2019). Anencephaly; the maternal experience of continuing with the pregnancy: Incompatible with life but not with love. *Midwifery, 71,* 12-18. doi:10.1016/j.midw.2018.12.016

Olsson, M. B., & Hwang, C. P. (2008). Socioeconomic and psychological variables as risk and protective factors for parental well-being in families of children with intellectual disabilities. Journal of Intellectual Disability Research, 52(12), 1102–1113. Doi:j.1365-2788.2008.01081.x

Oulton, K, Sell, D, Gibson, F. (2020). Hospitalized children with intellectual disability: Parents as partners in their care. *Journal of Applied Research in Intellectual Disabilities, 33,* 917– 926. doi:10.1111/jar.12713

Ponterotto, J. G. (2002). Qualitative research methods: The fifth force in psychology. *The Counseling Psychologist, 30(3),* 394–406. doi:10.1177/0011000002303002

Pös, O., Budiš, J., & Szemes, T. (2019). Recent trends in prenatal genetic screening and testing. *F1000Research, 8,* F1000 Faculty Rev-764. doi:10.12688/f1000research.16837.1

Reid, C., Gill, F., Gore, N., & Brady, S. (2016). New ways of seeing and being: Evaluating an acceptance and mindfulness group for parents of young people with intellectual disabilities who display challenging behaviour. *Journal of Intellectual Disabilities, 20(1),* 5–17. doi:10.1177/1744629515584868

Richter, L., Slemming, W., Norris, S.A., Stein, A., Poston, L., & Pasupathy, D. (2020). Health pregnancy, healthy baby: Testing the added benefits of pregnancy ultrasound scan for child development in a randomised control trial. *Trials, 21(25).* doi:10.1186/s13063-019-3924-0

Rich-Edwards, J. W., Kleinman, K., Abrams, A., Harlow, B. L., McLaughlin, T. J., Joffe, H., & Gillman, M. W. (2006). Sociodemographic predictors of antenatal and postpartum depressive symptoms among women in a medical group practice. Journal of *Epidemiology and Community Health, 60(3),* 221–227. doi:10.1136/jech.2005.039370

Rini C., & Dunkel Schetter, C. (2010). The effectiveness of social support attempts in intimate relationships: "T ain't what you do, it's the way that you do it." In K. Sullivan, J. Davila (Eds.), *Support Processes in Intimate Relationships.* (pp. 26–70). New York: Oxford Univ. Press

Rohde, A., Woopen, C., & Gembruch, U. (2008). Development in prenatal diagnosis -- Changes in experiencing pregnancy and effects of pathological prenatal diagnosis. *Journal of Family Research, 20(1), 62–79.* doi: 10.3238/arztebl.2010.0857

Schwartz, H., McCusker, J., Law, S., Zelkowitz, P., Somera, J., & Singh, S. (2021). Perinatal mental healthcare needs among women at a community hospital. *Journal of Obstetrics and Gynaecology Canada, 43(3)*, 322–328.e1.

Slade, P., Dembinsky, M., Bristow, K., Garthwaite, K., Mahdi, A., James, A., Rahman, A., & Downe, S. (2021). Facilitating perinatal access to resources and support (PeARS): A feasibility study with external pilot of a novel intervention. *BMC Pregnancy and Childbirth, 21,* 1-14. doi:10.1186/s12884-021-04112-w

Statham, H., Solomou, W., & Chitty, L. (2000). Prenatal diagnosis of fetal abnormality: Psychological effects on women in low risk pregnancies.*Best Practice and Research Clinical Obstetrics Gynae-*

cology, 14(4),731–147. doi:10.1053/beog.2000.01

Stock, N. M., Costa, B., Williams, J. R., & Martindale, A. (2019). Breaking the news: Parents' experiences of receiving an antenatal diagnosis of cleft lip. T*he Cleft Palate-Craniofacial Journal, 56(9),* 1149–1156. doi:10.1177/1055665619830884

Thomas, J., Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology, 8,* 45. doi:10.1186/1471-2288-8-45

Tol, W. A., Ebrecht, B., Aiyo, R., Murray, S. M., Nguyen, A. J., Kohrt, B. A., Ndyanabangi, S., Alderman, S., Musisi, S., & Nakku, J. (2018). Maternal mental health priorities, help-seeking behaviors, and resources in post-conflict settings: a qualitative study in eastern Uganda. *BMC Psychiatry, 18(1),* 39. doi:10.1186/s12888-018-1626-x

Urizar, G. G., Jr, Caliboso, M., Gearhart, C., Yim, I. S., & Dunkel Schetter, C. (2019). Process evaluation of a stress management program for low-income pregnant women: The SMART moms/mamás LÍSTAS project. *Health Education & Behavior, 46(6),* 930–941. doi:10.1177/1090198119860559

Vacaru, S., Beijers, R., Browne, P.D., Cloin, M., van Bakel, H., van den Heauvel, M., & de Weerth, C. (2021). The risk and protective factors of heightened prenatal anxiety and depression during the COVID-19 lockdown. *Scientific Reports, 11,* 20261. doi:10.1038/s41598-021-99662-6

Van der Steen, S.L., Riedijk, S.R., Verhagen-Visser, J., Govaerts, L.C.P., Srebniak, M.I., van Opstal, D., Joosten, M., Tibben, A., Diderich, K.E.M., & Galjaard, R.J.H. (2016). The psychological impact of prenatal diagnosis and disclosure of susceptibility loci: First impressions of parents' experiences. *Journal of Genetic Counseling, 25(6),* 1227–1234. doi:10.1007/s10897-016-9960-y

Winn, P., Acharya, K., Peterson, E., & Leuthner, S. (2018). Prenatal counseling and parental decision-making following a fetal diagnosis of trisomy 13 or 18. *Journal of Perinatology, 38(7),* 788-796. doi: 10.1038/s41372-018-0107-x

Wittman, A. T., Hashmi, S. S., Mendez-Figueroa, H., Nassef, S., Stevens, B., & Singletary, C. N. (2016). Patient perception of negative noninvasive prenatal testing results. *AJP reports, 6(4),* e391–e406. doi:10.1055/s-0036-1594243

Woolf-King, S., Anger, A., Arnold, E., Weiss, S., & Teitel, D. (2017). Mental health among parents of children with critical congenital heart defects: A systematic review. *Journal of the American Heart Association, 6(2).* http://dx.doi.org/10.1161/jaha.116.004862

World Health Organization. (2012). Screening the Genes. Bulletin of the *World Health Organization, 90(8)*:557–632. Retrieved November 4th, 2020, from https://www.who.int/bulletin/volumes/90/8/12-030812/en/

Table 1

Characteristics of Included Studies

Author, Year, and Country	Stated Aim	n	Fetal Diagnosis	Qualitative data source	Quality
Carlsson *et al.* (2015) Sweden	Explore parental experiences and need for information following a prenatal diagnosis	11	Congenital anomaly	Qualitative interviews	*High*
Carlsson *et al.* (2017) Sweden	Explore written statements where parents communicate their experience with a prenatal diagnosis	18	Congenital anomaly	Search of online discussion responses	*High*
Clark *et al.* (2020) United States	Describe the adjustment process of parents after receiving a prenatal diagnosis	42	Down Syndrome	Qualitative interviews	*High*
Côté-Arsenault & Denney-Koelsch (2011) United States	Investigate the parents' pregnancy experiences to gain insight into their needs	8	Lethal Fetal Diagnosis (e.g. Trisomy 13 or 18, Anencephaly)	Qualitative interviews	*High*
Goff *et al.* (2013) United States	Explore parent experiences upon receiving a prenatal diagnosis	46 (prenatal sample)	Down Syndrome	Online survey with open-ended questions	*Moderate*
Hickerton *et al.* (2011)	Investigate how parents experience a pregnancy with a	9	Genetic conditions	Qualitative interviews	*High*
How *et al.* (2019) Australia	Explore the views of fathers towards prenatal diagnosis	5	Down Syndrome	Qualitative interviews	*High*
Im *et al.* (2018) South Korea	Explore the pregnancy experience of mothers who receive a prenatal diagnosis	12	Congenital heart defect	Qualitative interviews	*High*
Johnson *et al.* (2018) United Kingdom	Explore parent experiences with a prenatal diagnosis and understand how family care could be improved	20	Dysmelia	Qualitative interviews	*High*
Lokmic *et al.* (2017) Australia	Investigate the views and experiences of parents who received a prenatal diagnosis	5	Lymphatic malformation	Qualitative interviews	*Moderate*
Lou *et al.* (2020) Denmark	Explore how parents with a prenatal diagnosis experience the diagnostic process and make decisions	9	Down Syndrome	Qualitative interviews	*High*
McKechnie & Pridham (2012) United States	Examine the retrospective accounts of parents who experienced a prenatal diagnosis	16	Congenital heart defect	Qualitative interviews	*High*
McKechnie *et al.* (2015) United States	Examine how parenting develops after receiving a prenatal diagnosis	37	Fetal anomaly	Qualitative interviews	*High*
O'Connell *et al.* (2019) Ireland	Investigate the experiences of pregnancy and prenatal care of mothers who continue pregnancy after receiving a fetal diagnosis	4	Anencephaly	Qualitative interviews	*High*

Table 2

Summary of Themes

Study	Initial Reactions					Processing Period		
	Shock	Grief	Anger/Confusion	Fear	Guilt	Acceptance	Anxiety	Attachment/ Hope
Carlsson *et al* (2015)		X	X	X		X	X	
Carlsson *et al* (2017)	X	X	X		X	X	X	X
Clark *et al* (2020)	X	X			X	X	X	X
Côté-Arsenault & Denney-Koelsch (2011)	X	X			X			X
Goff *et al* (2013)	X		X			X	X	
Hickerton *et al* (2011)	X	X	X		X	X	X	
How *et al* (2019)	X	X				X	X	
Im *et al* (2018)	X	X	X	X	X	X	X	X
Johnson *et al* (2018)	X	X					X	
Lokmic *et al* (2017)	X				X			
Lou *et al* (2020)		X		X		X	X	X
McKechnie & Pridham (2012)	X	X	X			X	X	X
McKechnie *et al* (2015)		X		X		X	X	X
O'Connell *et al* (2019)	X	X					X	X

Table 2 (continued)

Summary of Themes

Study	Influencing Factors							
	Social Interactions				Coping Strategies			
	Healthcare		Friends and Family		Information-Gathering	Experiential Knowledge	Healthcare Planning	Perspective-Taking
	Positive	Negative	Positive	Negative				
Carlsson *et al* (2015)	X	X			X	X	X	
Carlsson *et al* (2017)		X	X	X	X	X		
Clark *et al* (2020)		X	X	X	X	X		X
Côté-Arsenault & Denney-Koelsch (2011)	X	X	X	X				
Goff *et al* (2013)	X	X	X					X
Hickerton *et al* (2011)	X	X	X	X				
How *et al* (2019)	X	X		X		X		X
Im *et al* (2018)	X	X			X		X	X
Johnson *et al* (2018)		X			X	X	X	
Lokmic *et al* (2017)		X			X	X		
Lou *et al* (2020)	X	X			X			X
McKechnie & Pridham (2012)	X		X		X	X	X	
McKechnie *et al* (2015)	X	X	X	X	X	X	X	X
O'Connell *et al* (2019)	X	X	X	X	X	X		

Figure 1

PRISMA Flow Diagram

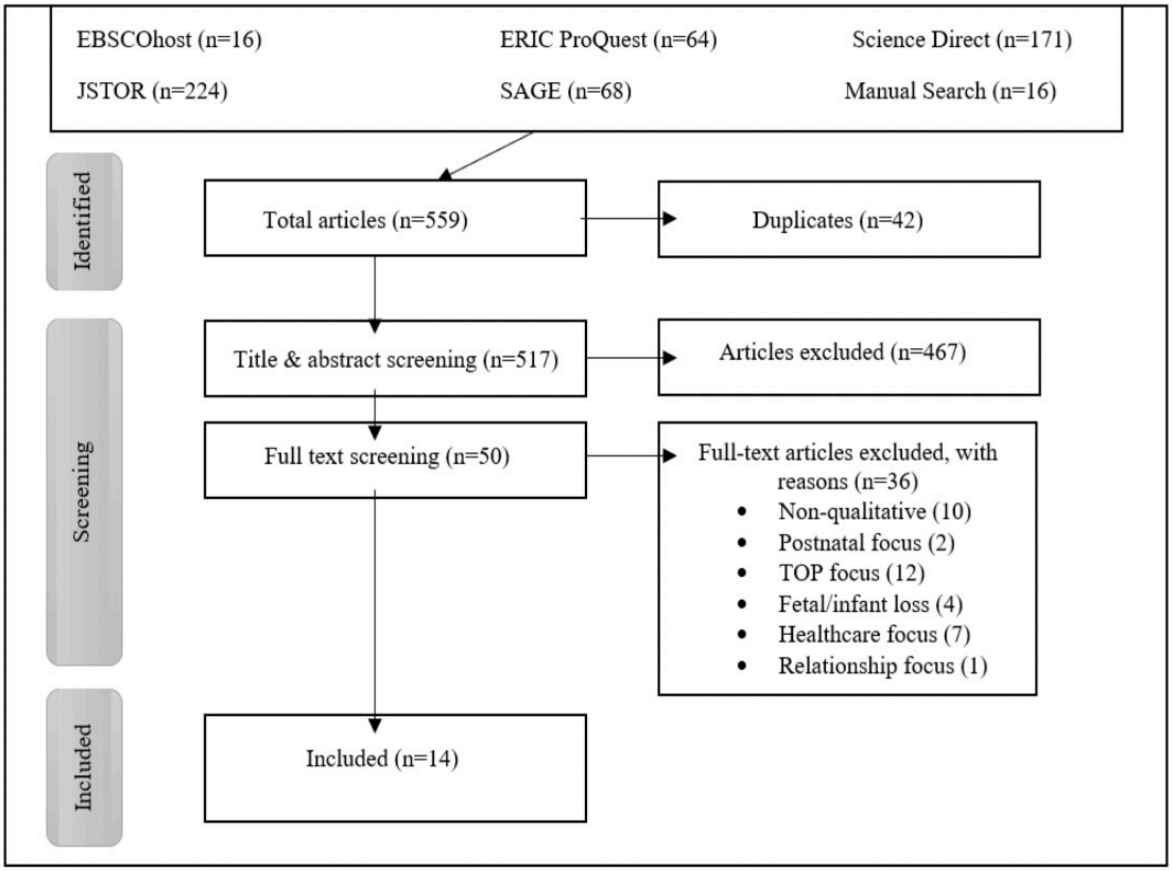
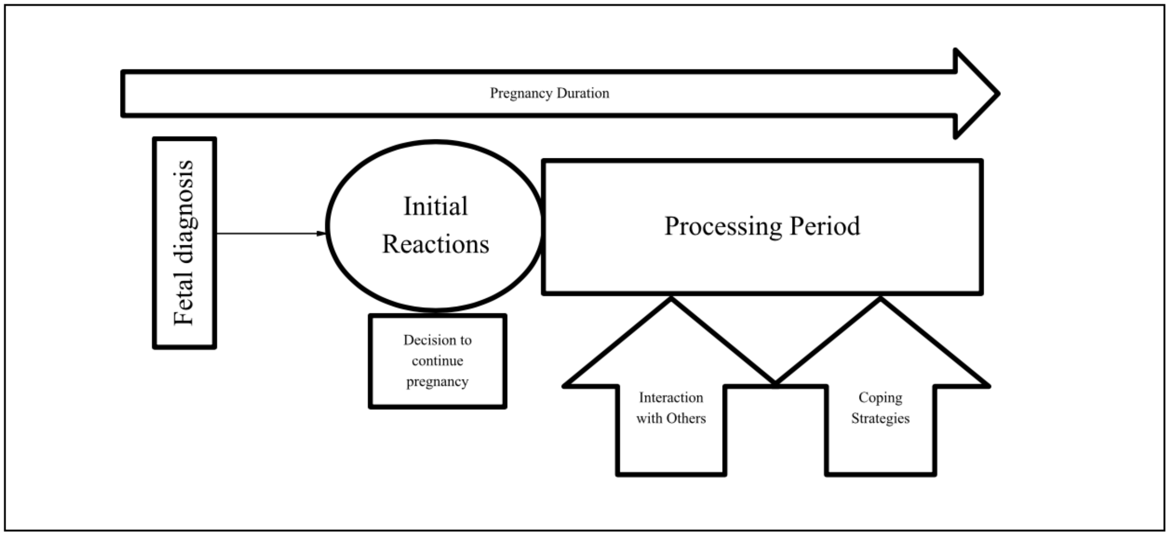


Figure 2

Model of Parent Psychological Processing After a Fetal Diagnosis



Graduate Student Journal of Psychology
2022 Vol. 19

Executive Function in Weight Loss Maintenance: The Moderating Role of Socioeconomic Status

Kathryn P. King,[1,2] Casie H. Morgan,[1] Gareth R. Dutton,[3] Sylvie Mrug,[1] Alena C. Borgatti,[1,3]
and Marissa A. Gowey,[2]
[1] Department of Psychology, University of Alabama at Birmingham
[2] Department of Pediatrics, University of Alabama at Birmingham
[3] Department of Medicine, University of Alabama at Birmingham

Few individuals with overweight/obesity maintain weight loss. Executive function (EF) and socioeconomic status (SES) contribute to weight loss maintenance (WLM). This study examined whether the relationship between EF and WLM differs across SES. Forty-four participants between 32-78 years of age were assessed ≥ 1-year post- behavioral obesity intervention. Those who achieved >5% weight loss during the program were recruited for the present study. Participants (N = 44) previously lost >5% of initial body weight. Hierarchical regressions tested the moderating role of SES in the relationship between performance-based EF [Iowa Gambling Task (IGT)] or self-report EF [Behavior Rating Inventory of Executive Function (BRIEF-A)] and %WLM. The relationship between performance-based EF and %WLM varied across SES ($p < .05$). For those with high SES, a 1-point T-score increase on IGT corresponded with 4.5% greater %WLM ($\beta = .52$, $p = .03$). No association was observed for those with low SES ($\beta = -.12$, $p = .54$). For those with low SES, greater EF may not benefit WLM. For those with high SES, greater EF may benefit WLM. Personalized WLM interventions accounting for levels of SES and EF may best facilitate WLM.

Keywords: executive function, weight loss maintenance, socioeconomic status, obesity

Approximately one-third of U.S. adults have obesity, with prevalence estimates increasing each year (Lundeen et al., 2018). Behavioral treatment for obesity is the gold-standard approach (Butryn et al., 2011). However, only half of the individuals achieve clinically significant weight loss (i.e., >5%) through these interventions (Ball & Crawford, 2002; Christian et al., 2010; Kraschnewski et al., 2010; Montesi et al., 2016). Further, only 20% of individuals maintain clinically significant weight loss >1-year post-treatment (Wing & Phelan, 2005), highlighting the significant challenge of weight loss maintenance (WLM).

A multitude of factors contribute to WLM, many of which relate to patients' socioeconomic and demographic characteristics (Fitzgibbon et al., 2012; Goode et al., 2017). One review demonstrated that occupation, education, and income all predicted weight change over time, with more socioeconomically disadvantaged participants having a greater risk of weight gain (Ball & Crawford, 2005). Because each of these constructs was related to poorer weight maintenance, it may be advantageous to utilize a measure of SES that captures the broader construct of SES related to weight maintenance (Ball & Crawford, 2005). These findings point toward the importance of identifying the combined influence of several measures of SES to understand the holistic influence of a disadvantaged background on WLM.

Additionally, several psychological variables, including executive function (EF), have been implicated in weight regain. EF refers to neuropsychological processing that controls and coordinates behaviors and cognitive abilities (Diamond, 2013). This typically includes skills pertaining to organization and regulation such as problem solving, decision making, reasoning, attention, planning, and time management. Deficits in impulse control (Giel et al., 2017) and related EF constructs have been repeatedly associated with reduced obesity treatment efficacy and greater weight regain (Montesi et al., 2016; Wing & Phelan, 2005; Elfhag & Rössner, 2005; Varkevisser et al., 2019). Further, constructs consistently related to executive dysfunction, such as binge eating (Boggiano et al., 2014; Striegel-Moore et al., 1998), eating in the absence of hunger, and emotional eating, have been associated with a greater weight regain (Giel et al., 2017; Elfhag & Rössner, 2005). Together, these studies suggest that EF plays a critical role in WLM.

SES and EF may interact to affect health outcomes as well. For example, in an intervention that trained EF skills, SES moderated improvement in EF skills, such that those from low SES families experienced greater improvement than those from high SES families, emphasizing the importance of including SES as a mod erator, rather than simply a covariate when examining the effects of EF on treatment (Schubert, 2016). Evidence from qualitative research supports this notion as well. One study exploring factors associated with dietary behavior indicated that low and mid-SES women emphasized the effect of employment-related time constraints on food preparation more than high-SES women (Inglis et al., 2005). Similarly, low-SES women, but not mid or high-SES women, named the cost of healthy food most frequently among food purchasing considerations (Inglis et al., 2005). These emphases reflect a high demand for resource management via organization and planning when preparing and purchasing foods (Inglis et al., 2005). Indeed, healthy food preparation can require a great deal of time and EF. Those with greater SES resources may be able to compensate for EF constraints by utilizing higher cost strategies (e.g., eating healthier quickly prepared foods due to lack of cost barrier and endorsing more opportunities to cook from home; Inglis et al., 2005) to accomplish EF-demanding health behaviors. Thus, these individuals may not experience the same degree of negative effects of EF on their WLM. Conversely, those with low SES may not be able to employ more costly coping strategies (Inglis et al., 2005) and subsequently experience greater negative effects of EF difficulties on WLM.

Although initial evidence suggests that SES may interact with EF to influence health behavior or WLM, the literature has yet to examine this moderation effect. Evaluating the interaction between SES and EF on WLM would elucidate risk and resilience factors in WLM and has the potential to inform precision medicine approaches to WLM (e.g. identifying who may benefit from interventions targeting resources and/or EF skills). As such, the present paper aims to examine whether SES moderates the relationship between EF and WLM in a racially-diverse group of individuals who lost a clinically significant amount of weight via lifestyle modification. We hypothesized that higher EF will be associated with greater WLM among those with low SES, but be unrelated to WLM among those with high SES.

Method

Participants

Forty-four participants between 32-78 years of age (M = 57.43 years, SD = 11.71) were recruited from previous participants of a behavioral obesity intervention. The original intervention, Improving Weight Loss Maintenance Through Alternative Schedules of Treatment (ImWeL, NCT02487121), consisted of weekly sessions involving evidence-based dietary modifications, increased physical activity, and behavioral strategies designed to promote adherence to these lifestyle changes, delivered by trained interventionists (for more information, see Gowey et al., 2021). For the original intervention, participants were recruited through the local newspaper, television, flyers, and the university-affiliated website and e-newsletter advertisements. For the current study, participants were contacted 2-4 years post-intervention on a rolling basis for six months. Individuals were eligible for recruitment if ≥ 5% weight loss was achieved during ImWeL. Eligibility was confirmed based on study records of weight loss history. Participants were excluded if they had (a) a history of bariatric surgery, (b) unintentional weight loss since participating in the previous weight loss trial, or (c) a medical condition influencing body weight. The current sample was predominantly female and racially diverse (93% female, 55% African American/other, 45% White, see Table 1). The study was approved by the university's Institutional Review Board.

Procedure

Individuals were recruited via mailed letters and telephone calls to assess eligibility. All 44 participants contacted for this study were interested and eligible to enroll in the study. They were scheduled for a two-hour study visit where informed consent procedures were conducted, after which anthropometry measurements were taken, surveys were completed, and EF testing was conducted by a trained graduate student under the supervision of a PhD-level clinical psychologist.

Measures

Demographic information

Participants self-reported their age, educational attainment, medical history, race, ethnicity, marital status, and household income.

Socioeconomic Status

SES was measured by averaging standardized income and education variables (e.g., Pu & Rodriguez, 2021; Rodriguez et al., 2021; Gardner et al., 2017). Education was reported on a 5-point scale, ranging from (1) Less than a high school diploma to (5) Graduate school. Annual total gross family income was reported on an 11-point scale, with the following values: 0) Under \$10,000, 1) \$10-20,000, 2) \$20-30,000, 3) \$30-40,000, 4) \$40-50,000, 5) \$50-60,000, 6) \$60-70,000, 7) \$70-80,000, 8) \$80-90,000, 9) \$90-100,000, 10) Over \$100,000. For interaction analyses, simple slopes were calculated at 1 standard deviation above and below the mean according to best practices for moderation analyses when there are no meaningful cut points available (Memon et al., 2019). Thus, "High SES" refers to an SES level one standard deviation above the mean, or the 84th percentile. "Low SES" refers to an SES level at one standard deviation below the mean, or the 16th percentile. For reference, an income one SD above the mean would be an income between \$80-90,000 and an income one SD below the mean would be an income of about \$30,000. For education, one SD above the mean represents a doctoral or professional degree, while one SD below the mean represents some college, but no degree.

Anthropometric measurements

Trained staff measured participants' height and weight with shoes removed using a wall-mounted stadiometer and digital scale.

Percent weight loss maintenance (%WLM)

To determine %WLM, the following data were self-reported by participants: the most weight they lost in their lifetime (initial weight loss; Krueger & Reither, 2015; Santos et al., 2017) how much they weighed prior to losing that weight (start weight), how much they weighed after losing that weight (post weight), and their current weight which was measured objectively (see anthropometric measurements section). The following formula is based on prior literature (Ryder et al., 2005) and was used to calculate %WLM:

$$\frac{\text{initial weight loss} - (\text{current weight} - \text{post weight})}{\text{initial weight loss}}$$

Performance-based EF

The Iowa Gambling Task (IGT; Bechara, 2007) was utilized to measure performance-based EF. The IGT measures decision-making using four virtual decks of cards. The participant is instructed to win as much money as possible and that cards will reward or penalize them. Participants are scored based on their use of good decks, which provide smaller rewards more often and have better net outcomes, versus bad decks, which provide larger rewards less often and have poorer net outcomes. A norm-referenced T-score (age-, gender-, race-, ethnicity-matched) is generated based on the total net score, with lower scores indicating more impaired decision making. Mixed results have been noted when comparing IGT performance to performance on other executive functions, decision making, and memory tasks, with impairments in cognitive skills more associated with "cold" decision making a likely cause for the inconsistencies (Buelow & Suhr, 2009). However, there is evidence to demonstrate that IGT shows good construct validity with some measures of executive function and decision-making, like the Wisconsin Sorting Card task (Brand et al., 2007, Buelow & Suhr, 2009)

Self-reported EF

The Behavior Rating Inventory of Executive Functioning (BRIEF-A) is a standardized self-report scale of EF that is well-validated and has demonstrated good internal consistency in adults with obesity (Roth et al., 2005; Rouel et al., 2016). There were moderate to high coefficient alphas for the nine clinical scales (α = 0.65–0.92), and high alphas for the three composite scales (α = 0.93, 0.95, and 0.97, respectively). Three subscales showed internal consistency below the expected value of 0.80 (α = 0.65, α = 0.78, α = 0.79). Participants rate the frequency with which certain behaviors have been a problem in the past month. Scoring of the 75-item questionnaire generates T-scores for the Global Executive Composite (GEC). Higher scores indicate more impaired EF.

Data Analyses

Descriptive statistics characterized key variables. Two candidate covariates (BMI, duration of WLM) were examined via correlations. Potential covariates that significantly correlated with %WLM were retained in the model. Moderation was tested in a hierarchical linear regression model. Step one included BMI as a covariate, step two added mean-centered SES and EF, and step three added the interaction between SES and EF. The hierarchical model was run separately for performance-based and self-reported EF. Significant interactions were followed up with simple slope testing at low and high SES (one standard deviation below and above the mean). All assumptions and analyses were tested via SPSS version 25.

Results

Preliminary Analyses

Descriptive statistics for key variables are reported in Table 1. Participants kept off approximately 13% of the total weight they lost in their lifetime on average. Spearman's correlations between potential covariates (BMI and duration of WLM) and primary variables of interest only revealed a negative correlation between BMI and %WLM ($r = -.41$, $p < .01$) (See Table 2). As expected, SES and Education were moderately associated ($r = .36$, $p < .05$) Thus, BMI was included as a covariate in the main analyses. All relevant assumptions for moderation using hierarchical multiple regression were tested and met.

Moderation Analyses

Performance-based EF (IGT)

The hierarchical regression model testing SES as a moderator of the relationship between IGT and %WLM was significant, $R2 = 0.24$, $F(4, 39) = 3.11$, $p < .05$; see Table 3. In the first step, higher BMI predicted lower %WLM, $\beta = -0.37$, $p < .05$; $R2 = 0.14$, $F(1, 42) = 6.72$, $p < .05$. In the second step, IGT and SES did not uniquely predict %WLM, $\Delta R2 = 0.02$, $\Delta F(2, 40) = 0.40$, $p = .11$. In step three, however, EF significantly interacted with SES in predicting %WLM, $\beta = 0.31$, $p < .05$; $\Delta R2 = 0.09$, $\Delta F(1, 39) = 4.47$, $p = < .05$, $b = 0.31$, $p < .05$. Simple slope analyses showed a positive effect of EF on WLM at higher levels of SES, a one-point t-score increase in IGT corresponded with a 4.5% increase in %WLM ($\beta = 0.52$, $p < .05$), while at lower levels of SES, there was no relationship between IGT and %WLM ($\beta = -0.12$, $p = .54$); see Figure 1. A post-hoc power analysis for the final model demonstrated that with $\Delta R2 = 0.09$, $f2 = 0.10$, $N = 44$, $\alpha = 0.05$, and four predictors, the achieved power to detect the moderation effect was 0.66.

Self-reported EF (BRIEF)

The hierarchical regression model testing SES as a moderator of the relationship between the BRIEF and %WLM was not significant, $R2 = 0.15$, $F(4, 39) = 1.66$, $p = .18$; see Table 3. In the first step, higher BMI predicted lower %WLM, $\beta = -0.38$, $p < .05$; $R2 = 0.14$, $F(1, 42) = 6.72$, $p < .05$. In the second step, the BRIEF and SES did not significantly predict %WLM, $\Delta R2 = 0.01$, $\Delta F(2, 40) = 0.16$, $p = .85$. In step three, the BRIEF did not significantly interact with SES to predict %WLM, $\beta = .03$, $p = .86$; $\Delta R2 = 0.001$, $\Delta F(1, 39) = .03$, $p = .86$, $b = .026$, $p = .86$. A post-hoc power analysis for the final model demonstrated that with $\Delta R2 = 0.001$, $f2 = 0.001$, $N = 44$, $\alpha = 0.05$, and four predictors, the achieved power to detect the moderation effect was 0.08.

Discussion

The goal of the present study was to examine the degree to which SES moderates the relationship between EF and WLM to address gaps in the WLM literature that may inform precision medicine approaches. Given recent studies demonstrating relationships between SES, EF, and weight loss outcomes, we examined whether individuals from different SES backgrounds showed unique relationships between EF and %WLM. EF was measured via a performance-based test and self-reports, as these methods provide unique information about EF and do not correlate highly with each other (Garcia et al., 2013; Toplak et al., 2013). As expected, findings indicated that the relationship between performance-based EF and %WLM was dependent on SES; contrary to our expectation, however, those with high SES experienced a greater benefit of performance-based EF on %WLM than individuals with low SES. Regarding self-reported EF, our hypothesis was not supported, as SES and self-reported EF did not interact to affect an individual's %WLM.

Access to high-cost coping strategies in high-SES individuals may best explain the unique relation between EF and SES in high-SES individuals. It is likely that for high SES individuals, having access to an abundance of weight management resources (e.g., grocery stores, gym memberships/classes, meal preparation services, smartphone applications, and gadgets, etc.) may be more efficiently accessed and utilized for an individual with stronger EF skills. For example, individuals with higher SES may be more likely to own a wearable device to monitor activity, and those with stronger EF skills may be more likely to utilize the tracking features (e.g., weight, food, and exercise tracking) on the device or its associated phone app. Alternatively, EF deficits could also be uniquely hindering those with high SES, perhaps due to increased access to unhealthy foods and mismanagement of extra resources.

In contrast, lower SES individuals often lack basic access to these same resources (Ailshire & House, 2011); thus, EF may manifest differently in each of these scenarios. For higher SES individuals with abundant opportunities, there is a need to organize options, utilize self-control with grocery shopping, and manage memberships efficiently and accurately. Alternatively, for lower SES individuals maintaining weight loss, there are fewer resources through which to apply EF skills of coordinating, organizing, and managing, so EF abilities may have a more limited “range” of impact. In fact, for low SES individuals, the weight-loss intervention program itself may be the primary resource accessible to this group for healthy eating and activity. Once the program ends, these individuals may not have the community structures (e.g., gyms, healthy food markets, etc.) in place to support previous efforts. This interpretation is supported by the recent emphasis on the relationship between social determinants of health and adverse health outcomes (Medvedyuk et al., 2018).

The use of a performance-based EF task is a notable strength of the study design. IGT is specifically designed to detect decision-making deficits and does so in the context of financial gains and losses (Bechara, 2007). One interpretation of these outcomes could imply unique interactions between SES and a financially-oriented EF-dependent task. Although the correlation between IGT and SES was weak and nonsignificant (see Table 2), there are financial patterns across different SES groups that are worthy to note. For example, individuals with low SES experience frequent financial uncertainties which often present as stressors and constraints, rather than solvable complications (Chen & Miller, 2013). In the context of weight management, which can be characterized as a stressor due to the extensive behavior change, resource allotment, and commitment required to maintain success, if low-SES families are attempting to balance weight-related stressors with financial stressors, a “spiral of resource loss” (Hobfoll, 2001) can occur (e.g., a parent misses work to take care of a sick child, loses a job, can’t afford gym membership). This financial uncertainty may lead to buying cheaper, unhealthy foods or lower quantities of healthy foods. Thus, real-world decisions about money, food choices, and healthy access to food could be influencing behavior during this performance-based measure and influencing lifestyle choices in the real-world setting, amplifying the significant difference for higher SES individuals compared to lower SES individuals.

Self-reported EF was measured using the BRIEF-A questionnaire and is considered more of a global composite of different behaviors pertaining to EF abilities. Subjective rating scales tend to have more ecological validity than performance-based testing but can be stifled if someone has severe enough impairments that they are not aware of their deficits or of the impact these deficits have on everyday behavior (Barkley, 2012; Chan, 2008). A self-rating scale that requires insight into one’s own cognitive abilities may be inherently difficult for someone with impairment in self-awareness as compared to performance-based testing which is rated by a trained observer (Buchanan, 2016), which could explain some of the discrepancies between the performance-based and self-reported EF results.

Limitations

This study has some limitations that should be mentioned. One of the most important limitations is the sample size, which reduced statistical power and did not now allow more complex modeling techniques, such as additional predictors or covariates. The post hoc analyses revealed low power to detect effects, supporting the notion that a larger sample size may improve power and allow for more complex modeling. Given that this study enrolled only those who lost a clinically significant amount of weight, future studies using similar designs may benefit from over-recruitment during a weight loss intervention to allow for a larger recruitment pool of those who lose a clinically significant amount of weight. Alternatively, future research could consider more large-scale designs, such as that of the National Weight Control Registry (Hill et al., 2005). However, with this approach, measurements would need to be adapted for remote data collection, which would introduce another limitation in exchange for an increased sample size. A second limitation was the composition of the sample. The majority of the sample was female, limiting the generalizability of findings to weight loss experiences for males. Despite these limitations, the current study represents an important step toward prioritizing SES and EF in weight management interventions and considering the impacts individual differences and characteristics have on WLM.

For the present study, the best two factors to capture SES included educational history and race, however, it is understood that other variables can be included to strengthen SES as a construct. One recent study acknowledged the complexity of measuring and conceptualizing SES and included a sam ple of additional criteria to be considered in future research (Rodríguez-Hernández et al., 2020). Specifically, they highlight parental education, family income, parental occupation, household resources, and neighborhood resources. Alternatively, SES can also be considered subjective, with perceived SES demonstrating its own separate impact on health outcomes (Nobles et al., 2013) compared to objective components of SES. Therefore, future research should also carefully consider the conceptualization and measurement of SES when studying weight management and could consider the influences of both perceived SES and more objective SES factors related to actual income, occupation, and education status.

Future studies may benefit from expanding upon the present findings. For example, efforts could be made to recruit males and examine sex differences in the studied relationships. Additionally, it may be advantageous to recruit a mix of individuals with varying degrees of success with WLM, including those experiencing weight regain. This allows for more variance in weight maintenance outcomes and allows for an improved investigation of potential barriers to WLM. One final consideration includes isolating the different clinical domains captured in the BRIEF to examine unique associations between individual EF domains, WLM, and SES. Continuation of this line of research could ultimately inform the development of precision medicine strategies that take such relationships into account in treatment selection and delivery.

Conclusion

The present findings suggest that for those with high SES, who already possess basic financial and community resources, higher EF may facilitate the ability to organize, prioritize, and efficiently access available weight management tools and strategies. For those with lw SES who may lack financial and community resources, research should examine the benefit of reducing barriers to such resources via local programs or providing continuous, free, or low-cost access to WLM treatment programs. Overall, if individuals are provided with personalized WLM support aligned with their levels of SES and EF, they may develop the necessary skills to succeed in lifetime WLM.

References

Ailshire, J. A., & House, J. S. (2011). The unequal burden of weight gain: An intersectional approach to understanding social disparities in BMI trajectories from 1986 to 2001/2002. *Social Forces, 90(2),* 397–423. https://doi.org/10.1093/sf/sor001

Ball, K., Brown, W., & Crawford, D. (2002). Who does not gain weight? Prevalence and predictors of weight maintenance in young women. International *Journal of Obesity, 26(12)*, 1570–1578. https://doi.org/10.1038/sj.ijo.0802150

Ball, K., & Crawford, D. (2005). Socioeconomic status and weight change in adults: A review. *Social Science & Medicine, 60(9),* 1987–2010. https://doi-org.ezproxy3.lhl.uab.edu/10.1016/j.socscimed.2004.08.056

Barkley, R. A. (2012). *Executive functions: What they are, how they work, and why they evolved*. Guilford Press.

Bechara, A. (2007). *Iowa gambling task.* Psychological Assessment Resources.

Boggiano, M. M., Burgess, E. E., Turan, B., Soleymani, T., Daniel, S., Vinson, L. D., ... & Morse, A. (2014). Motives for eating tasty foods associated with binge-eating. Results from a student and a weight-loss seeking population. *Appetite, 83,* 160–166. https://doi.org/10.1016/j.appet.2014.08.026

Brand, M., Recknor, E. C., Grabenhorst, F., & Bechara, A. (2007b). Decisions under ambiguity and decisions under risk: correlations with executive functions and comparisons of two different gambling tasks with implicit and explicit rules. Journal of Clinical and Experimental *Neuropsychology, 29,* 86–99. https://doi.org/10.1080/13803390500507196

Buelow, M. T., & Suhr, J. A. (2009). Construct validity of the Iowa gambling task. *Neuropsychology Review, 19(1),* 102-114. https://doi.org/10.1007/s11065-009-9083-4

Butryn, M. L., Webb, V., & Wadden, T. A. (2011). Behavioral treatment of obesity. *Psychiatric Clinics, 34(4),* 841–859. https://doi.org/10.1016/j.psc.2011.08.006

Buchanan, T. (2016). Self-report measures of executive function problems correlate with personality, not performance-based executive function measures, in nonclinical samples. *Psychological Assessment, 28(4),* 372–385. https://doi.org/10.1037/

pas0000192

Chan, R. C., Shum, D., Toulopoulou, T., & Chen, E. Y. (2008). Assessment of executive functions: Review of instruments and identification of critical issues. *Archives of Clinical Neuropsychology, 23(2),* 201–216. https://doi-org.ezproxy3.lhl.uab.edu/10.1016/j.acn.2007.08.010

Chen, E., & Miller, G. E. (2013). Socioeconomic status and health: mediating and moderating factors. *Annual Review of Clinical Psychology, 9,* 723–749. https://doi.org/10.1146/annurev-clinpsy-050212-185634

Christian, J. G., Tsai, A. G., & Bessesen, D. H. (2010). Interpreting weight losses from lifestyle modification trials: using categorical data. *International Journal of Obesity, 34(1),* 207–209. https://doi.org/10.1038/ijo.2009.213

Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64,* 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Elfhag, K., & Rössner, S. (2005). Who succeeds in maintaining weight loss? A conceptual review of factors associated with weight loss maintenance and weight regain. *Obesity Reviews, 6(1),*67–85.https://doi.org/10.1111/j.1467-789X.2005.00170.x

Fitzgibbon, M. L., Tussing-Humphreys, L. M., Porter, J. S., Martin, I. K., Odoms-Young, A., & Sharp, L. K. (2012). Weight loss and African–American women: a systematic review of the behavioural weight loss intervention literature. *Obesity Reviews, 13(3),* 193–213. https://doi.org/10.1111/j.1467-789X.2011.00945.x

Garcia, S., Fedor, A., Spitznagel, M. B., Strain, G., Devlin, M. J., Cohen, R. A., ... & Gunstad, J. (2013). Patient reports of cognitive problems are not associated with neuropsychological test performance in bariatric surgery candidates. *Surgery for Obesity and Related Diseases, 9(5)*, 797–801. https://doi.org/10.1016/j.soard.2012.10.008

Gardner, M. H., Mrug, S., Schwebel, D. C., Phipps, S., Whelan, K., & Madan-Swain, A. (2017). Benefit finding and quality of life in caregivers of childhood cancer survivors: the moderating roles of demographic and psychosocial factors. *Cancer Nursing, 40(5),* E28. https://doi.org/10.1097/NCC.0000000000000419

Giel, K. E., Teufel, M., Junne, F., Zipfel, S., & Schag, K. (2017). Food-related impulsivity in obesity and binge eating disorder—a systematic update of the evidence. *Nutrients, 9(11),* 1170. https://doi.org/10.3390/nu9111170

Goode, R. W., Styn, M. A., Mendez, D. D., & Gary-Webb, T. L. (2017). African Americans in standard behavioral treatment for obesity, 2001-2015: what have we learned?. *Western Journal of Nursing Research, 39(8),* 1045–1069. https://doi.org/10.1177%2F0193945917692115

Gowey, M. A., Neumeier, W. H., Henry, S., Wadley, V. G., Phillips, J., Hayden, K. M., Espeland, M. A., Coday, M., Lewis, C. E., & Dutton, G. R. (2021). Executive function in individuals with clinically significant weight loss via behavioral intervention. *Obesity Science & Practice, 7(1)*, 25–34. https://doi.org/10.1002/osp4.458

Hill, J. O., Wyatt, H., Phelan, S., & Wing, R. (2005). The National Weight Control Registry: is it useful in helping deal with our obesity epidemic? *Journal of Nutrition Education and Behavior, 37(4),* 206-210. https://doi.org/10.1016/S1499-4046(06)60248-0

Hobfoll, S. E. (2001). The influence of culture, community, and the nested-self in the stress process: Advancing conservation of resources theory. A*pplied Psychology, 50(3),* 337–421. https://doi-org.ezproxy3.lhl.uab.edu/10.1111/1464-0597.00062

Inglis, V., Ball, K., & Crawford, D. (2005). Why do women of low socioeconomic status have poorer dietary behaviours than women of higher socioeconomic status? *A qualitative exploration. Appetite, 45(3),* 334–343. https://doi.org/10.1016/j.appet.2005.05.003

Kraschnewski, J. L., Boan, J., Esposito, J., Sherwood, N. E., Lehman, E. B., Kephart, D. K., & Sciamanna, C. N. (2010). Long-term weight loss maintenance in the United States. *International Journal of Obesity, 34(11),* 1644–1654. https://doi.org/10.1038/ijo.2010.94

Krueger, P. M., & Reither, E. N. (2015). Mind the gap: Race/ethnic and socioeconomic disparities in obesity. *Current Diabetes Reports, 15(11),* 1–9. https://doi.org/10.1007/s11892-015-0666-6

Lundeen, E. A., Park, S., Pan, L., O'Toole, T., Matthews, K., & Blanck, H. M. (2018). Obesity prevalence among adults living in metropolitan and nonmetropolitan counties—United States, 2016.

Morbidity and Mortality Weekly Report, 67(23), 653–658. https://doi.org/10.15585/mmwr.mm6723a1

Medvedyuk, S., Ali, A., & Raphael, D. (2018). Ideology, obesity and the social determinants of health: A critical analysis of the obesity and health relationship. *Critical Public Health, 28(5),* 573–585. https://doi.org/10.1080/09581596.2017.1356910

Memon, M. A., Cheah, J. H., Ramayah, T., Ting, H., Chuah, F., & Cham, T. H. (2019). Moderation analysis: issues and guidelines. *Journal of Applied Structural Equation Modeling, 3(1)*, 1-11.

Montesi, L., El Ghoch, M., Brodosi, L., Calugi, S., Marchesini, G., & Dalle Grave, R. (2016). Long-term weight loss maintenance for obesity: a multidisciplinary approach. *Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy, 9,* 37–46. https://doi.org/10.2147/DMSO.S89836

Nobles, J., Weintraub, M. R., & Adler, N. E. (2013). Subjective socioeconomic status and health: relationships reconsidered. *Social Science & Medicine, 82,* 58-66. https://doi.org/10.1016/j.socscimed.2013.01.021

Pu, D. F., & Rodriguez, C. M. (2021). Spillover and crossover effects: Mothers' and fathers' intimate partner violence, parent-child aggression risk, and child behavior problems. *Child Maltreatment, 26(4),* 420-430. https://doi.org/10.1177/1077559520985936

Rodriguez, C. M., Granger, D. A., & Leerkes, E. M. (2021). Testosterone associations with parents' child abuse risk and at-risk parenting: a multimethod longitudinal examination. *Child Maltreatment, 26(1),* 50-62. https://doi.org/10.1177/1077559520930819

Rodríguez-Hernández, C. F., Cascallar, E., & Kyndt, E. (2020). Socio-economic status and academic performance in higher education: A systematic review. *Educational Research Review, 29,* 100305. https://doi.org/10.1016/j.edurev.2019.100305

Roth, R. M., Gioia, G. A., & Isquith, P. K. (2005). *BRIEF-A: Behavior Rating Inventory of Executive Function-adult Version.* Psychological Assessment Resources.

Rouel, M., Raman, J., Hay, P., & Smith, E. (2016). Validation of the Behaviour Rating Inventory of Executive Function–Adult Version (BRIEF-A) in the obese with and without binge eating disorder. *Eating Behaviors, 23,* 58–65. https://doi.org/10.1016/j.eatbeh.2016.07.010

Ryder, J. R., Gross, A. C., Fox, C. K., Kaizer, A. M., Rudser, K. D., Jenkins, T. M., ... & Inge, T. H. (2018). Factors associated with long-term weight-loss maintenance following bariatric surgery in adolescents with severe obesity. International *Journal of Obesity, 42(1),* 102–107. https://doi.org/10.1038/ijo.2017.193

Santos, I., Sniehotta, F. F., Marques, M. M., Carraça, E. V., & Teixeira, P. J. (2017). Prevalence of personal weight control attempts in adults: a systematic review and meta-analysis. *Obesity Reviews, 18(1),* 32–50. https://doi.org/10.1111/obr.12466

Schubert, E. C. (2016). *Examining moderators of response to executive function reflection training: initial skill and socioeconomic status* (Doctoral dissertation, University of Minnesota).

Striegel-Moore, R. H., Wilson, G. T., Wilfley, D. E., Elder, K. A., & Brownell, K. D. (1998). Binge eating in an obese community sample. International J*ournal of Eating Disorders, 23(1),* 27–37. https://doi.org/10.1002/(SICI)1098-108X(199801)23:1%3C27::AID-EAT4%3E3.0.CO;2-3

Toplak, M. E., West, R. F., & Stanovich, K. E. (2013). Practitioner review: Do performance-based measures and ratings of executive function assess the same construct? *Journal of Child Psychology and Psychiatry, 54(2),* 131–143. https://doi.org/10.1111/jcpp.12001

Varkevisser, R. D. M., Van Stralen, M. M., Kroeze, W., Ket, J. C. F., & Steenhuis, I. H. M. (2019). Determinants of weight loss maintenance: a systematic review. *Obesity Reviews, 20(2),* 171–211. https://doi.org/10.1111/obr.12772

Wing, R. R., & Phelan, S. (2005). Long-term weight loss maintenance. *The American Journal of Clinical Nutrition, 82(1),* 222S–225S. https://doi.org/10.1093/ajcn/82.1.222S

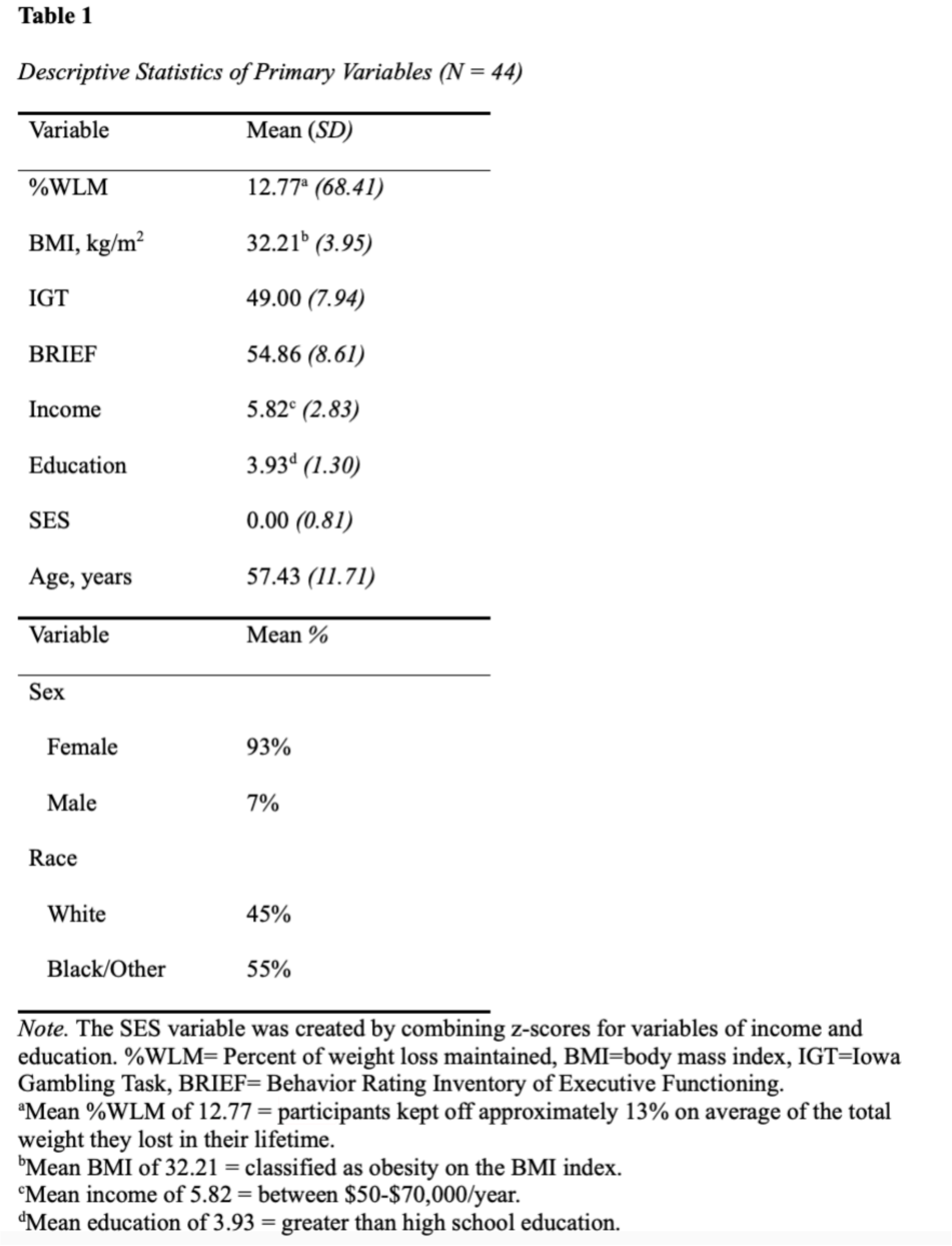


Table 1

Descriptive Statistics of Primary Variables (N = 44)

Variable	Mean (*SD*)
%WLM	12.77[a] *(68.41)*
BMI, kg/m^2	32.21[b] *(3.95)*
IGT	49.00 *(7.94)*
BRIEF	54.86 *(8.61)*
Income	5.82[c] *(2.83)*
Education	3.93[d] *(1.30)*
SES	0.00 *(0.81)*
Age, years	57.43 *(11.71)*

Variable	Mean %
Sex	
Female	93%
Male	7%
Race	
White	45%
Black/Other	55%

Note. The SES variable was created by combining z-scores for variables of income and education. %WLM= Percent of weight loss maintained, BMI=body mass index, IGT=Iowa Gambling Task, BRIEF= Behavior Rating Inventory of Executive Functioning.

[a] Mean %WLM of 12.77 = participants kept off approximately 13% on average of the total weight they lost in their lifetime.

[b] Mean BMI of 32.21 = classified as obesity on the BMI index.

[c] Mean income of 5.82 = between $50-$70,000/year.

[d] Mean education of 3.93 = greater than high school education.

Table 2

Spearman's Correlations between EF, SES, and Weight Variables (N = 44)

	1	2	3	4	5	6	7	8
1. %WLM	–							
2. Income	.12	–						
3. Education	.05	.36*	–					
4. SES	.08	.83**	.79**	–				
5. BRIEF	-.01	.18	-.22	-.03	–			
6. IGT	.25	.19	.11	.16	-.04	–		
7. BMI	-.41**	-.15	-.12	-.14	-.01	-.34*	–	
9. #Years Maintained	-.24	.01	.16	.06	-.24	.25	-.09	–

Note. * *p* < *.05;* ** *p* < *.01*

Table 3

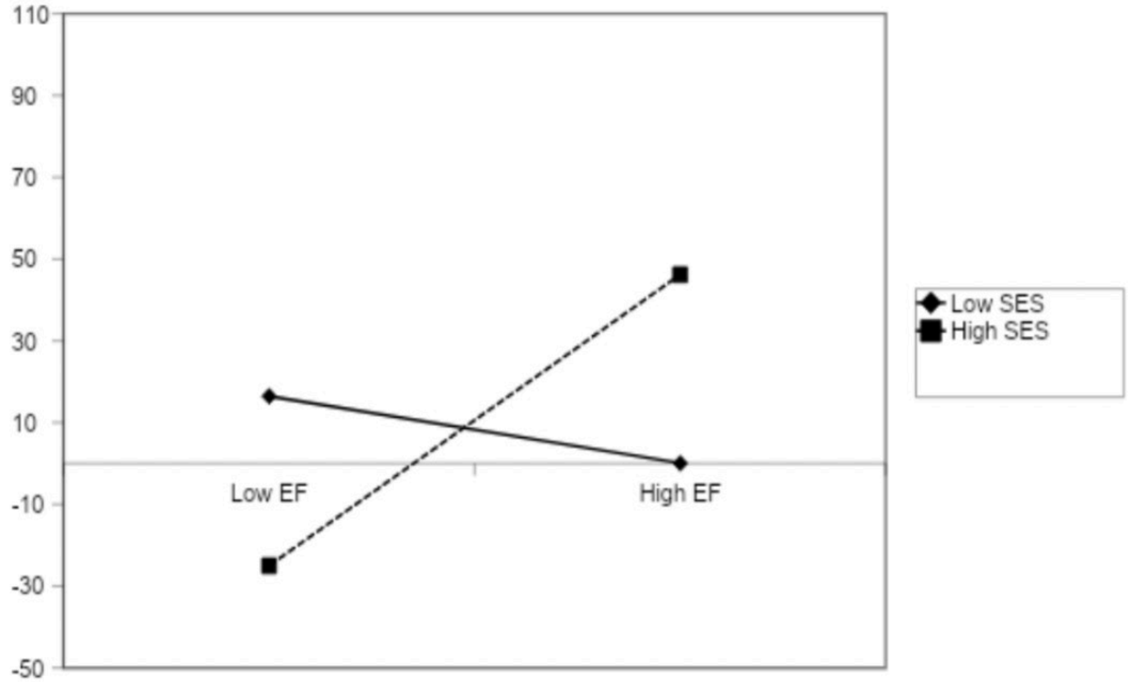
Summary of Hierarchical Multiple Regression Analyses of BMI, EF, and SES as Predictors of %WLM (N = 44)

	IGT (Performance-based)		BRIEF (Self-reported)	
Variable	β	ΔR^2	β	ΔR^2
Step 1		.138*		.38*
BMI	-.37*		-.37*	
Step 2		.017		.007
BMI	-.32*		-.38*	
SES	.02		.04	
EF	.14		.08	
Step 3		.087*		.001
BMI	-.25		-.38*	
SES	.02		.03	
EF	.20		.08	
EFxSES	.31*		.03	

Note. *$p < .05$.

Figure 1

Interaction Effect Between Performance-based EF and SES on % Weight Loss Maintenance (N = 44)



Note. Performance-based EF significantly interacted with SES in predicting %WLM.

Graduate Student Journal of Psychology
2022, Vol. 19

Memory Self-Efficacy and Community Participation

Genna M. Mashinchi and Craig Ravesloot,
Rural Institute for Disabilities, University of Montana

Objective: This study examines the relationships between working memory, memory self-efficacy (MSE), and community participation among older adults.
Method: 203 United States older adults (age 55+) were recruited through MTurk to complete surveys and a memory task. A multiple linear regression was used to regress MSE and community participation.
Main Findings: Entering all variables into the model explained 45% of the variance in community participation (R2 = .45, Adj. R2 = .41, F(7, 141) T = 16.26, $p < .001$). MSE was positively related to community participation ($\beta = .38$, $p < .001$), as predicted. However, contrary to hypotheses, worse working memory ($\beta = -.22$, $p = .001$) and greater difficulty remembering/concentrating ($\beta = .26$, $p < .001$) predicted higher levels of community participation.
Conclusions: It is reasonable to conclude that one's beliefs about their memory ability is an important consideration when one chooses to engage in community events.

Keywords: memory self-efficacy, community participation, aging, dementia, Alzheimer's disease.

As the world's older adult population grows, older individuals are susceptible to a heightened risk of experiencing age-related cognitive challenges, such as dementia (Aartsen et al., 2002, Fritsch et al., 2005). To help combat this growing problem, researchers have worked to identify factors that can support or improve cognitive functioning. One such factor that has gathered extensive support in the literature is participation in community leisure activities, such as ones that include social engagement (e.g., attending events with friends, having dinner with friends, participating in a card game group; Fratiglioni et al., 2004; Scarmeas et al., 2001; Sobral & Paúl, 2013; Verghese et al., 2003). However, participating in these activities can be difficult if one has difficulty remembering or concentrating, which might impair their ability to track a conversation, answer questions, or carry out multi-step instructions to complete an activity (Cowan, 2014). Additionally, if one feels that their cognitive abilities are poor, they might be more likely to avoid these activities, instead choosing to isolate themselves to bypass experiencing difficulties or feelings (Nieboer et al, 2020). To examine this idea further, the present study examined the relationships between memory self-efficacy, working memory, and community participation among older adults.

Memory Self-Efficacy

Memory self-efficacy (MSE) is defined as the beliefs an individual holds about their memory ability (Lalitha & Aswartha Reddy, 2021; Pearman & Trujillo, 2013; Sawin, 2021). The concept of MSE stems from Bandura's self-efficacy theory, which refers to an individual's self-perception of their ability to organize and execute tasks under given conditions (Bandura, 1997). In line with past research that has supported the positive relationship between MSE and memory performance, Bandura hypothesized that those with low self-efficacy perform poorer on tasks, compared to those with higher self-efficacy (Bandura, 1989; Beaudoin & Desrichard, 2011). As Bandura notes, this poorer performance occurs as those who doubt their ability to carry out a task are less invested in the tasks, which results in less effort, persistence, and motivation; setting lower goals for themselves; experiencing higher anxiety; and committing less to accomplish these goals (Beaudoin & Desrichard, 2011; Lalitha & Aswartha Reddy, 2021).

MSE is an important construct within metamemory that has been used to explain the cognitive decline that occurs with aging (Beaudoin & Desrichard, 2011; Hertzog et al.,1987). Past research has identified a correlation between MSE and memory task performance for older adults, with higher MSE associated with better memory performance (Lalitha & Aswartha Reddy, 2021; Pearman & Trujillo, 2013; Sawin, 2021). Other studies have found evidence that higher MSE is predictive of memory performance in cases of both laboratory and simulated-everyday episodic memory tasks (Turvey et al., 2000; West et al., 1996). In addition to being positively related to memory performance, such that low MSE is related to poor memory performance (Beaudoin & Desrichard, 2011; Lalitha & Aswartha Reddy, 2021; Pearman & Trujillo, 2013; Sawin, 2021), MSE has been found to be negatively related to beliefs about forgetting, such that stronger beliefs about forgetting and aging being related was associated with lower MSE (Vallet et al., 2015). In addition to being related to one's ability to remember, MSE has been shown to be related to one's ability to complete self-care tasks. Salinas (2021) found that self-efficacy is related to self-management behaviors in older adults who live alone. Similarly, Vellone et al. (2016) found that in those with impaired cognition, one's self-efficacy in being able to care for themselves was an important factor that influenced self-care abilities. In fact, one's self-efficacy mediated the relationship between working memory ability and self-care ability, illustrating the influence of self-efficacy even when working memory is poor. Given these findings, Vellone et al. recommended interventions that could increase self-efficacy and, in turn, improve self-care. Additionally, past research appears to suggest a gender difference in MSE, with Fallan and Opstad (2016), Huang (2012), and West et al. (2002) finding that males reported higher self-efficacy, and were more likely to overestimate their abilities, when completing cognitive tasks.

Working Memory

Working memory is a memory/executive functioning ability that allows one to actively hold and manipulate information for a brief amount of time, such as reordering numbers or completing mathematical problems (Aben et al., 2012; Cowan, 2008; McCabe et al., 2010; Miyake & Shah, 1999). Working memory has been measured in multiple ways, primarily with mental arithmetic or digit span tasks (Wechsler, 2008).

Due to its role in planning, working memory ability is needed to complete vital self-care tasks, such as remembering to take medications or remembering that one took the medications. In fact, Insel et al. (2006) found that working memory tasks and executive function were the only significant predictors in a model used to predict medication adherence. Poor working memory can also affect an individual's participation in the community, as working memory is necessary for several aspects of social-cognitive information processing, including tracking a conversation, the information presented, relationships between others, names just learned, and others' feelings towards topics (Meyer et al., 2012).

Present Study

Although past research has explored the relationship between MSE and memory ability in general (Lalitha & Aswartha Reddy, 2021; Pearman & Trujillo, 2013; Sawin, 2021), to the best of the authors' knowledge, only two studies have appeared to examine the relationship between MSE and working memory. One study, conducted by Hoffman and Schraw (2007), investigated the influence of self-efficacy and working memory on mathematical problem-solving performance. Hoffman and Schraw found that self-efficacy was beneficial as demands on working memory increased, and these findings proposed that one's ability to efficiently and strategically solve problems increased with self-efficacy. Additionally, Mashinchi et al. (2022) used a hierarchical regression analysis and found that MSE explained a large, unique portion of variance in working memory ability after controlling for age, depression, and anxiety.

Further, the authors of the present study are unaware of any research that has investigated the relationship between MSE and community participation. This study sought to fill this gap in the literature by examining the relationship between MSE and community participation. Hypotheses are as follows: 1) MSE and working memory ability will be positively correlated with community participation while difficulty in completing self-care tasks and difficulty remembering and concentrating will be negatively correlated with community participation, with all variables having statistically significant independent effects on community participation; and 2) noticing memory changes, difficulty in completing self-care tasks, and difficulty remembering and concentrating will all be negatively correlated with MSE while working memory ability will be positively correlated with MSE, with all variables having statistically significant independent effects on MSE.

Method

Participants

Participants were all United States residents over the age of 55 and were recruited through MTurk, an online community that completes surveys for monetary rewards. An MTurk filter was applied to ensure that all participants were United States residents. An MTurk age filter was applied, such that only participants that fit within the age parameter of 55 and older were able to participate in the study. In order to participate in the study, potential participants had to achieve an approved task completion rate (HIT rate) of 95%, meaning that they had to demonstrate worker quality by being approved by 95% of the researchers that they had completed studies for in the past. Participants were excluded if they reported an age that was not over 55, if more than 5% of their data were missing, or if their data did not appear to be of high quality. Two hundred and three eligible participants consented to participate in the study. One participant did not complete the demographic questionnaire but was included in the final analysis. A $0.50 monetary incentive was awarded to participants in exchange for their time. An a priori power analysis for a linear multiple regression, fixed model, single regression coefficient was conducted on G*Power 3.1. This power analysis was two-tailed, the alpha error probability was set to .05, and the desired power was set to .95. Results of this analysis yielded a sample size of at least 89 participants would be needed to achieve these parameters.

Assessments and Measures

Demographics

Self-reported demographic information regarding age, gender, ethnicity, and educational attainment was collected from each participant. In the analyses, gender was binary coded, with 1 = male and 2 = female. Additionally, participants were asked to indicate (yes/no) if they had noticed their ability to remember things had changed over the years. Further, participants completed the Washington Group Short Set of Questions on Disability (Centers for Disease Control and Prevention [CDC], 2015), which queries any difficulty with seeing, hearing, walking, remembering/concentrating, communicating, and/or completing self-care tasks using a 4-point Likert scale with 1 = no difficulty and 4 = cannot do at all. The present study analyzed the data for items pertaining to remembering/concentrating and completing self-care tasks.

Memory Self-Efficacy

To assess participants' MSE, the present study used an adaptation of the Memory Self-Efficacy Questionnaire (MSEQ; Berry et al., 1989), which is designed to assess participants' prediction of their memory ability. The authors of the present study adapted the MSEQ by including only the MSEQ's digit recall items. Further, the items were modified to specifically ask participants to predict their memory ability on the forward, backward, and sequencing conditions of the digit span memory task (adapted from Wechsler, 2008). The present study's measure of MSE was found to be highly reliable (α = .93), with alpha levels for each of the conditions as follows: forward (α = .78), backward (α = .87), and sequencing (α = .79). See Table 1 for items and item means

Community Participation

To assess community participation, participants were administered the ten-item "Undertaking Activities" section of the Maastricht Social Participation Profile (MSPP; Mars et al., 2009). This section examines both consumptive participation — which is defined as activities that allow an individual to benefit from the offerings of society (e.g., participating in a course or eating at a restaurant) — and formal social participation — which is defined as activities that allow an individual to offer a contribution to society (e.g., participating in organized volunteer work or organized clubs; Mars et al., 2009). The ten items asked participants to indicate the frequency with which they participated in specific community activities in the past four weeks. The present study used a 4-point Likert scale as follows: 1 = not at all, 2 = less than once a week, 3 = once or twice a week, and 4 = more than twice a week. The MSPP has been found to have strong convergent validity and discriminant validity with the Frenchay Activities Index, which is a measure of participation, similar to the MSPP (Mars et al., 2009)..

Working Memory Ability

Given the novelty of this project, the authors sought to use a reliable working memory task that has strong psychometric properties (Wechsler, 2008). Thus, a digit span task was used. This digit span task was similar to the Wechsler Adult Intelligence Scale-Fourth Edition Working Memory Index (WAIS-IV; Wechsler, 2008). In the present study, the digit strings presented to participants differed from the strings presented in the WAIS-IV, but the procedure was similar. The digits for this study's task were presented on screen, making this a visual working memory task, whereas the WAIS-IV's Digit Span Task is a verbal memory task. This change in format of the digit span task was made to avoid predicted difficulties participants might encounter when completing an auditory digit span task, such as the need for working speakers, headphones, or assistive audio technology.

Participants were asked to remember a set of numbers under three varied conditions: forward, backward, and sequencing. In the first condition, digit span forward, participants were instructed to recall the numbers in the same order in which they were presented. In the second condition, digit span backward, participants were instructed to recall the numbers in the reverse order with which they were presented (e.g., if presented 2-3, asked to recall it as 3-2). In the third condition, digit span sequencing, participants were asked to recall the digits presented in order from least to greatest in value (e.g., if presented 4-1-8, asked to recall it as 1-4-8).

The string of numbers was presented one by one in the middle of the screen for one second. The numbers and timing were programmed to auto advance on the screen by a timer feature. An extra number was added to the digit string with each additional trial. Once all digits of a string were presented, the screen changed to include a text box in which participants were instructed to type in each number string with one space between each number. The text box was programmed to recognize the correct answer. If correct, participants auto advanced to a digit string with an additional digit included. If not, participants were auto advanced to the second trial string, in which they were given another chance to answer a string with the same digit amount, identical to the WAIS-IV's Digit Span. If participants answered this string incorrectly, participants auto advanced to the next condition (e.g., backward). Scores were summed automatically by the software. The total Digit Span score ranges from 0-48, with each condition's score ranging from 0-16. The present study's digit span task had an internal reliability score of .92. Additionally, the internal consistency reliability for the three conditions are as follows: Digit Span Forward = .75, Digit Span Backward = .81, Digit Span Sequencing = .78.

Procedure

The Institutional Review Board at the University of Montana approved this study prior to data collection. Data were collected online using a Qualtrics-based survey that was posted on MTurk. First, participants reviewed the consent form and consented to participate in the study. Once written consent was obtained, all participants completed the MSE items, and then completed the digit span task. Next, participants completed the Washington Group Short Set, the MSPP, and the demographics questionnaire. Finally, all participants reviewed a debriefing form, outlining the purpose of the study, and received a code to input into MTurk to receive the monetary incentive for their participation.

Results

Participants

The age of participants ranged from 55 to 80 years ($M = 65.25$, $SD = 4.90$) and were predominantly female (67%) and Caucasian (66%). Ninety-two percent of participants had an education greater than a high school degree. Thirty-three participants reported that they noticed that their ability to remember things had changed over the years. Seventy-eight participants reported that they did not have difficulty completing self-care tasks, while 64% of participants reported that they did not have difficulty remembering or concentrating. See Table 2 for the full demographic statistics of the sample.

Hypothesis Tests

The assumptions of linearity, normally distributed errors, and uncorrelated errors were assessed for all variables. The Shapiro-Wilks tests for each variable was significant ($ps <.05$), and a visual analysis of the data revealed a slight positive skew in the distribution, as well as positively skewed clustering. Thus, the data for these analyses deviate somewhat from a normal distribution warranting caution for interpreting the significance of inferential test statistics.

A collinearity analysis was conducted to examine any problematic correlations between predictor variables. In accordance with Denis (2016), which stated that a VIF score of 10 suggests that a study's parameter β was not being precisely estimated due to a large standard error, the present study used a VIF cutoff score of 10. VIF scores for all variables passed this cutoff for both regression analyses (all VIFs < 1.93).

Prior to computing the regression analyses, binary Pearson r correlations were computed to examine the relationships between each variable (see Table 3). Contrary to the first hypothesis, results revealed that the relationships between community participation and the following variables were both statistically significant ($ps < .001$) and positive in direction: MSE, difficulty completing self-care tasks, and difficulty remembering/concentrating. Working memory ability was negatively correlated with community participation ($p < .001$).

Community Participation

A multiple linear regression analysis was conducted to examine how MSE, difficulty completing self-care tasks, difficulty with remembering/concentrating, and working memory ability are associated with community participation. Entering all variables into the equation explained 45% of the variance in community participation (R2 = .45, Adj. R2 = .42, F(7, 141) = 16.26, p < .001; see Table 4). Greater MSE was positively related to more community participation (β = .38, p < .001), as predicted. However, contrary to the first hypothesis, worse working memory ability (β = -.22, p = .001) and having greater difficulty remembering/concentrating (β = .26, p < .001) predicted higher levels of community participation.

Memory Self-Efficacy

A multiple linear regression analysis was conducted to examine how difficulty completing self-care tasks, difficulty with remembering/concentrating, and noticing memory changes are associated with MSE. Entering all variables explained 13% of the variance in MSE, R2 = .13, Adj. R2 = .09, F(7, 140) = 2.95, p = .007; see Table 5). Contrary to the second hypothesis, greater difficulty completing self-care tasks (β = 0.24, p = .02), and gender (β = -0.22, p = .01) predicted higher levels of MSE.

Discussion

Findings

Consistent with the first hypothesis, greater MSE was positively associated with community participation. This finding might suggest that one's beliefs about their memory can influence whether they choose to participate in community activities or not. For example, if an individual does not think that their memory ability is strong, especially compared to their friends or to others they might interact with, they might choose to stay home to avoid embarrassment or the stress of attending the event. However, contrary to the first hypothesis, worse working memory ability and greater difficulty remembering/concentrating also predicted higher levels of community participation. This is surprising, as the literature has proposed that those experiencing cognitive difficulty are more likely to experience social isolation (DiNapoli et al., 2014; Shankar et al., 2013).

The findings of the present study might suggest that individuals who experience greater cognitive difficulty might tend to seek out alternative, pleasurable activities or activities that might enhance their social support, which is contrary to the idea that this increased difficulty would lead individuals to stay at home due to the extra effort it takes to leave their home. Further, an explanation might be found in Greenglass et al. (2006), which examined how proactive coping — a form of coping in which an individual views stressors as challenges rather than threats — was related to greater functional independence and lower depression in older adults. It is possible that older adults with difficulty view community activities as challenges, motivating them to participate in order to achieve and maintain independence (Greenglass et al., 2006)

With regard to MSE, a surprising finding that contradicted the second hypothesis was that greater difficulty completing self-care tasks was found to predict higher levels of MSE. It is possible that this study's results were an example of the role that social desirability can play in MSE and memory performance, as explored by Sawin (2021). Social desirability refers to one's desire to minimize negative and enhance positive attributes of themselves, which can result in underreporting negative behaviors while overreporting positive behaviors (Latkin et al., 2017; Sawin, 2021). In these cases, as potentially suggested by the present study's data, one might overreport their memory ability in an effort to conceal their experience of difficulty completing self-care tasks. This would also provide an explanation as to why working memory ability was negatively related to community participation, as it is possible that memory ability is not as influential as social desirability is. If correct, this idea would contrast with the literature that examines the threat of stereotypes associated with aging (e.g., older adults are not as cognitively able strictly due to their age), on an individual's abilities on memory tasks, such that ability is reduced due to buy-in of these stereotypes (Chasteen et al., 2005; Stein et al., 2002). If older adults instead choose to overreport their memory ability, attempting to appear better off than they are, then this social desirability factor might have a greater effect than stereotypes associated with age or difficulty completing self-care tasks could have on ability..

Further, our findings illustrated that men had higher MSE compared to women. This was similar to the findings of Fallan and Opstad (2016) and Huang (2012), which observed that males reported higher self-efficacy on math tasks. It is possible that the working memory task was considered by participants to be a math task because it involved numbers, which could help explain the gender differences revealed in the present study.

Finally, given past research findings that MSE was positively related to working memory, it was surprising that MSE was not positively related to working memory. It is possible that the executive functioning component of working memory caused the relationship between MSE and memory that was observed in past research to not be true of working memory. It is also possible that the present study's limitations, which are discussed in detail below, led to a positive relationship between MSE and working memory remaining undetected.

Limitations and Suggestions for Future Research

The present study was subject to three primary limitations: 1) a lack of diversity in the sample, 2) recruiting participants via MTurk, 3) participants were asked to self-report their levels of community participation.

Speaking to the first limitation, most participants identified as female (66.8%), Caucasian (65.6%), and achieved higher than a high school education (91.5%). This lack of variation could decrease the external validity of this study's findings. Future research should retest these hypotheses with a larger and more diverse sample to increase the ability to generalize results.

Second, all participants were recruited using MTurk, which is an online survey platform created by Amazon. This could have limited the external validity of the findings, as participating in the present study would have required technical skills to get on MTurk, as well as awareness about MTurk's monetary incentives in exchange for participation in research studies. This could particularly affect the population that is being examined in this study, as older adults are less likely to possess technological skills and be aware of MTurk, compared to younger populations. It is possible that the older adults who participated in this study possess characteristics that might serve as a latent, confounding variable that might have affected the present study's findings, although there is evidence, using the same method and population sample as the present study, to suggest that the memory performance of an MTurk sample of older adults does not differ from a normative, traditional sample of older adults (Mashinchi et al., 2021). Future research should use one sample to conduct data collection through two methods: 1) via MTurk data collection, and 2) via in-person data collection, and then compare the findings to determine if there are differences in the experiences of participants due to the difference in data collection method.

Further, the age filters that MTurk allows are preset and unmodifiable. Fifty-five and older is the oldest age grouping that can be selected. This means that those 55-64 years old were included in the study. There is evidence that this age range is when subjective memory complaints begin (Jenkins et al., 2019), which would result in a low reported MSE, and thus would be an important age sample to include in this study. Despite this evidence, the 55-64 age grouping is not often subject to concerns about cognitive decline (Aartsen et al., 2002). It is possible that this minimum age could have negatively skewed the results from the present study and might serve as an explanation as to why a positive relationship between MSE and working memory was not found. Similarly, the median age of participants was 65 years old, which is on the younger end of the age range, and might have led to an inability to detect a relationship between MSE and working memory. Future research should seek to recruit older participants and compare findings to the present study in order to determine if there is a point at which MSE, working memory ability, or community participation changes (e.g., comparing 65-year-old scores to 85-year-old scores). Given that the human population is getting older (Crimmins, 2015; Semenova & Stadtlander, 2016), this distinction will be important to determine.

Additionally, it is important to note that 23 (6.5%) participants were excluded because their reported age was younger than 55. Although the MTurk filter was used to recruit only those 55 and older, it is possible that some participants have found ways to bypass the filter in order to complete tasks for monetary incentives. Given these issues, the authors of the present study echo the recommendation made by Chmielewski and Kucker (2019) to screen data for completion, validity, and reliability prior to conducting analyses. Researchers using MTurk might need to plan to recruit more participants that an a priori power analysis suggests are needed for the study.

Third, the community participation data for this study was collected using a self-report measure. This option could allow for participants to erroneously recall how often they have engaged in community activities, either by over or underestimating. Future research should include measures beyond self-re port, such as asking participants to receive a stamp or initial on a card every time they participate in a community event. It is also possible that variables not accounted for in the present study, such as the geographic location and the socioeconomic status of both individuals and the community, might be related to community participation. Future research should seek to include these variables in their explorations.

In addition to future research suggestions to address limitations, future research should also further examine how proactive coping might have explained the present study's results and determine whether older adults with difficulties view community activities as motivated challenges to secure independence, as suggested by Greenglass et al. (2006).

Conclusion

Based on the findings of the present study, greater MSE is positively related to participation in community activities. This study serves as a first step to understanding how one's beliefs in their memory can be related to their decision of engaging in their community.

References

Aartsen, M. J., Smits, C. H. M., van Tilburg, T., Knipscheer, K. C. P. M., & Deeg, D. J. H. (2002). Activity in older adults: Cause or consequence of cognitive functioning? A longitudinal study on everyday activities and cognitive performance in older adults. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 57(2),* P153–P162. https://doi.org/10.1093/geronb/57.2.P153

Aben, B., Stapert, S., & Blokland, A. (2012). About the distinction between working memory and short-term memory. *Frontiers in Psychology, 3.* https://doi.org/10.3389/fpsyg.2012.00301

Bandura, A. (1989). Regulation of cognitive processes through perceived self-efficacy. *Developmental Psychology, 25(5),* 729–735. https://doi.org/10.1037/0012-1649.25.5.729

Bandura, A. (1997). *Self-efficacy: The exercise of control.* W.H. Freeman.

Beaudoin, M., & Desrichard, O. (2011). Are memory self-efficacy and memory performance related? A meta-analysis. *Psychological Bulletin, 137(2),* 211–241. https://doi.org/10.1037/a0022106

Berry, J. M., West, R. L., & Dennehey, D. M. (1989). Reliability and validity of the Memory Self-Efficacy Questionnaire. *Developmental Psychology, 25(5),* 701–713. https://doi.org/10.1037/0012-1649.25.5.701

Centers for Disease Control and Prevention [CDC]. (2015). *Census questions on disability endorsed by the Washington Group.* Centers for Disease Control and Prevention. https://www.cdc.gov/nchs/data/washington_group/WG_Short_Measure_on_Disability.pdf

Chasteen, A. L., Bhattacharyya, S., Horhota, M., Tam, R., & Hasher, L. (2005). How feelings of stereotype threat influence older adults' memory performance. *Experimental Aging Research, 31(3),* 235–260. https://doi.org/10.1080/03610730590948177

Chmielewski, M., & Kucker, S. C. (2020). An MTurk crisis? Shifts in data quality and the impact on study results. *Social Psychological and Personality Science, 11(4),* 464–473. https://doi.org/10.1177/1948550619875149

Cowan, N. (2008). Chapter 20: What are the differences between long-term, short-term, and working memory? *In Progress in Brain Research (Vol. 169,* pp. 323–338). Elsevier. https://doi.org/10.1016/S0079-6123(07)00020-9

Crimmins, E. M. (2015). Lifespan and healthspan: Past, present, and promise. *The Gerontologist, 55(6),* 901–911. https://doi.org/10.1093/geront/gnv130

Denis, D. J. (2016). *Applied univariate, bivariate, and multivariate statistics.* Wiley.

Desrichard, O., & Köpetz, C. (2005). A threat in the elder: The impact of task-instructions, self-efficacy and performance expectations on memory performance in the elderly. *European Journal of Social Psychology, 35(4),* 537–552. https://doi.org/10.1002/ejsp.249

DiNapoli, E. A., Wu, B., & Scogin, F. (2014). Social isolation and cognitive function in Appalachian older adults. *Research on Aging, 36(2),* 161–179. https://doi.org/10.1177/0164027512470704

Fallan, L., & Opstad, L. (2016). Student self-efficacy and gender-personality interactions. International *Journal of Higher Education, 5(3),* p32. https://doi.org/10.5430/ijhe.v5n3p32

Fratiglioni, L., Paillard-Borg, S., & Winblad, B. (2004). An active and socially integrated lifestyle in late life might protect against dementia. *The Lancet Neurology, 3(6),* 343–353. https://doi.org/10.1016/S1474-4422(04)00767-7

Fritsch, T., Smyth, K. A., Debanne, S. M., Petot, G. J., & Friedland, R. P. (2005). Participation in novelty-seeking leisure activities and Alzheimer's disease. *Journal of Geriatric Psychiatry and Neurology, 18(3),* 134–141. https://doi.org/10.1177/0891988705277537

Greenglass, E., Fiksenbaum, L., & Eaton, J. (2006). The relationship between coping, social support, functional disability and depression in the elderly. Anxiety, *Stress & Coping, 19(1),* 15–31. https://doi.org/10.1080/14659890500436430

Hertzog, C., Dixon, R. A., Schulenberg, J. E., & Hultsch, D. F. (1987). On the differentiation of memory beliefs from memory knowledge: The factor structure of the metamemory in adulthood scale. *Experimental Aging Research, 13(2),* 101–107. https://doi.org/10.1080/03610738708259308

Hoffman, B., & Schraw, G. (2009). The influence of self-efficacy and working memory capacity on problem-solving efficiency. *Learning and Individual Differences, 19(1),* 91–100. https://doi.org/10.1016/j.lindif.2008.08.001

Huang, C. (2013). Gender differences in academic self-efficacy: A meta-analysis. *European Journal of Psychology of Education, 28(1)*, 1–35. https://doi.org/10.1007/s10212-011-0097-y

Insel, K., Morrow, D., Brewer, B., & Figueredo, A. (2006). Executive function, working memory, and medication adherence among older adults. The *Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 61(2),* P102–P107. https://doi.org/10.1093/geronb/61.2.P102

Lalitha, K. & Aswartha Reddy, A. (2021). Everyday memory self-efficacy and self-reported health among elderly persons. *Indian Journal of Gerontology, 35(3),* 360–374.

Latkin, C. A., Edwards, C., Davey-Rothwell, M. A., & Tobin, K. E. (2017). The relationship between social desirability bias and self-reports of health, substance use, and social network factors among urban substance users in Baltimore, Maryland. *Addictive Behaviors, 73,* 133–136. https://doi.org/10.1016/j.addbeh.2017.05.005

Mars, G. M. J., Kempen, G. I. J. M., Post, M. W. M., Proot, I. M., Mesters, I., & van Eijk, J. T. M. (2009). The Maastricht social participation profile: Development and clinimetric properties in older adults with a chronic physical illness. *Quality of Life Research, 18(9),* 1207–1218. https://doi.org/10.1007/s11136-009-9537-4

Mashinchi, G. M., Cotter, K. A., & Williams, G. A. (2021). Creating an online digit span using Qualtrics: Ingenuity during the COVID-19 pandemic. *Internal University of Montana Report: Unpublished.*

Mashinchi, G. M. (2022). An examination of the relation between memory self-efficacy and working memory within the cognitive reserve framework. *University of Montana: Graduate Student Theses, Dissertations, & Professional Papers.* https://scholarworks.umt.edu/etd/11904/

McCabe, D. P., Roediger, H. L., McDaniel, M. A., Balota, D. A., & Hambrick, D. Z. (2010). The relationship between working memory capacity and executive functioning: Evidence for a common executive attention construct. *Neuropsychology, 24(2),* 222–243. https://doi.org/10.1037/a0017619

Meyer, M. L., Spunt, R. P., Berkman, E. T., Taylor, S. E., & Lieberman, M. D. (2012). Evidence for social working memory from a parametric functional MRI study. *Proceedings of the National Academy of Sciences, 109(6),* 1883–1888. https://doi.org/10.1073/pnas.1121077109

Miyake, A., & Shah, P. (Eds.). (1999). *Models of working memory: Mechanisms of active maintenance and executive control (1st ed.).* Cambridge University Press. https://doi.org/10.1017/CBO9781139174909

Nieboer, A. P., Hajema, K., & Cramm, J. M. (2020). Relationships of self-management abilities to loneliness among older people: A cross-sectional study. *BMC Geriatrics, 20(1),* 184. https://doi.org/10.1186/s12877-020-01584-x

Pearman, A., & Trujillo, A. (2013). Changes in memory prediction accuracy: Age and performance effects. *Educational Gerontology, 39(7),* 467–475. https://doi.org/10.1080/03601277.2012.701132

Sawin, K. G. (2021). Painting a pretty picture: the role of social desirability in the memory self-efficacy of young and older adults. B*utler Journal of Undergraduate Research, 7,* 272–286.

Scarmeas, N., Levy, G., Tang, M.-X., Manly, J., & Stern, Y. (2001). Influence of leisure activity on the incidence of Alzheimer's Disease. N*eurology, 57(12),* 2236–2242. https://doi.org/10.1212/WNL.57.12.2236

Semenova, V., & Stadtlander, L. (2016). Death anxiety, depression, and coping in family caregivers. *Journal of Social, Behavioral, and Health Sciences, 10(1).* https://doi.org/10.5590/JSBHS.2016.10.1.05

Shankar, A., Hamer, M., McMunn, A., & Steptoe, A. (2013). Social isolation and loneliness: relationships with cognitive function during 4 years of follow-up in the English Longitudinal Study of Ageing. *Psychosomatic Medicine, 75(2),* 161–170. https://doi.org/10.1097/PSY.0b013e31827f09cd

Sobral, M., & Paúl, C. (2013). Education, leisure activities and cognitive and functional ability of Alzheimer's disease patients: A follow-up study. *Dementia & Neuropsychologia, 7(2),* 181–189. https://doi.org/10.1590/S1980-57642013DN70200008

Stein, R., Blanchard-Fields, F., & Hertzog, C. (2002). The effects of age-stereotype priming on the memory performance of older adults. E*xperimental Aging Research, 28(2)*, 169–181. https://doi.org/10.1080/03610730252800184

Turvey, C. L., Schultz, S., Arndt, S., Wallace, R. B., & Herzog, R. (2000). Memory complaint in a community sample aged 70 and older. *Journal of the American Geriatrics Society, 48(11),* 1435–1441. https://doi.org/10.1111/j.1532-5415.2000.tb02634.x

Vallet, F., Agrigoroaei, S., Beaudoin, M., Fournet, N., Paignon, A., Roulin, J.-L., & Desrichard, O. (2015). Older adults' beliefs about forgetting and aging predict memory self-efficacy above and beyond actual memory performance and mental health. *Revue Internationale de Psychologie Sociale, 28(4)*, 57–79.

Vellone, E., Pancani, L., Greco, A., Steca, P., & Riegel, B. (2016). Self-care confidence may be more important than cognition to influence self-care behaviors in adults with heart failure: Testing a mediation model. *International Journal of Nursing Studies, 60,* 191–199. https://doi.org/10.1016/j.ijnurstu.2016.04.016

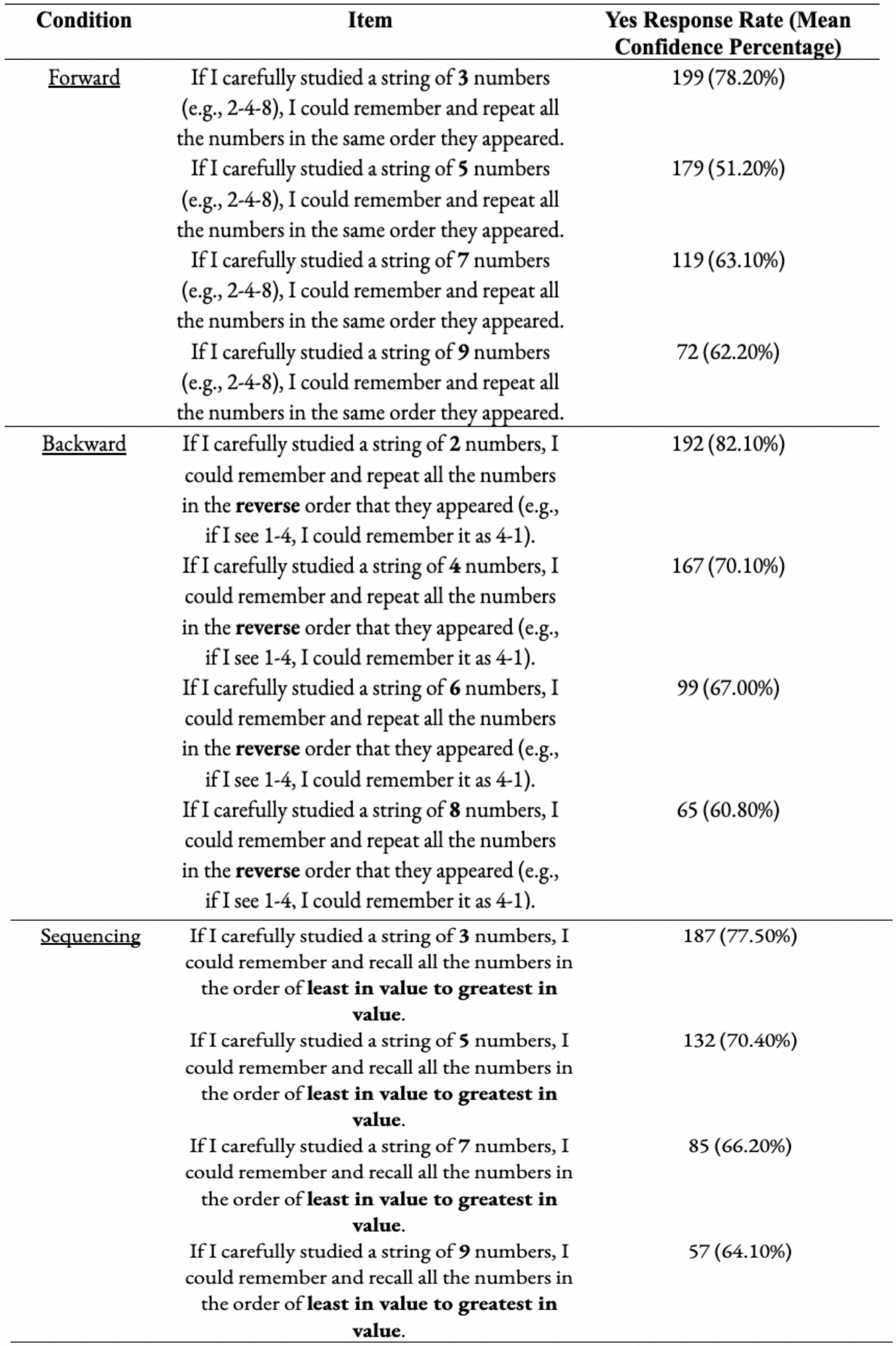
Verghese, J., Lipton, R. B., Katz, M. J., Hall, C. B., Derby, C. A., Kuslansky, G., Ambrose, A. F., Sliwinski, M., & Buschke, H. (2003). Leisure activities and the risk of dementia in the elderly. *New England Journal of Medicine, 348(25)*, 2508–2516. https://doi.org/10.1056/NEJMoa022252

Wechsler, D. (2008). *WAIS-IV technical and interpretive manual.* Pearson/PsychCorp.

West, R. L., Welch, D. C., & Knabb, P. D. (2002). Gender and aging: Spatial self-efficacy and location recall. *Basic and Applied Social Psychology, 24(1),* 71–80. https://doi.org/10.1207/S15324834BASP2401_7

Table 1

Memory Self-Efficacy Questionnaire Items



Condition	Item	Yes Response Rate (Mean Confidence Percentage)
Forward	If I carefully studied a string of **3** numbers (e.g., 2-4-8), I could remember and repeat all the numbers in the same order they appeared.	199 (78.20%)
	If I carefully studied a string of **5** numbers (e.g., 2-4-8), I could remember and repeat all the numbers in the same order they appeared.	179 (51.20%)
	If I carefully studied a string of **7** numbers (e.g., 2-4-8), I could remember and repeat all the numbers in the same order they appeared.	119 (63.10%)
	If I carefully studied a string of **9** numbers (e.g., 2-4-8), I could remember and repeat all the numbers in the same order they appeared.	72 (62.20%)
Backward	If I carefully studied a string of **2** numbers, I could remember and repeat all the numbers in the **reverse** order that they appeared (e.g., if I see 1-4, I could remember it as 4-1).	192 (82.10%)
	If I carefully studied a string of **4** numbers, I could remember and repeat all the numbers in the **reverse** order that they appeared (e.g., if I see 1-4, I could remember it as 4-1).	167 (70.10%)
	If I carefully studied a string of **6** numbers, I could remember and repeat all the numbers in the **reverse** order that they appeared (e.g., if I see 1-4, I could remember it as 4-1).	99 (67.00%)
	If I carefully studied a string of **8** numbers, I could remember and repeat all the numbers in the **reverse** order that they appeared (e.g., if I see 1-4, I could remember it as 4-1).	65 (60.80%)
Sequencing	If I carefully studied a string of **3** numbers, I could remember and recall all the numbers in the order of **least in value to greatest in value**.	187 (77.50%)
	If I carefully studied a string of **5** numbers, I could remember and recall all the numbers in the order of **least in value to greatest in value**.	132 (70.40%)
	If I carefully studied a string of **7** numbers, I could remember and recall all the numbers in the order of **least in value to greatest in value**.	85 (66.20%)
	If I carefully studied a string of **9** numbers, I could remember and recall all the numbers in the order of **least in value to greatest in value**.	57 (64.10%)

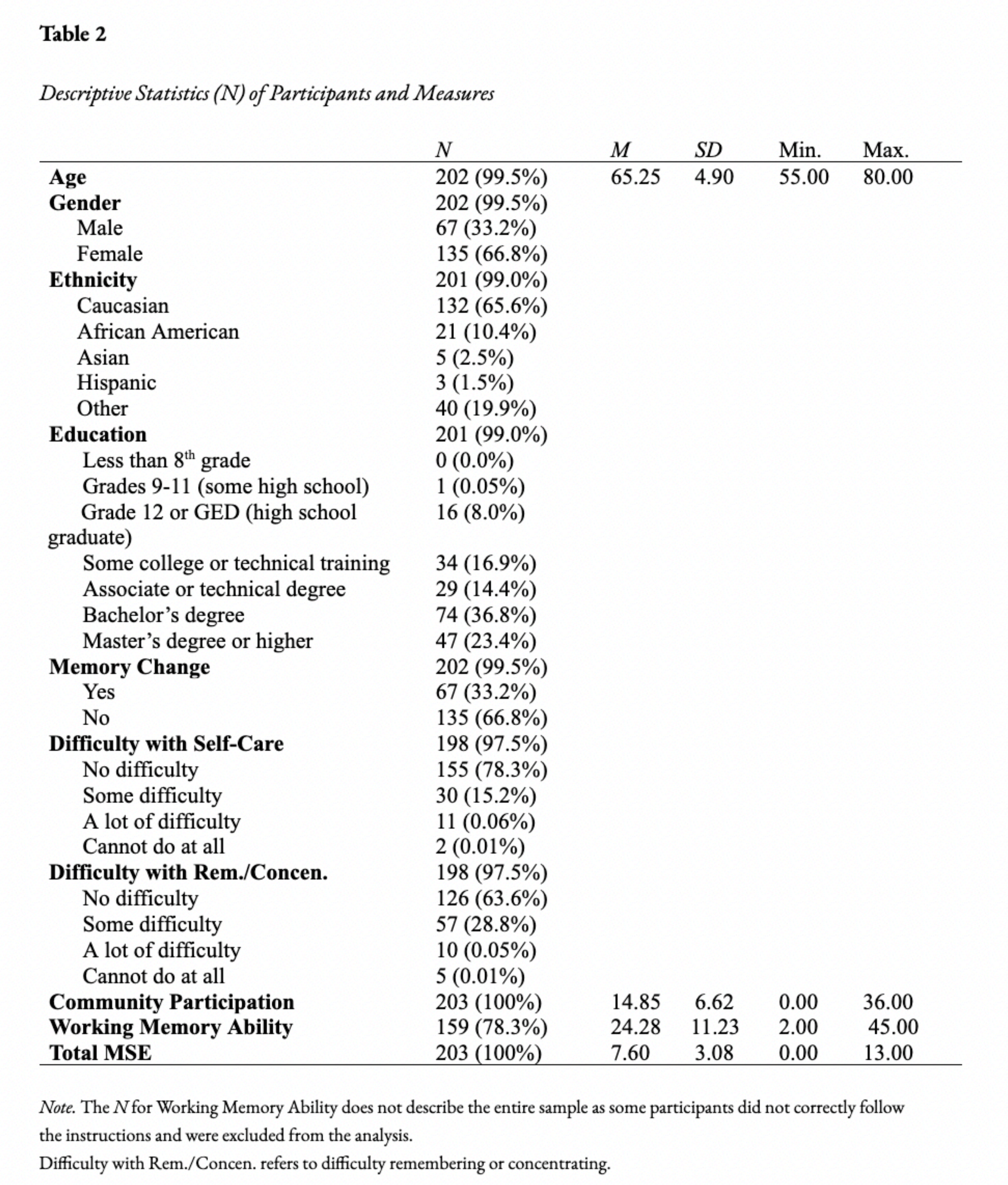


Table 2

Descriptive Statistics (N) of Participants and Measures

	N	*M*	*SD*	Min.	Max.
Age	202 (99.5%)	65.25	4.90	55.00	80.00
Gender	202 (99.5%)				
Male	67 (33.2%)				
Female	135 (66.8%)				
Ethnicity	201 (99.0%)				
Caucasian	132 (65.6%)				
African American	21 (10.4%)				
Asian	5 (2.5%)				
Hispanic	3 (1.5%)				
Other	40 (19.9%)				
Education	201 (99.0%)				
Less than 8th grade	0 (0.0%)				
Grades 9-11 (some high school)	1 (0.05%)				
Grade 12 or GED (high school graduate)	16 (8.0%)				
Some college or technical training	34 (16.9%)				
Associate or technical degree	29 (14.4%)				
Bachelor's degree	74 (36.8%)				
Master's degree or higher	47 (23.4%)				
Memory Change	202 (99.5%)				
Yes	67 (33.2%)				
No	135 (66.8%)				
Difficulty with Self-Care	198 (97.5%)				
No difficulty	155 (78.3%)				
Some difficulty	30 (15.2%)				
A lot of difficulty	11 (0.06%)				
Cannot do at all	2 (0.01%)				
Difficulty with Rem./Concen.	198 (97.5%)				
No difficulty	126 (63.6%)				
Some difficulty	57 (28.8%)				
A lot of difficulty	10 (0.05%)				
Cannot do at all	5 (0.01%)				
Community Participation	203 (100%)	14.85	6.62	0.00	36.00
Working Memory Ability	159 (78.3%)	24.28	11.23	2.00	45.00
Total MSE	203 (100%)	7.60	3.08	0.00	13.00

Note. The *N* for Working Memory Ability does not describe the entire sample as some participants did not correctly follow the instructions and were excluded from the analysis.

Difficulty with Rem./Concen. refers to difficulty remembering or concentrating.

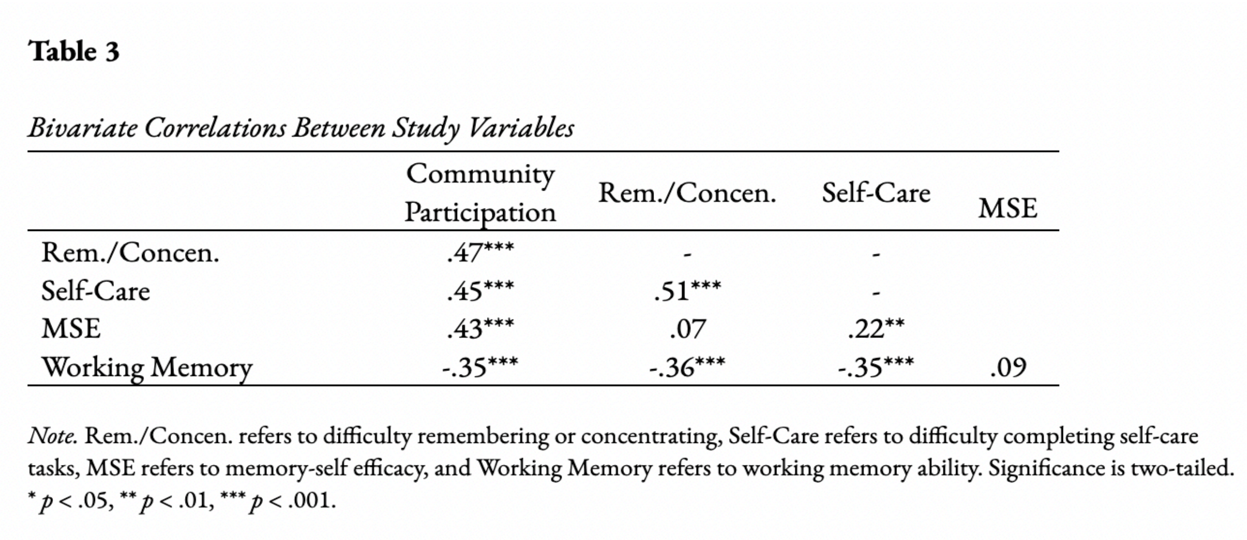


Table 3

Bivariate Correlations Between Study Variables

	Community Participation	Rem./Concen.	Self-Care	MSE
Rem./Concen.	.47***	-	-	
Self-Care	.45***	.51***	-	
MSE	.43***	.07	.22**	
Working Memory	-.35***	-.36***	-.35***	.09

Note. Rem./Concen. refers to difficulty remembering or concentrating, Self-Care refers to difficulty completing self-care tasks, MSE refers to memory-self efficacy, and Working Memory refers to working memory ability. Significance is two-tailed. * $p < .05$, ** $p < .01$, *** $p < .001$.

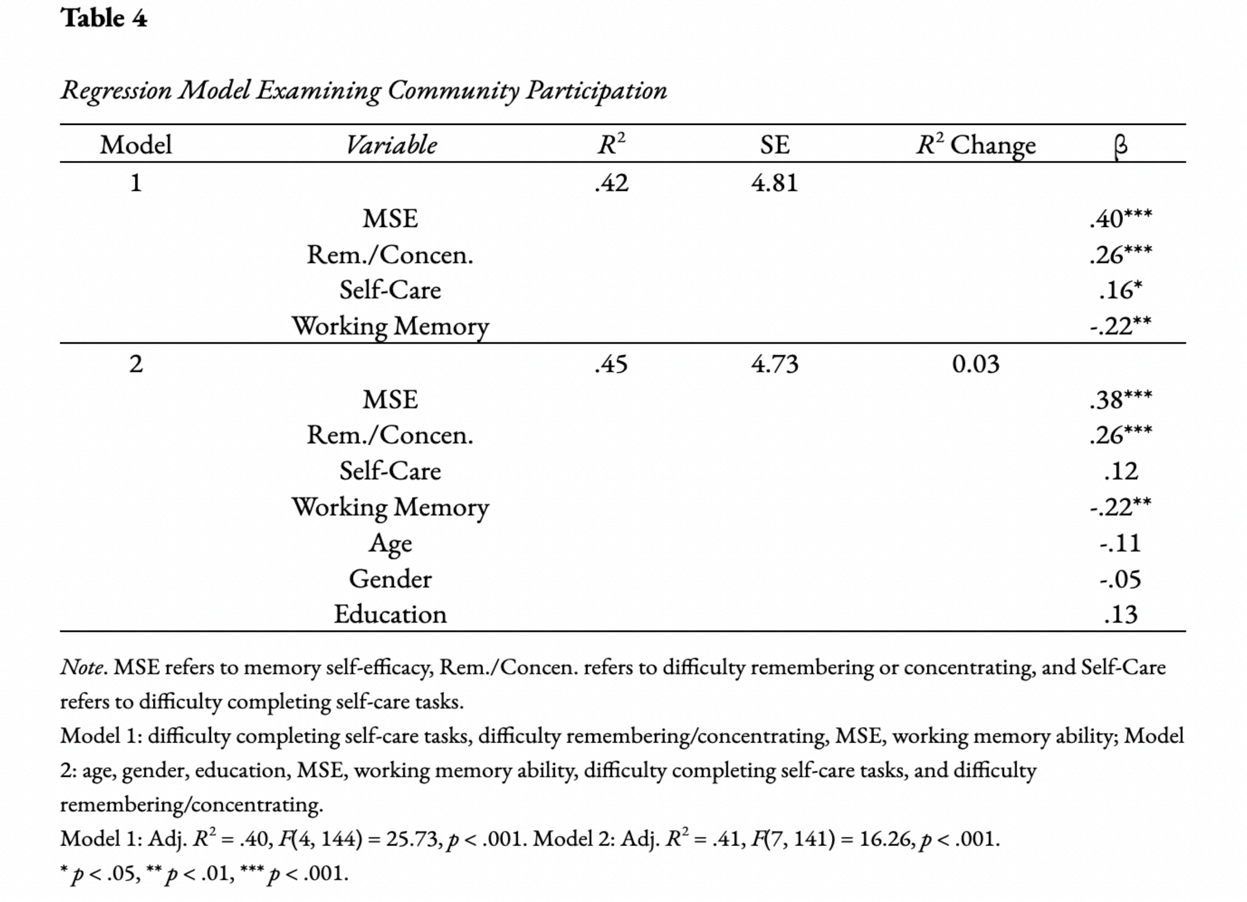


Table 4

Regression Model Examining Community Participation

Model	*Variable*	R^2	SE	R^2 Change	β
1		.42	4.81		
	MSE				.40***
	Rem./Concen.				.26***
	Self-Care				.16*
	Working Memory				-.22**
2		.45	4.73	0.03	
	MSE				.38***
	Rem./Concen.				.26***
	Self-Care				.12
	Working Memory				-.22**
	Age				-.11
	Gender				-.05
	Education				.13

Note. MSE refers to memory self-efficacy, Rem./Concen. refers to difficulty remembering or concentrating, and Self-Care refers to difficulty completing self-care tasks.

Model 1: difficulty completing self-care tasks, difficulty remembering/concentrating, MSE, working memory ability; Model 2: age, gender, education, MSE, working memory ability, difficulty completing self-care tasks, and difficulty remembering/concentrating.

Model 1: Adj. $R^2 = .40$, $F(4, 144) = 25.73$, $p < .001$. Model 2: Adj. $R^2 = .41$, $F(7, 141) = 16.26$, $p < .001$.

* $p < .05$, ** $p < .01$, *** $p < .001$.

Table 5

Regression Model Examining Memory Self-Efficacy

Model	*Variable*	R^2	SE	R^2 Change	β
1		.08	3.04		
	Memory Changes				.001
	Rem./Concen.				-.14
	Self-Care				.30**
	Working Memory				.17
2		.13	2.98	.05	
	Memory Changes				-.02
	Rem./Concen.				-.15
	Self-Care				.24*
	Working Memory				.16
	Age				-.05
	Gender				-.22**
	Education				.05

Note. Memory Changes refers to noticing memory changes, Rem./Concen. refers to difficulty remembering or concentrating, and Self-Care refers to difficulty completing self-care tasks.

Model 1: difficulty completing self-care tasks, difficulty remembering/concentrating, noticing memory changes, and working memory ability; Model 2: age, gender, education, working memory ability, memory changes, difficulty completing self-care tasks, and difficulty remembering/concentrating.

Model 1: Adj. $R^2 = .05$, $F(4, 143) = 2.98$, $p = .02$. Model 2: Adj. $R^2 = .11$, $F(7, 140) = 2.95$, $p = .01$.

* $p < .05$, ** $p < .01$.

Graduate Student Journal of Psychology
2022, Vol. 19

Mask-wearing and Facial Emotion Recognition: A Preliminary Analysis of the Relevance of Depressive Symptoms

Mateus Morante Mazzaferro,
Teachers College, Columbia University

Depression is a psychiatric condition that impairs the life of millions of people around the globe. Previous research has shown that depressed individuals tend to present deficits in facial emotion perception. For instance, perception accuracy may be reduced and biases in perceived intensity may be enhanced. Mask-wearing practices initiated in response to the COVID-19 pandemic have become a new social norm often enforced by local mandates. Preliminary studies have shown that mask-wearing may impair facial emotion recognition. In this study, we aimed at understanding how facial emotion recognition impairment interacts with depressive symptoms in a sample of German adolescents and adults (N = 91, 56% female, mean age of 32.7 years) by utilizing a mixed-effects linear regression analysis. We found evidence that mask-wearing may be a limiting factor for facial emotion recognition, as well as for emotion intensity ratings. However, a significant association of depressive symptoms with these outcome variables was not detected. Still, larger sample sizes may have the potential to substantiate a trend toward an interaction of depressive symptoms and mask-wearing for the rating of happy faces. Future research should be committed to psychophysiological processes and to improving the quality of the stimulus material.

Theoretical Background

Depression and Facial Emotion Recognition

In 2017, the WHO issued a report estimating that, globally, over 300 million people suffer from depression, equivalent to 4.4% of the world's population (WHO, 2017). A review from 2014 shows that the average 12-month prevalence estimate of DSM-IV major depressive episodes varies from 5.5% to 5.9% in most countries (Kessler et al., 2014). In the U.S. alone, around 30 million adults have met the criteria for major depressive disorder at least once in their lifetime (Haro et al., 2006). The core symptoms of clinically relevant depression include first and foremost a persistently depressed mood and loss of interest or pleasure in almost all activities (APA, 2013). In addition, there are several other affective, cognitive, somatic, and motivational-behavioral features that play a role in depression, such as sleep disturbances, a sense of guilt, and impaired executive function (Berking & Rief, 2011). The presentation of each symptom can vary greatly in nature and intensity so the overall picture can differ considerably from case to case.

Early theories of depression, such as Beck's (1976) schema and Bower's (1981) network theory, proposed that, in depressive patients, cognitive biases operate throughout information processing, including perception, attention, and memory. Beck (1976) proposes that those with depression develop cognitive distortions and tend to overlook positive attributes of reality and selectively attend to the negative. The associative network theory of emotion and memory, outlined by Bower (1981), suggests that depressed mood leads to biases favoring the perception of mood-congruent information. Both models assume that affected individuals selectively process information related to sadness, loss, and failure. Furthermore, poor interpersonal relationships have been proposed as a critical aspect in both the etiology and maintenance of depression (Finch & Zautra, 1992; Platt et al. 2013). Impaired emotion recognition may contribute to the interpersonal difficulties and avoidance behavior seen in depressed patients (Persad & Polivy, 1993). Since deficits in emotion recognition could contribute to the maintenance of depressive symptoms, investigating the relationship between these two variables has important implications for both existing and the development of novel cognitive-behavioral interventions.

For a long time now, research has shown that people suffering from affective disorders have difficulties correctly interpreting human emotions from the perception of facial expressions. In adults, the facial emotion recognition (FER) impairments have been associated, for example, with bipolar disorder (Derntl et al. 2009) and also very commonly with depression (e.g., Demenescu et al., 2010; Bistricky et al 2011, Rubinow & Post, 1992). However, findings in this area seem to vary strongly based on symptom severity, psychiatric comorbidity, and the nature of stimuli (Bistricky et al., 2011). A meta-analysis conducted by Dalili et al. (2015) using 22 independent samples found significant recognition deficits in depressed participants in all basic emotions except sadness. But the main feature of facial emotion recognition in depression is not a deficit in the ability to identify emotions per se, it is rather a bias in emotional valence rating. Many studies have reported a negative bias in depressed patients (e.g., Schmid & Schmid Mast, 2010; Milders et al., 2010, Kan et al., 2004), which means the affected populations sometimes perform better in recognizing sad faces, tend to interpret neutral or ambiguous faces as sad more often and label happy faces as less happy.

The literature presents similar evidence for adolescents and younger children. Nyquist & Luebbe, (2020) reviewed 26 studies and identified three trends for FER in youth that were either depressed or qualified as high-risk populations for depression: sensitivity to sadness (higher accuracy in recognizing sadness and biased rating towards sad faces), over-perception of anger (falsely recognizing ambiguous or neutral faces as angry), and under-perception of happiness (lower accuracy in recognizing happiness and biased rating towards happiness). Investigating features of depression in families, Lopez-Duran et al. (2013) observed sensitivity to sadness as a potential mechanism of risk among boys at familial risk for depression. In a similar vein, Kluczniok et al. (2015) found that children from mothers with remitted depression displayed depressive-like emotion recognition bias that correlated with their mothers' performance. As suggested by the literature, emotion recognition bias is a common correlate of depressive symptomatology in all age groups. It is, therefore, essential to investigate how these deficits can impair social functioning and contribute to the pathological behavior of clinically relevant populations.

Masks and Emotion Recognition

Since March of 2020, the COVID-19 pandemic has brought significant changes to the daily lives of everyday citizens. To help stop the spread of the virus, the wearing of a protective face mask, which has been shown to drastically reduce viral transmission (Cheng et al., 2020), has become a largely accepted norm (e.g., Cheng et al., 2020; Hong et al., 2020; Howard et al., 2020). Albeit necessary for keeping the virus under control, masks may have important psychological impacts on social interactions, for example, by muffling speech and other forms of communicative vocalization (Mheidly et al., 2020). Social distancing and mask-wearing can also impair interpersonal communication by limiting physical touch and body language, as well as by hindering visibility of the lower half of the face.

Regarding facial emotional recognition, recent studies have found that mask-wearing can make emotions less well recognized or interpreted (e.g., Carbon, 2020a; Carbon, 2020b), faces less well-remembered, or critically impair holistic processing (Freud et al., 2020). New data, nonetheless, shows that emotions can still be recognized both in adults (e.g., Calbi et al., 2021; Kastendieck, et al., 2021) and in children (e.g., Ruba & Pollak, 2020), even if the intensity of the emotion decreases (Kastendieck et al., 2021).

Studies also show that sad and angry faces cause more fixation on the eyes, while happy faces attract more fixation on the mouth, due to the way different emotions mobilize different facial musculature (Eisenbarth, 2011; Schurgin, 2014) - curiously, psychiatric patients seem to have less fixation overall (Eisenbarth, 2011). These facts may account for findings reporting that when only the upper part of the face is visible, participants perceive and recognize negative emotions, like anger and fear, better than positive ones (Marta et al., 2021; Fischer et al., 2012). In sum, it is possible that mask-wearing does not impair or even relatively increase the perception of negative facial expressions and, at the same time, diminishes the perception of positive ones (Spitzer, 2020). Therefore, it may have significant effects on day-to-day activities that rely on interpersonal communication and social interaction (Freud et al., 2020).

The factors that influence facial emotion recognition in laboratory settings remain unclear when individuals wear masks. The type of stimuli, for instance, is discussed as a possible moderator: so far, mainly adult faces have been used (e.g., Ruba & Pollak, 2020), whereas in real-life interactions children as well as adolescents and adults represent their counterpart. For this reason, emotional child and adult faces will be used in the current study. Moreover, static photos are mostly used (e.g., Carbon, 2020a, Carbon, 2020b; Ruba & Pollak, 2020), whereas closer proximity to reality can be assumed when videos are used (Rymarczyk et al., 2016). For that reason, this study is going to use video stimulus material.

Research Gap and Current Research Purpose

As to this point, we are unaware of any research projects trying to fill the newly created research gap of how mask-wearing interacts with psychopatholo gy in impairing social function. Considering that the COVID-19 pandemic has introduced a new norm in face-to-face interactions (which now almost always include the wearing of face masks), research on FER must rise to the challenge of understanding how this phenomenon can impact the nature of social exchange, especially for populations that already suffer from social difficulties, like a significant amount of patients diagnosed with mental disorders (Lehmann et al., 2019) – since these people may be particularly challenged by the pandemic (Druss, 2020). This preliminary study was conducted as a part of the *Gesichter lesen*[1] project, an initiative seeking to investigate how mask-wearing differentially impacts social exchange in children, adolescents, and adults. Here, we aim to understand how depressive symptoms interact with the effects of mask-wearing in facial emotion recognition processes in adolescents and adults.

In the future, the *Gesichter lesen* project intends to analyze data from multiple psychopathological variables and attitudes in emotion recognition and facial mimicry in children and adolescents. Considering that children's and early adolescents' development are largely influenced by the social context within which they interact (Richards & Light, 1986), the impact of mask-wearing might be particularly significant for them, notably if they struggle with any form of psychopathology. For this first experiment, we selected a more accessible sample of adults recruited online and decided to focus on a single psychopathology-related construct (depressive symptoms, continuous) and mask-wearing (binary) as independent variables. We aim to examine how the wearing of face masks interacts with the biases in facial emotion reading found to be associated with depressive symptoms in adolescents and adults (see above). The dependent variables of interest are, therefore, emotion recognition and emotion intensity rating.

Hypotheses

The first outcome being examined is emotion recognition. We hypothesize that participants' emotion recognition of anger, sadness, and happiness is negatively affected by the presence of a face mask (H1.1). In addition, recognition of anger, sadness, and happiness are associated with the participant's depressive symptoms (H1.2). In line with the presented literature, we expect sadness and anger to be overperceived and happiness to be underperceived by participants with elevated depressive symptoms. Finally, there is an interaction effect between the presence of a face mask and the participant's depressive symptoms in emotion recognition for anger, sadness, and happiness (H1.3).

For the second outcome, emotion intensity rating, we hypothesize that all emotions (anger, sadness, and happiness) are rated as less intense in masked faces than emotions in unmasked faces (H2.1). Additionally, the intensity rating of all perceived emotions (anger, sadness, and happiness) is associated with the participant's depressive symptoms (H2.2). In line with the presented literature, we expect sadness and anger to be rated more intensively and happiness to be rated less intensively by participants with elevated depressive symptoms. Lastly, there is an interaction effect between the presence of a face mask and the participant's depressive symptoms in the rating of emotion intensity for all emotions (anger, sadness, and happiness) (H2.3).

Methods

Sample. Inclusion criteria for the online experiment consisted of currently living in Germany, speaking fluent German, and being at least 14 years old. Recruiting was conducted between January and June 2021 through multiple channels, including Prolific[2], different social media platforms (Facebook and Twitter), and student recruiting resources from the Institute of Psychology at the Humboldt University Berlin, which consist of mailing lists and flyer distribution. Participants received 5 EUR on average for their participation, which is classified as good payment in Prolific. Compensation was carried out either via Prolific, bank deposit, or as test person points for undergraduate credit. A total of 108 participants (adults and adolescents 14 years and over) completed the online experiment and gave their consent via Sosci Survey. After the exclusion of subjects that did not complete psychopathology measures or dropped out before the end of the experiment, the final sample consisted of data from 91 participants (51 women, 56%) with a mean age of 32.7 years (SD = 15.6, range 14-71), who reported having seen the video stimuli and for whom at least one video upload

---

[1] For more information on the Gesichter lesen project, see https://www.kinderprojekte-psychologie.de/projekte/gesichter-lesen/

[2] For more information on the recruiting engine Prolific, see https://www.prolific.co/#how-it-works

was successful. We did not collect information on race, ethnicity, income level, or educational attainment.

Material. For the diagnostic material, we selected the PHQ-9 for the recording of depressive symptoms (Patient Health Questionnaire - 9; Richardson et al., 2010). The PHQ-9 is a versatile instrument for screening and measuring the severity of depression using only 9 items. It incorporates diagnostic criteria from the DSM-IV and other important major depressive symptoms into a short self-report tool (Kroenke et al., 2001). This instrument has been shown to have adequate internal consistency (>0.8; Kim & Lee, 2019; Titov et al., 2010), a one-factor replicable structure (Kim & Lee, 2019; Titov et al., 2010), and satisfactory convergent validity (Kim & Lee, 2019; Beard et al., 2016). In this study, the PHQ-9 had a Cronbach's alpha of 0.86. For a different set of analyses not relevant to the present study, the experiment materials also included the Inclusion of Other in the Self scale (IOS, Aron et al., 1992), the GAD-7 (Spitzer et al., 2006), and a short questionnaire on the participant's attitudes towards mask-wearing (An et al, 2021).

As video stimuli, we selected validated child and adult faces from the online Radbound Faces Database (Bijsterbosch et al., 2020), which were presented as videos through a morphing process created using FantaMorph5. Facial expressions ranged from neutral to emotional. Selected examples of the stimuli are presented in Figure 1. A surgical mask was added to half of the stimuli with the help of the video editing software Lens Studio by Snap Inc. The final videos were then composed with Adobe AfterEffects as follows: 1.5 seconds fixation cross, 5 seconds for the morph sequence, 1.5 seconds post-stimuli neutral background. For the sad videos, the morph sequence is 6 seconds long instead of 5 because research has shown that the onset and development of a sad facial expression take somewhat longer than happiness or anger (Fayolle & Droit-Volet, 2014). The pool of stimuli consists of 48 videos of 16 agents (8 adults, 8 children; 50% of all agents being female) showing three selected emotions (happiness, anger, and sadness). Half of all videos show agents wearing face masks. A superficial qualitative evaluation of how well a face mask would fit the morph was conducted to determine which videos would feature face masks. Each agent appears on 3 videos (one for each emotion), with either 1 out of 3 or 2 out of 3 videos being masked. The stimuli were randomized using a block design to prevent the same gender from being presented more than twice in a row.

Figure 1: On the left, a woman with a mask expressing happiness. On the right a child without a mask expressing anger.

Procedure. The study was conducted entirely online. For adolescents 14 years and over and adults, participation was completely self-sufficient, as approved by the Ethics Committee of the Humboldt University Berlin. Participants were informed that they would be required to watch 48 short videos, that the task would take about 40 minutes, and that participation in the experiment would only be possible if they had a webcam-enabled computer/laptop and agree to a webcam recording of their face during the experiment. Informed consent included standard details on compensation, confidentiality, and contact information. Participants who agreed to participate were instructed to set up their webcam to allow recording, to arrange sufficient lighting, and to refrain from eating or covering their face during the experiment.

The first set of questions asked for demographic information. Participants then watched the video stimuli while their facial activity was recorded. Following each video, using 7-point Likert-scales, participants rated the targets' emotion expressions using an emotion profile (happiness, sadness, fear, anger, disgust, and surprise) and were asked to indicate how close they felt to the person shown using the IOS Scale and how much they would like to meet the displayed agent. Following the video task, participants filled out psychopathology measures (PHQ-9 and GAD-7) and were asked about their attitudes toward mask-wearing and questioned about how the pandemic had affected their lives with the scale adapted from An et al. (2021). Finally, participants were given the opportunity to opt-out by requiring their videos to be erased (no participant used this option), informed about the purpose of the experiment, thanked for their participation, and received the necessary instructions for payment.

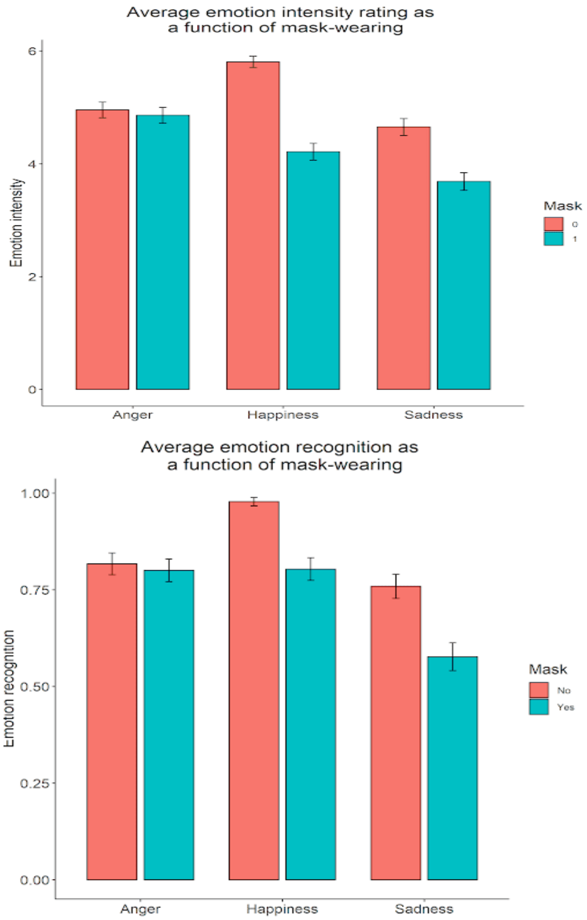
Statistical Analysis. Hypothesis for both outcome variables (emotion recognition and emotion intensity rating) were analyzed with linear mixed models (LMMs) using the statistical programming language R and its packages lme4 and lmerTest. The main effect of mask, depression score and the interaction between the two were estimated by random intercept/ random slope models. Cohen's d effect sizes were computed for every significant coefficient. The data structure characterized a repeated measures design since we collected many observations on the same variables from every participant. Thus, participant identification (ID) was included in the model as a cluster variable. Hence, the full model was computed as follows: Outcome ~ Mask * Depression + (1+Mask+Depression | ID).

Results

Emotion intensity rating. The main effect of the mask factor on emotion intensity rating can be observed in the descriptive bar plot in Figure 2. The mask factor (after controlling for depression score) yielded significant beta coefficients for happiness (β = -1.57, p < 0.001, Cohen's d = -0.41, CI95% = [-0.47; -0.35]), and for sadness (β = -0.99, p < 0.001, Cohen's d = -0.24, CI95% = [-0.28; -0.19]), but not for anger. The main effect for happiness, however, was qualified by the interaction mask*depression score and thus was interpreted in this context. These effects were also tested via likelihood-ratio test model comparison of intercept-only models and the results were maintained. Depression scores did not yield significant main effect regression coefficients for any of the selected emotions in the LMMs. It is important to mention that, for sadness, a non-significant small positive effect of depressive symptoms was observed (β = 0.04, p < 0.1, Cohen's d = 0.11, CI95% = [-0.02; -0.24]).

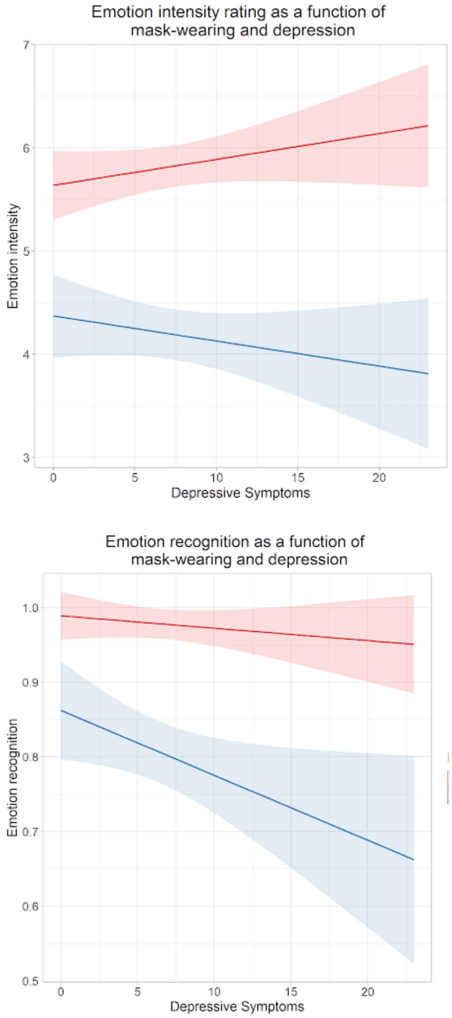
As for the interaction effects, the mask*depression score interaction was significant at the 5% level for happiness (β = -0.05, p < 0.05, Cohen's d = -0.07, CI95% = [-0.13; -0.01]) but not for sadness or anger. The significant interaction coefficients were plotted to allow a better interpretation of the effect and can be seen in Figure 5. The model for happiness, including the mask, depression scores, and the interaction between the two produced a marginal R2 of 0.174 (calculated according to Nakagawa & Schielzeth, 2013). A moderation analysis of depressive score in the mask effect was conducted with the help of simple slopes (with the package interactions in R) at three different points of the depression scale (-1SD, mean, +1SD). The mask effect was significant throughout the spectrum of depressive symptoms (p < 0.001). However, the 95% confidence intervals of the coefficients did overlap, that is, the simple slopes were not significantly different from one another. Thus, the moderation effect could not be interpreted as significant. Estimates for the effect of mask at the three respective levels of depression were β = -1.31 with CI95% = [-1.62; -1.00] at -1SD, β = -1.57 with CI95% = [-1.79; -1.36] at the mean and β = -1.84 with CI95% = [-2.14; -1.53] at +1SD

Emotion recognition. The main effect of the mask factor for emotion recognition can be observed in the descriptive bar plot in Figure 3. A hit (=1) was registered when the intensity rating of the target emotion was higher than the intensity rating of all other emotions. A miss (=0) was registered when one of the distractor emotions was rated higher than or equal to the target emotion. The second outcome variable, emotion recognition, produced similar results to emotion intensity ratings. The main effect of mask (after controlling for depression scores) was present for happiness (β = -0.17, p < 0.001, Cohen's d = -0.27, CI95% = [-0.34; -0.20]) and sadness (β = -0.19, p < 0.001, Cohen's d = -0.21, CI95% = [-0.26; -0.16]) but not for anger. Model comparison tests were consistent with this significance pattern. The variable depressive symptoms did not significantly predict the emotion recognition of any of the emotions displayed in this study.



Figures 2 (top) and 3 (bottom): Emotion intensity rating and recognition for each emotion as a function of mask. Error bars are 95% confidence intervals.

Interactions between mask and depression scores also failed to yield significant results. Still, it is worth mentioning that, in the case of happiness, a non-significant interaction effect consistent with the one found in the emotion intensity rating variable was reported (β = -0.006, p<0.1, Cohen's d = -0.05, CI95% = [-0.12; 0.01]), which is plotted in Figure 5. The model for emotion recognition of happiness including the mask, depression scores, and the interaction between the two produced a marginal R2 of 0.093 (calculated according to Nakagawa & Schielzeth, 2013). Moderation analysis through simple slopes was conducted for this outcome variable and showed similar results to the one of emotion intensity rating, meaning that the effect of the mask was significant throughout the depression scale and tended to increase as participants reported more symptoms, but 95% confidence intervals overlapped so that the difference between the simple slopes could not be interpreted as significant. Coefficients at the three respective points were β = -0.13 with CI95% = [-0.19; -0.07] at -1SD, β = -0.17 with CI95% = [-0.21; -0.12] at the mean and β = -0.20 with CI95% = [-0.26; -0.14] at +1SD.



Figures 4 (top) and 5 (bottom): Regression slopes of emotion intensity rating and emotion recognition of **happiness** by depression score grouped by mask (0: without mask, 1: with mask). Error bars are SDs.

Within the framework of the *Gesichter lesen* project, this study utilized morphed face stimuli to assess the effect of face masks and depressive symptoms on emotion recognition and emotion intensity rating.

The Effect of Mask-Wearing

In partial accordance with H1.1 and H2.1, the results suggest that mask-wearing has, according to Cohen's standards (Cohen, 1988), small to medium effects on both recognition and intensity rating of happy and sad faces, but no significant effects were found for angry faces. This pattern could, at least partially, be explained by the notion that, in comparison to happy and sad faces, the diagnostic face region of anger tends to be in the upper half of the observed face (Bassili, 1979). On a similar note, Calvo et al. (2018) evaluated eye movement patterns of participants looking at different facial emotions and found significantly more fixation on the eye region for angry faces. Thus, a face mask that covers only the lower region of the face may not impair the recognition of anger at all. However, it is important to point out that the fact that our study did not find a difference between the masked vs. unmasked condition for anger does not mean that this difference does not exist. In this sense, more research with larger samples is needed to further investigate the role of masks in the facial expression of anger.

Still, considering that the mask may, for example, also impair the understanding of speech (Mheidly et al., 2020), our findings for happiness and sadness may have important consequences for social interactions that include face masks. Social interaction partners may have to express their sadness or happiness in less ambiguous ways to avoid being misunderstood, for example, by speaking louder and incorporating gestures and body language when displaying these emotions. One type of interaction that could suffer considerably from these deficits is, for instance, a live psychotherapy session in which the client and the therapist are wearing face masks. Psychotherapists should pay attention to these impairments when trying to read the facial expression of patients and when expressing emotional reactions themselves. Failure in accounting for this shortfall may worsen interpersonal communication and hinder therapeutic progress.

The Relevance of Depressive Symptoms

The analysis of a possible main effect of depressive symptoms on emotion recognition and intensity rating as postulated by hypotheses H1.2 and H2.2 did not yield any significant results. This finding goes against some of the literature presented in the theoretical background of the present study. We could speculate in the following ways about the non-emergence of an effect. First, considering that we were expecting small effects, the study may have lacked the power to find it. However, a power analysis for linear mixed models, which needs simulation studies, was beyond the scope of this article. Second, our study did not apply psychiatric diagnostic of participants, and the number of participants presenting moderately severe to severe (>15) PHQ-9 scores was relatively low (11 out of 91, 12%). It could be that the effect in question manifests itself in clinical depression cases, as observed for example by Bistricky et al. (2011) and Bourke et al. (2010) but disappears when evaluating the non-clinical spectrum of depressive symptoms. Furthermore, many studies that observed FER deficits in depressive patients utilized ambiguous or neutral facial stimuli (e.g., Beevers et al., 2009; Kan et al., 2004, Bourke et al., 2010) whereas this study focused solely on a less ambiguous positive (happy) or less unambiguous negative (angry and sad) facial expressions. At the same time, it is relevant to note that there is still a debate in the literature as to whether facial emotion recognition biases are indeed a characteristic of depressive disorders. Wu et al. (2012), for example, reported normal performance by highly depressive patients when testing for accuracy in recognizing emotions.

Interaction of Mask-Wearing and Depression

Concerning H1.3 and H2.3, the data produced mixed results. For happiness, sadness, and anger, no consistent significant interaction was observed between the mask factor and depressive scores for any of the dependent variables. If at all, the pattern of the results was most suggestive in the case of happiness. Given the simple slope pattern, there may be some indication to speculate that there could be a significant difference once the sample size was larger. By taking a closer look at Figures 4 and 5 and the interaction coefficients, it seems that the presence of a face mask triggers a depressive response to the evaluation of happiness. In other words, when happy faces are masked, depressive scores correlate negatively in a slightly stronger fashion with emotion recognition and emotion intensity ratings. However, it is imperative to note that moderation analysis with the help of simple slopes revealed that the 95% confidence intervals of these coefficients were not significantly different from one another. One could speculate that with increased power and more reliable measurement, this effect could be found significant in future studies or the ongoing project *Gesichter lesen*. Still, effects are expected to be small, and no conclusions can be drawn from the present sample.

Implications of Findings

If one detected a significant interaction effect, for example, with a larger dataset, it could be interpreted in a few different ways. First, it is possible that, as reported by Leyman et al. (2008), participants with higher rates of depressive symptoms have more trouble fixating their gaze on happy faces and therefore present more difficulty in recognizing this emotion when the mask is present. Duque & Vázquez (2015) also observed attentional bias in depressed patients, in the sense that positive emotions attracted less attention than negative ones. The eye-tracking methodology could be utilized with the presence of a mask in future studies to further investigate this rationale. Second, one could argue that the mask may act as a social signal that triggers pandemic-related psychological burdens, and therefore impacts emotion interpretation performance. It is noteworthy to mention that the second explanation does not clarify why the interaction effect would be missing for sadness and anger.

The finding of an interaction effect on happiness and other emotions would have especially important implications. Many recently published studies highlight how pandemic-related safety measures interact with the mental health of vulnerable populations. One study conducted in Germany by Benke et al. (2020), for example, found that higher restrictions due to lockdown measures, a greater reduction of social contacts, and greater perceived changes in life were associated with higher mental health impairments. In Italy, Fiorenzato et al. (2021) found subjective cognitive functioning and mental health were strongly associated with enforcing social distancing measures. In China, Lai et al. (2020) reported particularly bad mental health outcomes for front-line professions such as doctors and nurses, people that consistently use face masks and protective equipment in their daily tasks. In this context, this preliminary study is a further account of how the impairments in facial emotion recognition brought about by the wearing of face masks can have a distinctive impact on people with a disposition to depressive symptoms.

The present investigation sheds light on one specific phenomenon that can exist in the universe of interactions between psychopathology and the burdens produced by the current pandemic, namely, that mask-wearing may significantly impair people's ability to interpret happy and sad facial expressions and that this impairment may present itself more strongly for individuals scoring higher on a depressive symptoms scale. Following the studies cited in this paragraph, the present data provides some evidence that the pandemic and the safety measures related to it do impact psychological functioning and may do so differently in dependence on people's mental health. Governments and responsible authorities should take these findings into consideration when dealing with the spread of the Coronavirus and planning future restrictions.

Limitations and Outlook

Although one should consider the time and accessibility constraints under which this study was conducted and evaluated, there are a few strengths worth mentioning. First, we built a time-effective online study that could be completed by anyone that spoke German and had a tablet or computer with up-to-date software and a webcam. Second, considering the complexity and time limitation usually associated with this investigation, this study conducted fairly sophisticated data analysis with the help of LMMs and its possibilities within the statistical software R, which help prevent false-positive associations due to population or relatedness structure and increase power by applying a correction that is specific to this structure. In addition, the preliminary character of this investigation allows for further scientific inquiry of the psychological processes addressed in this paper, both within and outside of the *Gesichter lesen* project.

Nonetheless, the present study contains several limitations worthy of critical evaluation. Because of the short period of time within which the data had to be collected for this thesis, we conducted the analysis on a relatively small and homogeneous sample. Having expected small effect sizes, a larger and more heterogeneous sample could account for more reliable results and possibly find effects that remained undisclosed in the present sample.

Dealing with the phenomenon of emotion perception and processing, the fact that this study was based solely upon psychological rating data is also a limitation. Including other forms of measurement to assess response to stimuli, such as psychophysiological data, can increase validity and further contribute to the understanding of the psychological response to emotions. The *Gesichter lesen* project is already in the process of collecting and analyzing this type of data. In a relat ed online experiment, Kastendieck et al. (2021) have found that facial mimicry - the perceiver's imitation of the other's emotional display - was reduced or absent in response to happy but preserved for sad mask-covered expressions. In the future, it would be interesting to see how mimicry and other physiological processes such as skin conductance and heart rate relate to psychopathology in emotion processing. For that, a study set up in laboratory settings would be the better approach.

Because of the COVID-19 pandemic, university facilities and therefore laboratories had their access restricted and did not allow studies to take place. In order to increase standardization (for example, using electromyography instead of OpenFace[3] video analysis), future studies should utilize superior methods for mimicry assessment and, if possible, go back to being conducted in laboratory environments.

As a last point, one could argue that there is still room for improvement in the quality of the stimuli. Here, we utilized morphed videos composed of static photos from neutral faces to faces displaying full emotions and added a face mask to it. To increase ecological validity, future stimuli should consist of recorded videos of people wearing face masks and expressing their respective emotions.

Conclusion

In summary, the present study found evidence that masks impair facial emotion recognition and bias the intensity rating of happy and sad faces but have no effect on angry faces. Depression scores were not associated with the outcomes in this sample. An interaction effect between mask-wearing and depressive symptoms may exist for happy facial expressions. The findings implicate that social interaction partners wearing face masks should pay additional attention to facial expressions to avoid misinterpreting emotions. In light of the results of this preliminary study and the relevance of precaution measures to contain the spread of COVID-19 at the current moment, authorities should consider the suggested deficits in emotion recognition and, in particular, how they associate with psychopathology when planning future pandemic-related public policy.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). https://doi.org/10.1176/appi.books.9780890425596

An, L., Hawley, S., Van Horn, M. L., Bacon, E., Yang, P., & Resnicow, K. (2021). Development of a coronavirus social distance attitudes scale. P*atient Education and Counseling, 104(6),* 1451–1459. https://doi.org/10.1016/j.pec.2020.11.027

Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of Other in the Self Scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology, 63(4)*, 596–612. https://doi.org/10.1037/0022-3514.63.4.596

Bassili, J. N. (1979). Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of the face. *Journal of Personality and Social Psychology, 37(11),* 2049–2058. https://doi.org/10.1037/0022-3514.37.11.2049

Baltrusaitis, T., Robinson, P., & Morency, L. P. (2016). *OpenFace: An open-source facial behavior analysis toolkit* (p. 10). https://doi.org/10.1109/WACV.2016.7477553

Beard, C., Hsu, K. J., Rifkin, L. S., Busch, A. B., & Björgvinsson, T. (2016). Validation of the PHQ-9 in a psychiatric sample. *Journal of Affective Disorders, 193,* 267–273. https://doi.org/10.1016/j.jad.2015.12.075

Beevers, C., Wells, T., Ellis, A., & Fischer, K. (2009). Identification of Emotionally Ambiguous Interpersonal Stimuli Among Dysphoric and Nondysphoric Individuals. *Cognitive Therapy and Research, 33*, 283–290. https://doi.org/10.1007/s10608-008-9198-6

Beck, A. T. (1976). *Cognitive therapy and the emotional disorders.* New York: Meridian.

Benke, C., Autenrieth, L. K., Asselmann, E., & Pané-Farré, C. A. (2020). Lockdown, quarantine measures, and social distancing: Associations with depression, anxiety, and distress at the beginning of the COVID-19 pandemic among adults from Germany. *Psychiatry Research, 293,* 113462. https://doi.org/10.1016/j.psychres.2020.113462

Berking, M., & Rief, W. (Eds.). (2012). *Klinische Psychologie und Psychotherapie für Bachelor:*

---

[3] The *Gesichter lesen* project is currently analyzing mimicry data using the OpenFace software. For more information on OpenFace, see Baltrusaitis et. al., (2016).

Band I: Grundlagen und Störungswissen. Lesen, Hören, Lernen im Web. Springer-Verlag. https://doi.org/10.1007/978-3-642-16974-8

Bijsterbosch, G., Mobach, L., Verpaalen, I. A. M., Bijlstra, G., Hudson, J. L., Rinck, M., & Klein, A. M. (2021). Validation of the child models of the Radboud Faces Database by children. *International Journal of Behavioral Development, 45(2),* 146–152. https://doi.org/10.1177/0165025420935631

Bistricky, S. L., Ingram, R. E., & Atchley, R. A. (2011). Facial affect processing and depression susceptibility: Cognitive biases and cognitive neuroscience. *Psychological Bulletin, 137(6)*, 998–1028. https://doi.org/10.1037/a0025348

Bourke, C., Douglas, K., & Porter, R. (2010). Processing of Facial Emotion Expression in Major Depression: A Review. Australian & New Zealand *Journal of Psychiatry, 44(8),* 681–696. https://doi.org/10.3109/00048674.2010.496359

Bower, G. H. (1981). Mood and memory. *American Psychologist, 36(2),* 129–148. https://doi.org/10.1037/0003-066X.36.2.129

Calbi, M., Langiulli, N., Ferroni, F., Montalti, M., Kolesnikov, A., Gallese, V., & Umiltà, M. A. (2021). The consequences of COVID-19 on social interactions: An online study on face covering. *Scientific Reports, 11(1),* 2601. https://doi.org/10.1038/s41598-021-81780-w

Calvo, M. G., Fernández-Martín, A., Gutiérrez-García, A., & Lundqvist, D. (2018). Selective eye fixations on diagnostic face regions of dynamic emotional expressions: KDEF-dyn database. S*cientific Reports, 8(1),* 17039. https://doi.org/10.1038/s41598-018-35259-w

Carbon, C. C. (2020a). Wearing Face Masks Strongly Confuses Counterparts in Reading Emotions. *Frontiers in Psychology, 11.* https://doi.org/10.3389/fpsyg.2020.566886

Carbon, C. C. (2020b). The Psychology of Wearing Face Masks in Times of the COVID-19 Pandemic. *SSRN Electronic Journal.* https://doi.org/10.2139/ssrn.3584834

Cheng, V. C.C., Wong, S.C., Chuang, V. W.M., So, S. Y.C., Chen, J. H.K., Sridhar, S., To, K. K.W., Chan, J. F.W., Hung, I. F.N., Ho, P.L., & Yuen, K.Y. (2020). The role of community-wide wearing of face mask for control of coronavirus disease 2019 (COVID-19) epidemic due to SARS-CoV-2. *The Journal of Infection, 81(1),* 107–114. https://doi.org/10.1016/j.jinf.2020.04.024

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences.* New York, NY: Routledge Academic

Dalili, M. N., Penton-Voak, I. S., Harmer, C. J., & Munafò, M. R. (2015). Meta-analysis of emotion recognition deficits in major depressive disorder. *Psychological Medicine, 45(6),* 1135–1144. https://doi.org/10.1017/S0033291714002591

Demenescu, L. R., Kortekaas, R., den Boer, J. A., & Aleman, A. (2010). Impaired Attribution of Emotion to Facial Expressions in Anxiety and Major Depression. *PLoS ONE, 5(12),* e15058. https://doi.org/10.1371/journal.pone.0015058

Derntl, B., Seidel, E.-M., Kryspin-Exner, I., Hasmann, A., & Dobmeier, M. (2009). Facial emotion recognition in patients with bipolar I and bipolar II disorder. *The British Journal of Clinical Psychology, 48(Pt 4)*, 363–375. https://doi.org/10.1348/014466509X404845

Druss, B. G. (2020). Addressing the COVID-19 Pandemic in Populations With Serious Mental Illness. *JAMA Psychiatry, 77(9)*, 891–892. https://doi.org/10.1001/jamapsychiatry.2020.0894

Duque, A., & Vázquez, C. (2015). Double attention bias for positive and negative emotional faces in clinical depression: Evidence from an eye-tracking study. *Journal of Behavior Therapy and Experimental Psychiatry, 46,* 107–114. https://doi.org/10.1016/j.jbtep.2014.09.005

Eisenbarth, H., & Alpers, G. W. (2011). Happy mouth and sad eyes: Scanning emotional facial expressions. *Emotion, 11(4),* 860–865. https://doi.org/10.1037/a0022758

Fayolle, S. L., & Droit-Volet, S. (2014). Time Perception and Dynamics of Facial Expressions of Emotions. *PLoS ONE, 9(5),* e97944. https://doi.org/10.1371/journal.pone.0097944

Fiorenzato, E., Zabberoni, S., Costa, A., & Cona, G. (2021). Cognitive and mental health changes and their vulnerability factors related to COVID-19 lockdown in Italy. *PLOS ONE, 16(1),* e0246204. https://doi.org/10.1371/journal.pone.0246204

Finch, J. F., & Zautra, A. J. (1992). Testing latent longitudinal models of social ties and depression among the elderly: A comparison of distri bution-free and maximum likelihood estimates with nonnormal data. *Psychology and Aging, 7(1),* 107–118. https://doi.org/10.1037/0882-7974.7.1.107

Fischer, A. H., Gillebaart, M., Rotteveel, M., Becker, D., & Vliek, M. (2012). Veiled Emotions: The Effect of Covered Faces on Emotion Perception and Attitudes. *Social Psychological and Personality Science, 3(3),* 266–273. https://doi.org/10.1177/1948550611418534

Freud, E., Stajduhar, A., Rosenbaum, R. S., Avidan, G., & Ganel, T. (2020). The COVID-19 pandemic masks the way people perceive faces. *Scientific Reports, 10(1),* 22344. https://doi.org/10.1038/s41598-020-78986-9

Haro, J. M., Arbabzadeh-Bouchez, S., Brugha, T. S., De Girolamo, G., Guyer, M. E., Jin, R., Lepine, J. P., Mazzi, F., Reneses, B., Vilagut, G., Sampson, N. A., & Kessler, R. C. (2006). Concordance of the Composite International Diagnostic Interview Version 3.0 (CIDI 3.0) with standardized clinical assessments in the WHO World Mental Health Surveys. *International Journal of Methods in Psychiatric Research, 15(4),* 167–180. https://doi.org/10.1002/mpr.196

Hong, L.X., Lin, A., He, Z.B., Zhao, H.H., Zhang, J.G., Zhang, C., Ying, L.J., Ge, Z.M., Zhang, X., Han, Q.Y., Chen, Q.Y., Ye, Y.H., Zhu, J.S., Chen, H.X., & Yan, W.H. (2020). Mask wearing in pre-symptomatic patients prevents SARS-CoV-2 transmission: An epidemiological analysis. *Travel Medicine and Infectious Disease, 36,* 101803. https://doi.org/10.1016/j.tmaid.2020.101803

Howard, J., Huang, A., Li, Z., Tufekci, Z., Zdimal, V., Westhuizen, H.M. van der, Delft, A. von, Price, A., Fridman, L., Tang, L.H., Tang, V., Watson, G. L., Bax, C. E., Shaikh, R., Questier, F., Hernandez, D., Chu, L. F., Ramirez, C. M., & Rimoin, A. W. (2020). *Face Masks Against COVID-19: An Evidence Review.* https://doi.org/10.20944/preprints202004.0203.v1

Kan, Y. (2004). Recognition of emotion from moving facial and prosodic stimuli in depressed patients. *Journal of Neurology, Neurosurgery & Psychiatry, 75(12),* 1667–1671.

Kim, Y. E., & Lee, B. (2019). The Psychometric Properties of the Patient Health Questionnaire-9 in a *Sample of Korean University Students. Psychiatry Investigation, 16(12),* 904–910. https://doi.org/10.30773/pi.2019.0226

Kessler, R. C., de Jonge, P., Shahly, V., van Loo, H. M., Wang, P. S. E., & Wilcox, M. A. (2014). *Epidemiology of depression.* In I. H. Gotlib & C. L. Hammen (Eds.), *Handbook of depression* (7–24). The Guilford Press. https://doi.org/10.1136/jnnp.2004.036079

Kastendieck, T., Zillmer, S., & Hess, U. (2021). (Un)mask yourself! Effects of face masks on facial mimicry and emotion perception during the COVID-19 pandemic. *Cognition and Emotion, 36(1),* 59–69. https://doi.org/10.1080/02699931.2021.1950639

Kluczniok, D., Hindi Attar, C., Fydrich, T., Fuehrer, D., Jaite, C., Domes, G., Winter, S., Herpertz, S. C., Brunner, R., Boedeker, K., & Bermpohl, F. (2016). Transgenerational effects of maternal depression on affect recognition in children. *Journal of Affective Disorders, 189,* 233–239. https://doi.org/10.1016/j.jad.2015.09.051

Kroenke, K., & Spitzer, R. L. (2002). The PHQ-9: A new depression diagnostic and severity measure. *Psychiatric Annals, 32(9),* 509–515. https://doi.org/10.3928/0048-5713-20020901-06

Lai, J., Ma, S., Wang, Y., Cai, Z., Hu, J., Wei, N., Wu, J., Du, H., Chen, T., Li, R., Tan, H., Kang, L., Yao, L., Huang, M., Wang, H., Wang, G., Liu, Z., & Hu, S. (2020). Factors Associated With Mental Health Outcomes Among Health Care Workers Exposed to Coronavirus Disease 2019. *JAMA Network Open, 3(3),* e203976. https://doi.org/10.1001/jamanetworkopen.2020.3976

Leyman, L., Raedt, R. D., Vaeyens, R., & Philippaerts, R. M. (2011). Attention for emotional facial expressions in dysphoria: An eye-movement registration study. *Cognition and Emotion, 25(1),* 111–120. https://doi.org/10.1080/02699931003593827

Lopez-Duran, N. L., Kuhlman, K. R., George, C., & Kovacs, M. (2013). Facial emotion expression recognition by children at familial risk for depression: High risk boys are oversensitive to sadness. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 54(5),* 565–574. https://doi.org/10.1111/jcpp.12005

Lehmann, K., Maliske, L., Böckler, A., & Kanske, P. (2019). Social Impairments in Mental Disorders: Recent Developments in Studying the Mecha nisms of Interactive Behavior. *Clinical Psychology in Europe, 1(2),* 1–15. https://doi.org/10.32872/cpe.v1i2.33143

Marta, C., Nunzio, L., Francesca, F., Martina, M., Anna, K., Vittorio, G., & Alessandra, U. M. (2020). *I see how you feel: Facial expressions' recognition and distancing in the time of COVID-19* [Preprint]. In Review. https://doi.org/10.21203/rs.3.rs-107145/v1

Mheidly, N., Fares, M. Y., Zalzale, H., & Fares, J. (2020). Effect of Face Masks on Interpersonal Communication During the COVID-19 Pandemic. *Frontiers in Public Health, 8:*582191. https://doi.org/10.3389/fpubh.2020.582191

Milders, M., Bell, S., Platt, J., Serrano, R., & Runcie, O. (2010). Stable expression recognition abnormalities in unipolar depression. *Psychiatry Research, 179(1),* 38–42. https://doi.org/10.1016/j.psychres.2009.05.015

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution, 4(2),* 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x

Nyquist, A. C., & Luebbe, A. M. (2020). An Emotion Recognition-Awareness Vulnerability Hypothesis for Depression in Adolescence: A Systematic Review. Clinical Child and Family Psychology Review, 23(1), 27–53. https://doi.org/10.1007/s10567-019-00302-3

Persad, S. M., & Polivy, J. (1993). Differences between depressed and nondepressed individuals in the recognition of and response to facial emotional cues. *Journal of Abnormal Psychology, 102(3),* 358–368. https://doi.org/10.1037//0021-843x.102.3.358

Platt, B., Kadosh, K. C., & Lau, J. Y. F. (2013). The Role of Peer Rejection in Adolescent Depression. *Depression and Anxiety, 30(9),* 809–821. https://doi.org/10.1002/da.22120

Richards, M., & Light, P. (Eds.). (1986). *Children of social worlds: Development in a social context* (p. 327). Harvard University Press.

Richardson, L. P., Rockhill, C., Russo, J. E., Grossman, D. C., Richards, J., McCarty, C., McCauley, E., & Katon, W. (2010). Evaluation of the PHQ-2 as a brief screen for detecting major depression among adolescents. *Pediatrics, 125(5),* e1097-1103. https://doi.org/10.1542/peds.2009-2712

Ruba, A. L., & Pollak, S. D. (2020). Children's emotion inferences from masked faces: Implications for social interactions during COVID-19. *PLOS ONE, 15(12),* e0243708. https://doi.org/10.1371/journal.pone.0243708

Rubinow, D. R., & Post, R. M. (1992). Impaired recognition of affect in facial expression in depressed patients. *Biological Psychiatry, 31(9),* 947–953. https://doi.org/10.1016/0006-3223(92)90120-o

Rymarczyk, K., Żurawski, Ł., Jankowiak-Siuda, K., & Szatkowska, I. (2016). Emotional Empathy and Facial Mimicry for Static and Dynamic Facial Expressions of Fear and Disgust. Fr*ontiers in Psychology, 7.* https://doi.org/10.3389/fpsyg.2016.01853

Schmid, P. C., & Schmid Mast, M. (2010). Mood effects on emotion recognition. *Motivation and Emotion, 34(3),* 288–292. https://doi.org/10.1007/s11031-010-9170-0

Schurgin, M. W., Nelson, J., Iida, S., Ohira, H., Chiao, J. Y., & Franconeri, S. L. (2014). Eye movements during emotion recognition in faces. *Journal of Vision, 14(13),* 14–14. https://doi.org/10.1167/14.13.14

Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A Brief Measure for Assessing Generalized Anxiety Disorder: The GAD-7. *Archives of Internal Medicine, 166(10),* 1092–1097. https://doi.org/10.1001/archinte.166.10.1092

Spitzer, M. (2020). Masked education? The benefits and burdens of wearing face masks in schools during the current Corona pandemic. *Trends in Neuroscience and Education, 20,* 100138. https://doi.org/10.1016/j.tine.2020.100138

Titov, N., Andrews, G., Davies, M., McIntyre, K., Robinson, E., & Solley, K. (2010). Internet treatment for depression: A randomized controlled trial comparing clinician vs. technician assistance. *PloS One, 5(6),* e10939. https://doi.org/10.1371/journal.pone.0010939

World Health Organization (2017). Depression and Other Common Mental Disorders. Global Health Estimates. Geneva: World Health organization.Retrieved from: https://apps.who.int/iris/handle/10665/254610 Licence: CC BY-NC-SA 3.0 IGO.

Graduate Student Journal of Psychology
2015, Vol. 16

Afraid to vomit? The relationship between temperamental fear, emetophobia symptoms, and the impact of sex

Jennifer A. Petell[1], Molly E. Wickenhauser,[1] and Danielle J. Maack,[1,2]
[1]Department of Psychology, University of Mississippi
[2] Delta Autumn Consulting

Emetophobia, the specific phobia of vomiting (SPOV), is an anxiety disorder associated with a persistent fear of emesis. The temperament of fear is evident in phobic and anxiety disorders, which behaviorally manifests through a sensitivity in the fight, flight, freeze system (FFFS). The present study aimed to test specific associations between these constructs. Undergraduate students (N= 186) completed self-report measures assessing symptoms of emetophobia, the FFFS, and general anxiety. Correlational analyses were used to test initial relations, followed by regression analyses to assess the unique contribution of temperamental fear to the emetophobia symptom experience. Significant positive correlations were found between emetophobia symptoms with anxiety symptoms and the overall temperament of fear. However, only the fight response of the FFFS was significantly associated with symptoms of emetophobia. Further regression analyses demonstrated that the fight response did not uniquely predict emetophobia symptoms above and beyond anxiety. However, post-hoc analyses illustrated sex moderated the relationship between temperamental fear and emetophobia symptoms, such that the relationship was significantly stronger for male participants than female participants. Results of this study expand upon the current conceptualization of emetophobia to incorporate underlying temperamental vulnerabilities and sex differences, which may function to exacerbate and/or maintain symptoms.

Keywords: emetophobia; specific phobia of vomiting (SPOV); fight, flight, freeze system (FFFS); fight, flight, freeze questionnaire (FFFQ); sex differences

The specific phobia of vomiting (SPOV), also known as emetophobia, is a specific phobia characterized by a marked fear of emesis (self or other). Though symptoms have a mean age of onset occurring between 7.5 to 15.7 years, symptoms may develop and frequently persist beyond young adulthood (Keyes & Veale, 2018). In fact, symptoms are pervasive and persist for an average duration of 25.9 years before individuals seek treatment (Keyes et al., 2017; Lipsitz et al., 2001). Although inconsistent (Veale et al., 2015), prevalence estimates of emetophobia demonstrate a much higher occurrence in females (6-7%) than in males (1.8-3.1%; Kirkpatrick & Berg, 1981; van Hout & Bouman, 2012). Those with emetophobia symptoms have severe, negative impairments in quality of life and may develop co-occurring anxiety disorders (e.g., generalized anxiety disorder (GAD), obsessive-compulsive disorder (OCD), social anxiety disorder (SAD); Boschen, 2007; Keyes et al., 2017).

In addition to the core symptom of pervasive fear of vomiting, there are a variety of fear emphases that an individual may also hold. For example, 41-75% of individuals with emetophobia present with a primary fear of vomiting themselves; while 18-45% of individuals are primarily debilitated upon seeing others vomit in their presence (Keyes & Veale, 2018). Other fears of vomiting surround the potential of choking on vomit leading subsequent to death, the act of vomiting in public (16-62%), and/or the physiological symptoms themselves associated with vomiting (Keyes & Veale, 2018). These impairing symptoms of emetophobia are supported through elaborate safety-seeking and avoidance behaviors, such as checking food expiration dates and avoiding contact with an ill person (Simons & Vloet, 2018; van Hout & Bouman, 2012). Through these consistent responses with acute reinforcement, the phobia is actively maintained and exacerbated.

Though symptoms overlap, it is evident that anxiety and fear are both unique constructs and components observed in anxiety disorders (Gullone et al., 2000; McNaughton & Corr, 2008; Perkins et al., 2007; Woody & Teachman, 2000). The psychological constructs of anxiety and fear have been conceptualized using the revised Reinforcement Sensitivity Theory (r-RST), a neuropsychological model of personality and motivation developed by Gray and McNaughton (2000). The r-RST consists of three motivational systems that work together to respond to threats: behavioral activation and inhibition systems (BIS and BAS) and the fight, flight, freeze system (FFFS; Corr & Krupić, 2017; McNaughton & Corr, 2008). The BIS and BAS are predominantly used to conceptualize anxiety, which is posited to occur upon approaching of perceived danger (Gray & McNaughton, 2000;

McNaughton & Corr, 2008). The third motivational system, the FFFS, is used to operationalize the temperament of fear, which operates upon active avoidance (i.e., leaving the vicinity) of perceived danger (Gray & McNaughton, 2000; McNaughton & Corr, 2008; Walker & Jackson, 2017). Indeed, the BIS and BAS, and to a lesser extent the FFFS (Kimbrel et al., 2008), have been used in the conceptualization of anxiety symptoms and disorders (e.g., SAD and OCD; Bijttebier et al., 2009; Katz et al., 2020). One area of limited research is the relationship of FFFS with anxiety disorders (Kambouropoulos et al., 2014; Kimbrel et al., 2008). As the r-RST motivational symptoms are concurrently activated (Corr & Krupić, 2017; McNaughton & Corr, 2008), it stands to reason that the temperament of fear, as operationalized by the FFFS, is implicated in the experience of anxiety symptoms, and likely emetophobia symptoms.

The FFFS is maintained through learned avoidance of aversive stimuli and is theorized to be highly sensitive to punishment (Kimbrel et al., 2008), thereby reinforcing avoidance behaviors seen in anxiety disorders and likely emetophobia. Specifically, the fight response is characterized by defensive aggression evoked in the presence of proximal, unescapable threats; while the flight response is conceptualized as a quick, direct escape from distal threats (Walker et al., 2017). Finally, the freeze response is considered a physiological panic response to distal threats involving physical immobility (Walker et al., 2017). Although the flight and freeze responses are classic avoidance behaviors seen in emetophobic responses to potential or distal fear (e.g., leaving the vicinity of an ill individual or shutting down mentally and physically upon the sight of vomit; Keyes et al., 2017, Veale et al., 2013), the fight response may also occur. For example, a threatening situation (e.g., the inability to vacate a room with ill colleagues) may provoke a fight response expressed as outright physical or verbal aggression or subtle nonverbal hostility. As is consistently demonstrated in phobic disorders, reinforcement and behavioral responses have a significant effect on the presentation and severity of emetophobia (Keyes & Veale, 2018; Wu et al., 2015).

The Present Study

To date, the symptoms of emetophobia have not been assessed within the r-RST framework or the specific FFFS motivational system. The primary aim of the present exploratory study was to assess the overall relationship between emetophobia symptoms, general symptoms of anxiety, and fight, flight, freeze tendencies (combined and individually) in an undergraduate, analog population. As fear is inherently involved in phobias and other anxiety disorders, the current study aimed to examine both symptoms of emetophobia and overall anxiety with the FFFS motivation system. Based on the r-RST theory and current conceptualization of emetophobia, it was hypothesized that emetophobia symptoms, general anxiety, and FFFS would be positively associated. A secondary aim of the study was to investigate the unique contribution of fear in the experience of emetophobia symptoms when controlling for the known association of general anxiety symptoms (Boschen, 2007).

Methods

Participants and Procedures

Participants (N = 186; Mage = 19.05 years, SD = 1.69) were undergraduate students, aged 18 to 36 years old, enrolled in an introductory Psychology course at a large South-Eastern university in the United States. Participants were recruited via class announcements and through SONA Systems (an online participant recruitment portal). No other inclusion or exclusion criteria were present as this was an exploratory study assessing potential associations between temperamental fear and symptoms of emetophobia in an analog sample. Further, symptoms of emetophobia may develop after the average age of onset and often persist through adulthood (Keyes et al., 2017), the current sample of undergraduate students is consistent with individuals with emetophobia who may seek treatment for persistent symptoms. The majority of participants were female (74.9%; 25.1% male) and White (71.7%), though participants also identified as African American (21.2%), Asian (1.6%), Native American or Alaskan Native (0.5%), or Multiracial (4.6%), and 5.2% identified as Hispanic.

As a part of a larger study, undergraduate students presented in person to the lab. Prior to study commencement participants provided verbal and written informed consent, and then completed a series of self-report questionnaires (non-randomized). Participants received research course credit for participation. The following three questionnaires were examined for the purposes of the current study. All procedures were approved by the University's Institutional Review Board.

Instruments

The Specific Phobia of Vomiting Inventory (SPOVI; Veale, et al., 2013). The SPOVI is a short, 14-item measure assessing fear of vomiting symptoms experienced within the last seven days. All items are rated on a Likert-type scale from 0 (symptoms not at all experienced) to 4 (symptoms experienced all the time). Scores are summed, with higher scores suggesting an increased experience of SPOV symptoms (range 0-35). A clinical cut-off of 10 has been previously proposed (Veale, et al., 2013), which was endorsed by 9.14% of the current study sample (n = 17). The SPOVI has displayed good internal consistency ($\alpha = .89$), good convergent and divergent validity in college samples, and has evidenced invariance across sex (Maack et al., 2017). In the current study, the SPOVI also demonstrated good internal consistency ($\alpha = .82$).

The Fight, Flight, Freeze Questionnaire (FFFQ; Maack et al., 2015). The FFFQ is a 21-item measure assessing the typical reaction of an individual to a threatening situation. All items are rated on a Likert-type scale from 1 (experienced almost never) to 5 (experienced almost always). The measure results in a total FFFQ score as well as three subscale scores individually assessing fight, flight, and freeze. The FFFQ has demonstrated reliable internal consistency overall ($\alpha = .92$) and with all subscale scores (fight: $\alpha = .91$; flight: $\alpha = .94$; freeze: $\alpha = .86$; Maack et al., 2015; Walker & Jackson, 2017). The current study found similarly good internal consistency for the overall FFFQ score ($\alpha = .93$) and individual subscales (fight: $\alpha = .90$; flight: $\alpha = .94$; freeze: $\alpha = .91$).

The Depression, Anxiety, Stress Scale (DASS-21; Lovibond & Lovibond, 1995). The DASS-21 assesses depression, anxiety, and stress symptoms experienced within the last week, consisting of 21 items. All items are on a Likert-type scale from 0 (not applicable) to 3 (applicable much of the time). The 7-item anxiety subscale was used to assess anxiety as a potential covariate apart from overall temperamental fear. Both the overall scale and anxiety subscale scores have demonstrated good internal consistency ($\alpha = .82$ and $\alpha = .81$, respectively; Crawford & Henry, 2003; Osman et al., 2012). The current study showed acceptable internal consistency for the subscale of interest (anxiety subscale: $\alpha = .71$) and overall scale ($\alpha = .74$).

Statistical Analysis

All data analyses were conducted using SPSS version 27. Data was cleaned prior to analyses. Of the original sample (N= 194), three participants were removed for missing more than 10% of data, five multivariate outliers were removed using Mahalanobis distance, and additional missing data was replaced with variable means. Preliminary analyses of the final sample (N= 186) included descriptive statistics and zero-order correlations for all variables of interest. Following this, a hierarchical regression analysis was run to assess the unique variance of predictors on the experience of emetophobia symptoms. Post-hoc analyses, including correlations, independent samples t-tests, and a moderation analysis, examined the potential impact and interactional effects of sex on the relationship between temperamental fear and emetophobia symptoms.

Results

Descriptive statistics and intercorrelations among variables of interest are presented in Table 1. As illustrated, SPOV symptoms were significantly related to anxiety ($r = .20$, $p < .01$) and both the overall temperament of fear ($r = .17$, $p < .05$) and the fight subscale ($r = .15$, $p < .05$). However, the subscales of freeze ($r = .13$, $p = .08$) and flight ($r = .13$, $p = .08$) were not significantly associated with SPOV symptoms.

To assess the unique contribution of the temperamental fear response to the experience of emetophobia symptoms, distinct from general anxiety symptoms, a hierarchical regression analysis was run with SPOV symptoms as the dependent variable (see Table 2). In the first step of the model, the fight subscale (the only significantly associated subscale of fear) was entered. This step of the model was significant ($F[1, 184] = 4.03$, $p < .05$, $\Delta R2 = .02$), with fight accounting for 2.1% of the variance in the prediction of emetophobia symptoms. Next, in the second step of the model, anxiety was added. The overall model was significant ($F[1, 183] = 5.72$, $p < .01$, $\Delta R2 = .03$); however, the contribution of the fight subscale no longer added unique predictive ability.

Post-hoc Analyses

Given the above unexpected findings and clear sex differences in the occurrence of emetophobia symptoms evidenced in previous studies (Kirkpatrick &

Berg, 1981; van Hout & Bouman, 2012), post-hoc analyses related to sex were conducted to elucidate potential unique contributions of sex in the present preliminary study. Additionally, although using a non-clinical sample for an initial exploratory study, an assessment of clinical elevations of SPOV symptoms occurred. In the current study, 17 participants (9.14%) met or exceeded the threshold for clinical cut-off scores on the SPOVI (score > 10). Consistent with the previously mentioned studies (Kirkpatrick & Berg, 1981; van Hout & Bouman, 2012), of the 17 participants who met the SPOVI clinical cut-off, 13 were females (6.99%), while only 4 were males (2.27%). Point-biserial correlations demonstrated that sex was only significantly associated with the flight subscale ($r = .17$, $p < .05$) and was not significantly associated with SPOV symptoms, anxiety, overall temperamental fear, or the fight or freeze subscales. Further, independent samples t-tests demonstrated no significant differences between sex among any of the variables (ps > .05).

A simple moderation analysis using PROCESS 3.5 by Hayes (Hayes, 2018) was conducted to test if sex acted as a moderator among temperamental fear and emetophobia symptoms. Temperamental fear was entered as the predictor variable, sex as the moderator, and SPOV symptoms as the outcome. Anxiety was entered as a covariate. The overall model was significant ($F [4, 178] = 3.66$, $p < .01$, $R2 = .08$). More importantly, the interaction between temperamental fear and sex was also significant ($F [1, 178] = 4.14$, $p < .05$, $R2 = .02$). Specifically, the relationship between temperamental fear and SPOV symptoms was significantly stronger for men compared to women (see Figure 1).

Discussion

The aim of the current study was to investigate the fight, flight, freeze motivational system and its relations with symptoms of emetophobia. Results provided initial evidence for the impact of fear, specifically the fight tendency, on symptoms of the specific phobia of vomiting. However, further extrication of fear and anxiety is needed.

Consistent with theory and as hypothesized, the overall temperament of fear was found to be significantly associated with symptoms of emetophobia (Harnett et al., 2013; Keyes & Veale, 2018; Perkins et al., 2007; Wu et al., 2015). However, contrary to the hypotheses, not all domains of the FFFS behavioral responses were related. Specifically, the flight and freeze responses, common reactions to potential threats across species (Roelofs, 2017), were not significantly associated with emetophobia symptoms in the current study. Notably, the fight response was the only fear domain significantly associated with SPOV symptoms. Regarding the salient role of the fight response in the overall sample, individuals may respond differently in situations wherein active avoidance (i.e., escape from the immediate and proximal threat) of vomit is not available (McNaughton & Corr, 2008). The perceived urgency of threat and inaccessible escape in these situations may prompt the fight response to override other FFFS reactions experienced more commonly among distal threats. However, upon further examination, the fight subscale did not add significant predictive utility above and beyond general anxiety symptoms. Additional research may examine the unique impact of the remaining motivational systems of the r-RST, the BIS and BAS, which are operationalized as the anxiety response (McNaughton & Corr, 2008), and provide a further understanding of these results. It is possible that SPOV symptoms are more closely linked to a preparatory response in anticipation of future negative consequences from the act of vomiting as opposed to a defensive response to vomiting alone (Barlow, 2002; Lang et al., 2000).

Potential explanations for the aforementioned discrepancies observed in the FFFS and emetophobia symptoms may involve documented hypervigilance regarding perceived vomit-related risks and threats (Boschen, 2007; Keyes & Veale, 2018). The flight and freeze responses are implicated in response to perceived threats that are distal to the individual, while the fight response uniquely combats proximal threats (Walker et al., 2017). Given the significant association of emetophobia symptoms with the fight response, this suggests participants in the current study perceived a proximal, immediate threat wherein active avoidance was unattainable (McNaughton & Corr, 2008). However, this perception may have been overshadowed by the strong connection between emetophobia symptoms and general anxiety about potential future threats that could result from vomiting (e.g., difficulty breathing, choking, embarrassment). Further research may benefit by clarifying the situational contexts unique to emetophobia which provoke the perception of proximal and distal threats and consequent arousal.

In addition to being impacted by assessed distance (i.e., proximal or distal) from threat (i.e., vomiting), sex differences have also been demonstrated to impact risk perception across species (Gruene et al., 2015; Gustafsod, 1998; Jones & Monfils, 2016). In the current study, the flight response was associated with being female, which suggests differential behavioral responses in the motivational system may occur. One possibility is that females may achieve more successful attempts of physically or mentally distancing from vomit. For instance, females experiencing emetophobia may employ flight responses (such as actively avoiding ill children at the local preschool or temporarily restricting food consumption) to avoid possible situations of vomit that may or may not occur in the distal future. Additional research should examine the frequency of the specific motivational systems of the r-RST among males and females, and whether those systems were successful at avoiding the anxiety-provoking stimuli.

Negative affect, including experience of anger and irritability, may also account for differences in behavioral responses. Emotions of negative affect (anger or irritability, anxiety, disgust, etc.) are expressed among various emotional disorders, such as anxiety and depression (Hofmann et al., 2012; Hundt et al., 2013) and have been linked to higher intensity in women (Fujita et al., 1991). Within the revised Reinforcement Sensitivity Theory (r-RST), the experience of negative affect has also been demonstrated to negatively impact the perception of daily events (Hundt et al., 2013). Further, the specific emotion of anger has been positively associated with heightened sensitivity to reward and aggression, and it is further associated with an elevated fight response to threatening situations (Hundt et al., 2013; Roelofs, 2017). As such, the fight response may be perceived as resulting in a larger, more salient reward (to evade the proximal, immediate threat) than flight or freeze (to evade the distal, potential threat). Thus, negative affect, particularly when accounting for sex differences, may alter the behavioral response of the FFFS in individuals with emetophobia.

Given the inconsistencies between the current study's findings and theory, post-hoc analyses were conducted to explore a potential missing link of the previous analyses: sex. Results from the current study demonstrated clinical levels of emetophobia symptoms were three times more prevalent in female participants than males, consistent with prior research (Kirkpatrick & Berg, 1981; van Hout & Bouman, 2012). Interestingly, the significant interaction effect found in the moderation model demonstrated a stronger relation among men for temperamental fear and emetophobia symptoms. That is, during moments wherein heightened fear symptoms are activated, men will likely experience increased severity of emetophobia symptoms. In comparison, this effect was not found in females in the current sample. Elucidating whether this demonstration is unique to the FFFS, as compared to the BIS and BAS, may provide further understanding regarding sex differences in emetophobia.

It is possible, potentially as a result of societal differences and behavioral expectations among sex, that males and females may display and engage in emetophobia symptoms differently. Sex differences have been evidenced across the lifespan among temperamental fear and anxiety symptoms (McLean & Anderson, 2009). For example, disgust sensitivity, or the degree an individual regulates the emotion of disgust, has been found to be higher in females than in males (Cisler et al., 2009; Connolly et al., 2008). As disgust sensitivity promotes avoidance of perceived disease (Connolly et al., 2008), this suggests females may engage in increased flight responses, as evidenced in the current study. Future studies are needed to elucidate sex differences in the experience of emetophobia to provide additional contextual information to better inform treatment approaches. Overall, the results of the current preliminary study add to the current conceptualization of emetophobia as its relationship with the temperament of fear combined with the impact of sex has not yet been documented in the literature.

Limitations

Although this preliminary study of temperamental fear and symptoms of emetophobia furthers the extant literature, it is not without limitations. Specifically, data was collected from an undergraduate population with no formal diagnostic assessment of emetophobia or other psychiatric disorders. As the study was considered exploratory to first determine the potential of a relation between fear tendencies and symptoms of emetophobia, there were no exclusion or inclusion criteria. It is important to note that substantial variance nical sample although having good internal validity (α = .82 in this sample) may not be representative of external validity (i.e., true diagnostic status). Moreover, replicating this study in a clinical sample with structured clinical assessments of emetophobia is needed to truly further the overall conceptualization of emetophobia.

Another limitation, as previously mentioned, was that the study relied upon self-report measures wherein the threat of immediate and proximal vomit was not present, thereby potentially impacting the innate behavioral responses of participants. Despite these limitations, the current study provides a novel, initial understanding of the FFFS and the impact of sex on emetophobia symptoms.

To address such limitations in the future, studies employing behavioral avoidance tasks (BATs) related to emetophobia stimuli may improve the understanding of natural behavioral responses across sex and in response to tangible vomit-specific threats. Additionally, research examining the FFFS in individuals formally diagnosed with emetophobia may contribute to a comprehensive conceptualization of this disorder. Additionally, clarity on possible confounding variables related to FFFS presentation may be provided by formal diagnostic assessment by a trained clinician and/or the addition of self-report measures. Finally, based on previous literature indicating anger is associated with an increased sensitivity to reward, the emotion of anger in relation to the overall temperament of fear, individual response tendencies (i.e. fight, flight, freeze), and symptoms of emetophobia is worth exploration (Hundt et al., 2013).

Conclusion

The present study assessed the associations of temperamental fear (fight, flight freeze systems) with the experience of emetophobia symptoms. Although symptoms of emetophobia were associated with the temperament of fear, specifically the fight subscale, none of the subscales offered predictive utility above and beyond general anxiety symptoms. Interestingly, post-hoc analyses revealed that sex moderated the relationship between the overall FFFS and emetophobia symptoms, such that this relationship was much stronger for male participants than for females. To advance the conceptualization of emetophobia, additional research is needed to first confirm the association between the FFFS and emetophobia behaviors (i.e. using behavioral tasks) and then assess/identify these specific fear behaviors in a clinical sample. Further, exploring whether the temperament of fear is generalized to vomit-specific stimuli and proximally threatening situations may provide information to assist with comprehensive treatment. Importantly, the current study expands the literature in the area of emetophobia by assessing the specific underlying temperamental vulnerability of fear, the impact of sex, and how this relationship significantly affects symptoms of this phobic experience.

References

Barlow, D. H. (2002). *Anxiety and its disorders: The nature and treatment of anxiety and panic* (2nd ed.). New York, NY: Guilford Press.

Bijttebier, P., Beck, I., Claes, L., & Vandereycken, W. (2009). Gray's Reinforcement Sensitivity Theory as a framework for research on personality–psychopathology associations. *Clinical Psychology Review, 29(5),* 421–430. https:// doi:10.1016/j.cpr.2009.04.002.

Boschen, M. J. (2007). Reconceptualizing emetophobia: A cognitive–behavioral formulation and research agenda. *Journal of Anxiety Disorders, 21(3),* 407-419. https://doi.org/10.1016/j.janxdis.2006.06.007

Cisler, J. M., Olatunji, B. O., & Lohr, J. M. (2009). Disgust sensitivity and emotion regulation potentiate the effect of disgust propensity on spider fear, blood-injection-injury fear, and contamination fear. *Journal of Behavior Therapy and Experimental Psychiatry, 40(2),* 219–229. https:// doi:10.1016/j.jbtep.2008.10.002

Connolly, K. M., Olatunji, B. O., & Lohr, J. M. (2008). Evidence for disgust sensitivity mediating the sex differences found in blood-injection-injury phobia and spider phobia. *Personality and Individual Differences, 44(4),* 898–908. https:// doi:10.1016/j.paid.2007.10.020

Crawford, J. R., & Henry, J. D. (2003). The Depression Anxiety Stress Scales (DASS): Normative data and latent structure in a large non-clinical sample. *British Journal of Clinical Psychology, 42,* 111–131. https:// doi:10.1348/014466503321903544

Fujita, F., Diener, E., & Sandvik, E. (1991). Gender differences in negative affect and well-being: The case for emotional intensity. *Journal of Personality and Social Psychology, 61(3),* 427–434. https://doi-org.umiss.idm.oclc.org/10.1037/0022-3514.61.3.427

Gray, J.A., & McNaughton, N. (2000). *The neuropsychology of anxiety.* Oxford: Oxford University

Press.

Gruene, T. M., Flick, K., Stefano, A., Shea, S. D., & Shansky, R. M. (2015). Sexually divergent expression of active and passive conditioned fear responses in rats. *Elife, 4,* e11352. https://doi: 10.7554/eLife.11352

Gullone, E., King, N. J., & Ollendick, T. H. (2000). The development and psychometric evaluation of the Fear Experiences Questionnaire: An attempt to disentangle the fear and anxiety constructs. *Clinical Psychology & Psychotherapy, 7(1),* 61-75. https://doi:10.1002/(SICI)1099-0879(200002)7:1<61::AID-CPP227>3.0.CO;2-P

Gustafsod, P. E. (1998). Gender Differences in risk perception: Theoretical and methodological perspectives. *Risk analysis, 18(6),* 805-811. https://doi.org/10.1023/B:RIAN.0000005926.03250.c0

Harnett, P. H., Loxton, N. J., & Jackson, C. J. (2013). Revised reinforcement sensitivity theory: Implications for psychopathology and psychological health. *Personality and Individual Differences, 54(3),* 432–437. https://doi:10.1016/j.paid.2012.10.019

Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (2nd edition). Guilford Press.

Hofmann, S. G., Sawyer, A. T., Fang, A., & Asnaani, A. (2012). Emotion dysregulation model of mood and anxiety disorders. *Depression and anxiety, 29(5),* 409-416. https:// doi:10.1002/da.21888

Hundt, N. E., Brown, L. H., Kimbrel, N. A., Walsh, M. A., Nelson-Gray, R., & Kwapil, T. R. (2013). Reinforcement sensitivity theory predicts positive and negative affect in daily life. *Personality and Individual Differences, 54(3)*, 350–354. https://doi:10.1016/j.paid.2012.09.021

Jones, C. E., & Monfils, M.-H. (2016). Fight, flight, or freeze? The answer may depend on your sex. *Trends in Neurosciences, 39(2),* 51–53. https://doi-org.umiss.idm.oclc.org/10.1016/j.tins.2015.12.010

Kambouropoulos, N., Egan, S., O'Connor, E. J., & Staiger, P. K. (2014). Escaping threat: Understanding the importance of threat sensitivity in social anxiety. *Journal of Individual Differences, 35(1),* 47.https://doi: 10.1027/1614-0001/a000126

Keyes, A., Gilpin, H. R., & Veale, D. (2017). Phenomenology, epidemiology, co-morbidity and treatment of a specific phobia of vomiting: A systematic review of an understudied disorder. *Clinical Psychology Review, 60,* 15-31. https://doi:10.1016/j.cpr.2017.12.002

Keyes, A., & Veale, D. (2018). Atypical eating disorders and specific phobia of vomiting: Clinical presentation and treatment approaches. In L. K. Anderson, S. B. Murray, W. H. Kaye, L. K. Anderson, S. B. Murray, W. H. Kaye (Eds.), Clinical handbook of complex and atypical eating disorders (pp. 270-289). New York, NY, US: Oxford University Press.

Kimbrel, N. A., Cobb, A. R., Mitchell, J. T., Hundt, N. E., & Nelson-Gray, R. O. (2008). Sensitivity to punishment and low maternal care account for the link between bulimic and social anxiety symptomology. Eating Behaviors, 9(2), 210–217. https:// doi:10.1016/j.eatbeh.2007.09.005

Kirkpatrick, D. R., & Berg, A. J. (1981). Fears of a heterogeneous, nonpsychiatric sample: A factor analytic study. Paper presented at the Am. psychol. Ass. Conf. Los Angeles, Calif.

Corr, P. J., & Krupić, D. (2017). Motivating personality: Approach, avoidance, and their conflict. In A. J. Elliot (Ed.), *Advances in motivation science (pp. 39–90).* Elsevier Academic Press.

Katz, B. A., Matanky, K., Aviram, G., & Yovel, I. (2020). Reinforcement sensitivity, depression and anxiety: A meta-analysis and meta-analytic structural equation model. *Clinical Psychology Review, 77,* 101842. https://doi.org/10.1016/j.cpr.2020.101842

Lang, P. J., Davis, M., & Ohman, A. (2000). Fear and anxiety: Animal models and human cognitive psychophysiology. *Journal of Affective Disorders, 61(3),* 137-159. https://doi.org/10.1016/s0165-0327(00)00343-8

Lipsitz, J. D., Fyer, A. J., Paterniti, A., & Klein, D. F. (2001). Emetophobia: Preliminary results of an Internet survey. *Depression and Anxiety, 14(2),* 149-152. https://doi:10.1002/da.1058

Lovibond, S.H. & Lovibond, P.F. (1995). *Manual for the Depression Anxiety & Stress Scales.* (2nd Ed.) Sydney: Psychology Foundation.

Maack D.J., Ebesutani C, Smitherman T.A. (2017) Psychometric investigation of the specific phobia of vomiting inventory: A new factor model. *International Journal of Methods in Psychiatry*

Research, 27(1), e1574. https://doi.org/10.1002/mpr.1574

Maack, D. J., Buchanan, E., & Young, J. (2015). Development and psychometric investigation of an inventory to assess fight, flight, and freeze tenden cies: the fight, flight, freeze questionnaire. *Cognitive Behavior Therapy, 44(2),* 117-127. https:// do i:10.1080/16506073.2014.972443

McLean, C. P., & Anderson, E. R. (2009). Brave men and timid women? A review of the gender differences in fear and anxiety. *Clinical Psychology Review, 29(6),* 496–505. https://doi-org.umiss.idm.oclc.org/10.1016/j.cpr.2009.05.003

McNaughton, N., & Corr, P. J. (2008). *The neuropsychology of fear and anxiety: A foundation for Reinforcement Sensitivity Theory.* In P. J. Corr (Ed.), The reinforcement sensitivity theory of personality. (pp. 44–94). Cambridge University Press. https://doi-org.umiss.idm.oclc.org/10.1017/CBO9780511819384.003

Osman, A., Wong, J. L., Bagge, C. L., Freedenthal, S., Gutierrez, P. M., & Lozano, G. (2012). The Depression Anxiety Stress Scales—21 (DASS-21): Further examination of dimensions, scale reliability, and correlates. *Journal of Clinical Psychology, 68(12),* 1322–1338. https://doi:10.1002/jclp.21908

Perkins, A. M., Kemp, S. E., & Corr, P. J. (2007). Fear and anxiety as separable emotions: An investigation of the revised reinforcement sensitivity theory of personality. *Emotion, 7(2),* 252–261. https://doi:10.1037/1528-3542.7.2.252

Roelofs, K. (2017). Freeze for action: Neurobiological mechanisms in animal and human freezing. Philosophical Transactions of the Royal Society: *Biological Sciences, 372(1718),* 20160206. https://doi:10.1098/rstb.2016.0206

Simons, M., & Vloet, T. D. (2018). Emetophobia—A metacognitive therapeutic approach for an overlooked disorder. Zeitschrift Für Kinder- Und Jugendpsychiatrie *Und Psychotherapie, 46(1),* 57-66. https:// doi:10.1024/1422-4917/a000464

Van Hout, W. J., & Bouman, T. K. (2012). Clinical features, prevalence and psychiatric complaints in subjects with fear of vomiting. *Clinical Psychology & Psychotherapy, 19(6),* 531-539. https://doi:10.1002/cpp.761

Veale, David & Hennig, Camilla & Gledhill, Lucinda. (2015). Is a specific phobia of vomiting part of the obsessive compulsive and related disorders? *Journal of Obsessive-Compulsive and Related Disorders, 7,* 1-6. https://doi: 10.1016/j.jocrd.2015.08.002.

Veale, D., Ellison, N., Boschen, M. J., Costa, A., Whelan, C., Muccio, F., & Henry, K. (2013). Development of an inventory to measure specific phobia of vomiting (emetophobia). *Cognitive Therapy and Research, 37(3),* 595-604. https:// doi:10.1007/s10608-012-9495-y

Walker, B. R., & Jackson, C. J. (2017). Examining the validity of the revised Reinforcement Sensitivity Theory scales. *Personality and Individual Differences, 106,* 90-94. https:// doi:10.1016/j.paid.2016.10.035

Walker, B. R., Jackson, C. J., & Frost, R. (2017). A comparison of revised reinforcement sensitivity theory with other contemporary personality models. P*ersonality and Individual Differences,* 109232-236. https://doi:10.1016/j.paid.2016.12.053

Woody, S. R., & Teachman, B. A. (2000). Intersection of disgust and fear: Normative and pathological views. *Clinical Psychology: Science and Practice, 7(3),* 291-311. https://doi:10.1093/clipsy/7.3.291

Wu, M. S., Rudy, B. M., Arnold, E. B., & Storch, E. A. (2015). Phenomenology, clinical correlates, and impairment in emetophobia. *Journal of Cognitive Psychotherapy, 29(4),* 356-368. https://doi:10.1891/0889-8391.29.4.356

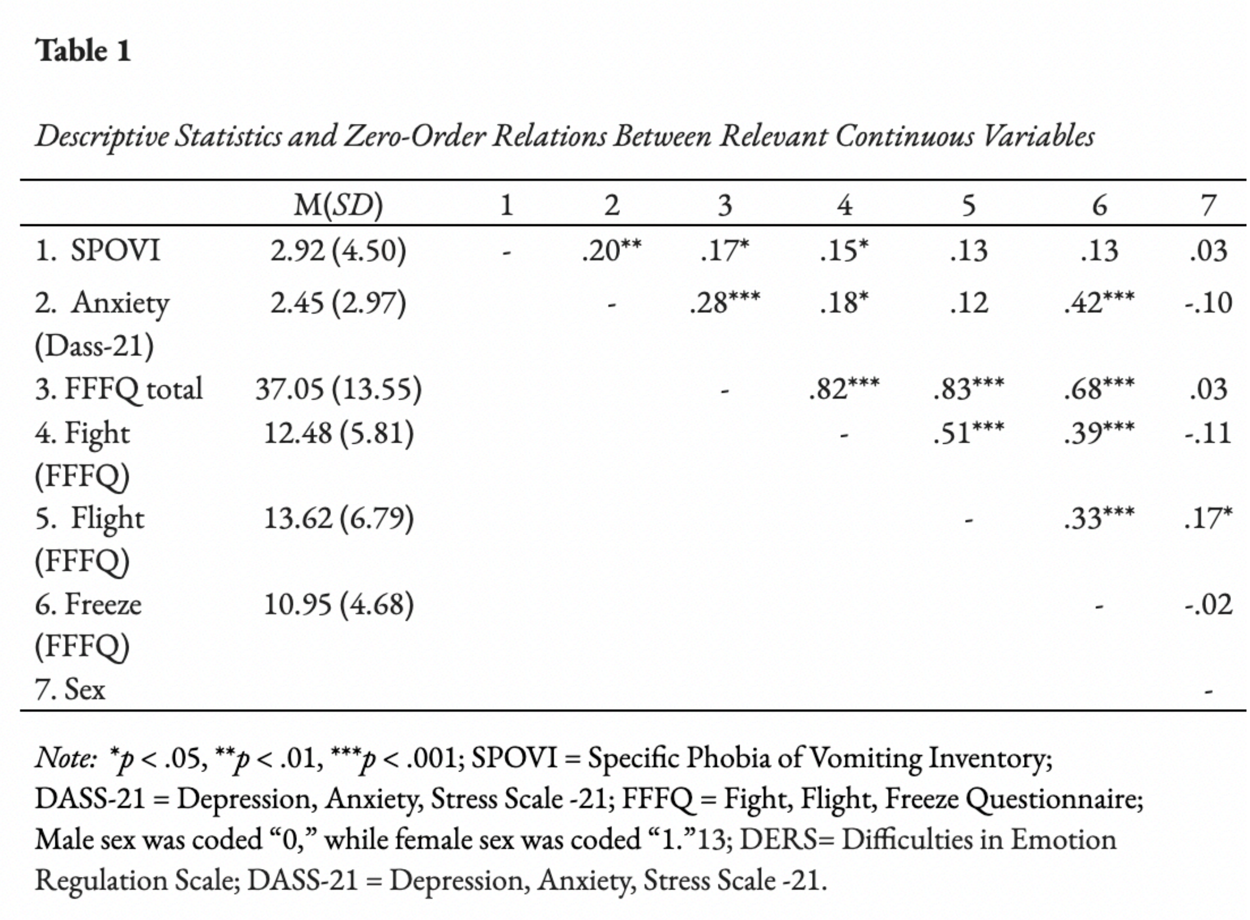


Table 1

Descriptive Statistics and Zero-Order Relations Between Relevant Continuous Variables

	M(*SD*)	1	2	3	4	5	6	7
1. SPOVI	2.92 (4.50)	-	.20**	.17*	.15*	.13	.13	.03
2. Anxiety (Dass-21)	2.45 (2.97)		-	.28***	.18*	.12	.42***	-.10
3. FFFQ total	37.05 (13.55)			-	.82***	.83***	.68***	.03
4. Fight (FFFQ)	12.48 (5.81)				-	.51***	.39***	-.11
5. Flight (FFFQ)	13.62 (6.79)					-	.33***	.17*
6. Freeze (FFFQ)	10.95 (4.68)						-	-.02
7. Sex								-

Note: *$p < .05$, **$p < .01$, ***$p < .001$; SPOVI = Specific Phobia of Vomiting Inventory; DASS-21 = Depression, Anxiety, Stress Scale -21; FFFQ = Fight, Flight, Freeze Questionnaire; Male sex was coded "0," while female sex was coded "1."13; DERS= Difficulties in Emotion Regulation Scale; DASS-21 = Depression, Anxiety, Stress Scale -21.

Table 2

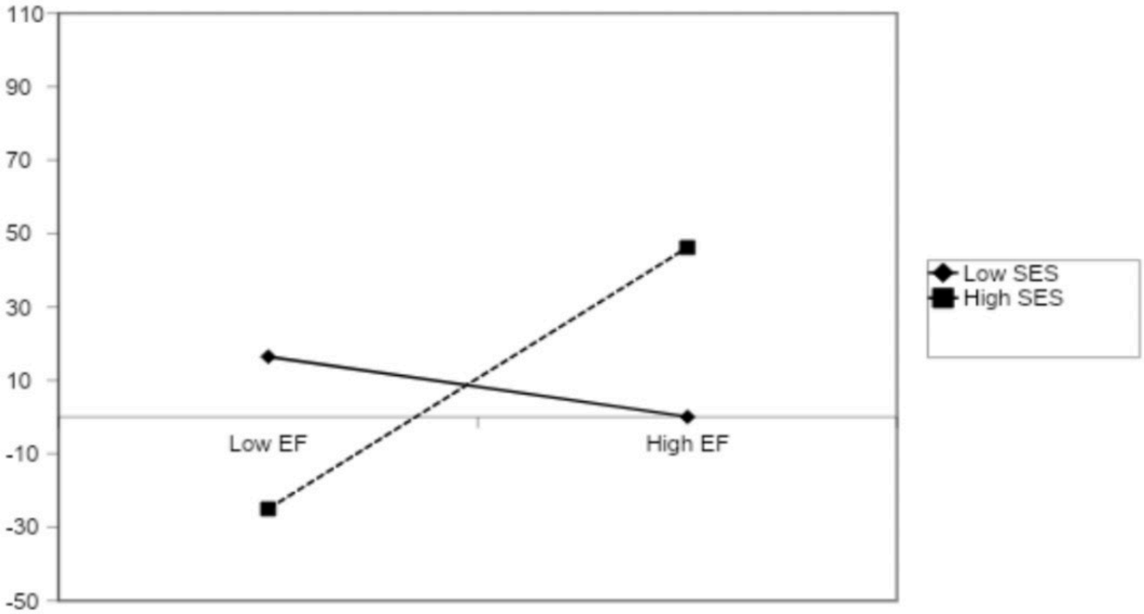
Summary of Regression Analysis with Predictors of SPOV Symptoms (N = 186)

Variable	*B*	95% CI for *B*		*SE B*	β	R^2	ΔR^2
		LL	UL				
Step 1						.02*	.02*
Constant	1.51*	-.02	3.04	.78			
Fight (FFFQ)	.11*	.00	.23	.06	.15*		
Step 2						.05**	.03**
Constant	1.12	-.38	2.70	.78			
Fight (FFFQ)	.09	-.02	.20	.06	.12		
Anxiety (DASS-21)	.27**	.05	.48	.11	.02**		

Note. **p* < .05; DASS-21 = Depression, Anxiety, Stress Scale -21; FFFQ = Fight, Flight, Freeze Questionnaire

Figure 1

Predicted Levels of SPOV Symptoms as a Result of the Interactive Effect of FFFQ (Temperamental Fear) and Sex



Note. SPOV = Specific Fear of Vomiting; FFFQ = Fight, Flight, Freeze, Questionnaire

Graduate Student Journal of Psychology
2022, Vol. 19

Effects of a Single-Session Intervention Targeting Perfectionism in College Students

Haley E. Ward and Michael G. Wheaton, Ph.D.,
Department of Psychology, Barnard College

Perfectionism is a well-established transdiagnostic factor that contributes to the maintenance and exacerbation of a wide range of psychopathology. Although effective, most treatments for perfectionism are inaccessible to the general population, and are time-, resource-, and cost-intensive. Online single-session interventions (SSIs) display promise in preventing and treating negative mental health symptoms; however, no studies have examined the effectiveness of an online SSI targeting perfectionism. To address this, the present study sought to evaluate a novel computerized, self-guided SSI. The two primary aims of the study were: 1) to assess the intervention's acceptability, likability, and utility, and 2) to examine whether the intervention reduces perfectionistic attitudes, depression, anxiety, and stress in a college student population. At baseline, students at a liberal arts college (N=79) were randomized to receive a 40-minute perfectionism SSI or a stress-management control. Participants rated the perfectionism intervention as highly helpful, acceptable, and likable. The perfectionism intervention was associated with greater improvements in perfectionism, depression, anxiety, and stress; however, the differences in reductions between the groups were not statistically significant. Given the small sample size, the study was likely underpowered to detect differences between conditions. Future research with larger sample sizes, a less potent control group, and clinical populations is needed to test the effects of the intervention more extensively. Nevertheless, this study demonstrates that single-session perfectionism interventions hold promise and merit future investigation.

Keywords: Perfectionism, Single-Session Interventions, Online Interventions, College Students

Perfectionism is defined as the strong need to perform at a flawless level and to meet excessively high standards (Flett & Hewitt, 2002). Although perfectionism is multidimensional and can have adaptive elements, it is considered a transdiagnostic factor associated with many forms of psychopathology (Shafran & Mansell, 2001; Egan, Wade, & Shafran, 2011). The type of perfectionism that contributes to psychopathology is known as maladaptive perfectionism, which is characterized by the pursuit of personally demanding standards and the basing of self-worth on performance outcomes in one or more important life domains, despite negative consequences (Shafran, Cooper, & Fairburn, 2002). Maladaptive perfectionism involves critical self-evaluations, concerns about making mistakes and being negatively evaluated by others, and fear of failure (Frost, Marten, Lahart, & Rosenblate, 1990; Egan, Wade, Shafran, & Antony, 2016). Many individuals with high levels of maladaptive perfectionism have core beliefs that they are weak, flawed, and unworthy (Besser, Flett, & Hewitt, 2004). Given these characteristics, it is unsurprising that perfectionism is elevated across eating (Limburg, Watson, Hagger, & Egan, 2017), anxiety (Gnilka, Ashby, & Noble, 2012), depressive (Hewitt & Flett, 1991), and personality disorders (Overholser & Dimaggio, 2020). Perfectionism has particular relevance to obsessive-compulsive personality disorder (OCPD), as one of the diagnostic criteria for OCPD involves perfectionism that interferes with normal task completion (American Psychological Association, 2013; Egan, Wade, & Shafran, 2011). Likewise, obsessive-compulsive disorder (OCD) often involves perfectionistic tendencies; the Obsessive-Compulsive Cognitions Working Group postulates that perfectionism is one of six central cognitive processes involved in obsessive-compulsive disorder (OCCWG, 1997).

In addition to the heightened perfectionism common among clinical populations, a substantial number of individuals who do not meet the diagnostic criteria for any given disorder have high levels of maladaptive perfectionism (Grzegorek et al., 2004). In nonclinical samples, perfectionism is highly correlated with subthreshold symptoms of various forms of psychopathology (Kearns, Forbes, & Gardiner, 2007). For instance, Ferrari (1995) found that perfectionism scores were significantly related to obsessional thoughts, compulsive acts, anger-suppression, anger expression, social desirability, and compulsive checking in a nonclinical college student population. Perfectionism levels are also associated with hopelessness, non-suicidal self-injury, and suicidal ideation and attempts (O'Connor, 2003; O'Connor, 2007). More broadly, perfectionism gives rise to high levels of stress and worry, behavioral impairments, physical health problems, impaired daily functioning, inferior academic performance, and interpersonal problems (Shafran et al., 2002; Flett & Hewitt, 2002).

Treatments for Perfectionism

Treatments for perfectionism are based on models that suggest cognitive and behavioral components may maintain the core biases and attitudes in perfectionism (Shafran et al., 2002). More specifically, maladaptive thinking patterns and cognitive biases such as dichotomous thinking, selective attention toward mistakes, rumination, self-criticism, self-blame, and overgeneralization play a role in maintaining perfectionism and its poor outcomes (Egan et al., 2016; Shafran et al., 2002). Such biases may contribute to procrastination behaviors, higher standard-setting, and inability to complete tasks (Egan et al., 2016; Sirois, 2014). Perfectionism is also antithetical to self-compassion (Brown, 2008; Nadeau, 2020), as internal self-dialogue is quite harsh and involves automatic perfectionism-themed thoughts about the self, judgments, and elements of self-neglect (Ferrari, 1995; Hewitt, 2020). Self-criticism may even mediate the relationship between perfectionism, psychological distress (James et al., 2015), and depression (Ferrari, Yap, Scott, Einstein, & Ciarrochi, 2018). Because of these cognitive and behavioral maintenance factors, a perfectionistic cycle is often quite resistant to change, thus some researchers suggest that perfectionism is a fixed part of personality (Hewitt et al., 2017). However, it might be more helpful to look at perfectionism as a cognitive vulnerability factor related to psychopathology that can change with targeted treatment (Egan et al., 2011). Indeed, perfectionistic beliefs can be changed through cognitive-behavioral therapy (CBT). Common elements of CBT target cognitive and behavioral processes in perfectionism via psychoeducation, cognitive restructuring, and self-compassion techniques (Egan et al., 2016; Nadeau, 2020).

A growing body of evidence shows that CBT for nonclinical and clinical populations reduces perfectionism, with medium to large effect sizes (Kearns et al., 2007; Lloyd, Schmidt, Khondoker, & Tchanturia, 2015). Targeting perfectionism may also lead to a reduction in other psychological disturbances characteristic of mental disorders (Egan et al., 2011; Lloyd et al., 2015). In particular, studies have found that depression, anxiety, and eating problems are often reduced with medium effect sizes in both clinical and nonclinical samples (e.g. Kearns, Forbes, & Gardiner, 2007; Steele & Wade, 2008; Shafran et al., 2017). There is also evidence supporting the utility of self-compassion interventions for perfectionism in nonclinical populations, although effects on reducing secondary disorders are less consistent (Rose, McIntyre, & Rimes, 2018; Nadeau, 2020). Therefore, combining more traditional cognitive-behavioral components with self-compassion techniques may be beneficial in alleviating perfectionism, as suggested in prior research (Fairweather-Schmidt & Wade, 2015).

Although there has been a recent surge in randomized controlled trials evaluating treatments for perfectionism (Lloyd et al., 2015), for many individuals psychotherapy interventions remain difficult to access given the face-to-face and individualized nature of treatment (Andersson, 2016). Cost, transportation issues, waiting lists, lack of trained mental health professionals, and stigma—typical barriers in facing treatment—might limit access to perfectionism interventions. And such barriers may be particularly strong for under-resourced populations (Andrade et al., 2014). Further, those with subclinical problems (e.g. someone who is perfectionistic and anxious but doesn't meet criteria for a disorder) would be unlikely to go to great lengths to seek out an intervention. This lack of access points to a clear need for more accessible mental health programs targeting perfectionism.

Internet-Based Interventions

As a response to lack of access, internet-based interventions have been rapidly developed. Internet-based interventions are effective for a variety of mental health disorders and problems (e.g. Parks et al., 2018), including perfectionism (e.g. Rozental et al., 2017). Online interventions for perfectionism utilize the same techniques present in in-person treatment (e.g. cognitive restructuring, self-compassion), are typically self-guided, and often report effect sizes comparable to in-person treatments (Suh et al., 2019). For example, Suh and colleagues' meta-analysis on 10 face-to-face and online interventions for perfectionism reported that there were no significant differences in delivery modality on effect sizes for perfectionism, depression, and anxiety (Suh et al., 2019). At present, there have been several randomized controlled trials evaluating online programs which have demonstrated great value in reducing the burden created by perfectionism (e.g. Egan et al., 2014).

However, one potential problem with both face-to-face and online perfectionism interventions is their length Lloyd and colleagues' (2015) meta-analysis of 8 perfectionism interventions reported that the number of treatment sessions ranged from 8 to 14, resembling the typical length of cognitive-behavioral therapy. Non-completion rates were high in many of the studies, which are reflective of treatment adherence norms; one in five adults will drop out of treatment before the recommended dose is complete (Olfson et al., 2009). Poor retention is even more common among online interventions. For instance, in the study conducted by Shafran and colleagues (2017), there was significant non-engagement and non-completion of modules, with 71% of participants completing fewer than half the modules and only 14.5% people completing five or more modules. Strikingly, even participants who had high rates of non-completion and non-engagement had reduced perfectionism. Change in perfectionism was largest at the beginning of the intervention, indicating that potential may lie in briefer, more compact interventions. This is consistent with other research reporting that the number of sessions is often unrelated to the magnitude of that treatment's effect (Weisz et al., 2017). Rozental and colleagues (2017) and Shafran et al (2017) noticed these effects in their 8-week long online interventions for perfectionism, and they suggested that brief, online interventions may have positive impacts on perfectionism and maximize scalability.

One form of brief intervention that has been developed in response to the high likelihood that an individual will access a treatment only once is single-session interventions (SSIs). SSIs are effective for adults (Campbell, 2012), youths (Schleider & Weisz, 2017), and college students (Samson & Tanner-Smith, 2015) with diagnosed psychiatric disorders and subclinical issues. SSIs have been demonstrated to improve anxiety, depression, problematic drug and alcohol use, and risk factors for mental health problems such as hopelessness, anxiety sensitivity, physiological stress self-hate, perceived control, fixed mindset, and agency (Tanner-Smith et al., 2015; Schleider, Dobias, Sung, Mumper, & Mullarkey, 2020; Schleider, Dobias, Sung, & Mullarkey, 2020). Gains are often maintained at follow-up, and participants typically rate single-session interventions as likable, acceptable, and useful (Duan & Bu, 2017; Wasil et al., 2021; Samson & Tanner-Smith et al., 2015; Schleider & Weisz, 2017). These interventions are particularly useful when delivered in an online, self-guided format because they can be made widely accessible, reducing the barriers to mental health care (Schleider et al, 2020a). Although short-term therapies have been shown to reduce perfectionistic tendencies and related problems and are comparable in effect sizes (Dodd et al., 2019; Lloyd et al., 2014; Fairweather-Schmidt & Wade, 2015), only one study has examined a single-session perfectionism intervention (LaSota, Ross, & Kearney, 2017), which lasted for several hours in-person and focused on psychoeducation, setting high standards, fear over mistakes, and reducing stress. To-date, no brief online perfectionism interventions have been tested. Given that online SSIs are effective in reducing mental health concerns (Schleider & Weisz, 2018b), and other types of brief interventions are effective in reducing perfectionism (e.g. Dodd et al., 2019; Fairweather-Schmidt & Wade, 2015), an online SSI aiming to decrease perfectionism might be particularly useful. The brevity and accessibility of such an intervention may amplify its potential impact, decreasing costs, time, and other problems associated with receiving a normal length and/or in-person intervention.

College students may be a good fit for a single-session intervention targeting perfectionism, considering that many students have problems with perfectionism, anxiety, and depression (LaSota et al., 2017), and mental health issues appear to be on the rise in college students (Curran & Hill, 2017). Perfectionism may be an etiological and maintenance factor of these mental health problems in college student populations for two primary reasons. First, work and studies are the two top domains of life in which people report being the most perfectionistic (Stoeber & Stoeber, 2009). Living in an academic environment could potentially lead to the exacerbation of perfectionistic tendencies. Second, college students are constantly evaluated based on their work and social skills and must meet certain self-presentation and performance-based standards (Klibert et al., 2014), which may contribute to the stress surrounding academic performance and the pressure to succeed that college students often report as a top stressor (Bedewey & Gabriel, 2015). An environment filled with excessive studying, academic pressure, and stress would undoubtedly influence perfectionism and may have particular relevance to highly competitive colleges that maintain a culture of perfectionism. Of relevance, Molnar and colleagues (2020) found that 14% of college students have extremely elevated levels of perfectionism, while other studies have found that two thirds can be categorized as perfectionists, with a quarter of those individuals meeting criteria for maladaptive perfectionism (Grzegorek et al., 2004). These rates have increased over the past decade, concurrent with rates of other mental health problems (Curran & Hill, 2017).

The increasing rates of perfectionism in college students may be contributing to the increase in mental illness among college populations, treatment-seeking behaviors, and other stressors (Pacewicz, Gotwals, and Blanton, 2018). Considering this information, there is an urgent need for brief, online interventions that reduce perfectionism and protect against the development of anxiety and depression among college students (Klibert et al., 2014). Targeting perfectionism in an online single-session intervention might help reduce the burden placed on college counseling centers, reduce distress among students, and prevent future dysfunction. Although traditional perfectionism programs are effective for various clinical and nonclinical concerns, no research to date has assessed the impact of a brief, online perfectionism program in a college population.

The Present Study

The present study extends prior research by conducting the first randomized controlled trial on a self-administered, online SSI designed to reduce perfectionism in a college student population. To assess the effectiveness of the program, which is based on psychoeducation, cognitive restructuring, and self-compassion, a randomized controlled trial was conducted comparing the program with a stress-management control. It was hypothesized that the online single session intervention will decrease perfectionism in college students, with secondary effects on depression, anxiety, and stress. Participants were also asked to rate the intervention's acceptability and utility and we hypothesized it would be rated as acceptable, likable, and useful.

Methods

Participants

Participants were undergraduate liberal arts college students who were enrolled in introductory psychology courses. Participants were recruited via an online platform (i.e. SONA) that allows students to sign up for studies in order to gain research credits. In total, 91 students participated in the present study; however, five participants were removed for failing both attention checks in each assessment or—admitting they did not watch the videos or give their best effort in completing the perfectionism intervention, leaving a total of 86 participants with usable pre-and post-data. Additionally, only 79 participants completed and passed the follow-up assessment. Participants were randomized to either the perfectionism (N=45) or control intervention (N=41). In total, participants were 89.5% female and 4.7% other, 80% were in their first and second year of college, 37.2% non-Hispanic White, 5.8% African American, 37.2% Asian/Pacific Islander, and 19.8% Other or Hispanic. As displayed in Table 1, the two groups did not differ on any demographic variables.

Materials

Dysfunctional Attitudes Scale-Perfectionism (DAS-PA; Weissman & Beck, 1978). The DAS was originally developed to study depression and has several subscales, including perfectionism, which was identified as a subscale by Beck and colleagues (1991). The subscale measures perfectionistic attitudes and has excellent internal consistency. Scores on the subscale are predictive of later depression and anxiety (Jacobs et al., 2009). The Likert Scale is a continuum from 0 (Fully Disagree) to 6 (Fully Agree).

Big Three Perfectionism Scale (BTPS; Smith, Saklofske, Stoeber, & Sherry, 2016). The BTPS evaluates three higher-order factors (rigid perfectionism, self-critical perfectionism, and narcissistic perfectionism) through 10 lower-order perfectionism facets (self-oriented perfectionism, self-worth contingencies, concern over mistakes, doubts about actions, self-criticism, socially prescribed perfectionism, other-oriented perfectionism, hypercriticism, grandiosity, entitlement). This scale was selected over older, more traditional perfectionism measures (MPS and FMPS) due to accessibility and its assessment of the perfectionistic attitudes targeted more directly in the intervention. Other-oriented perfectionism, socially prescribed perfectionism, and parental standards, which are assessed in other common measures of perfectionism, are not addressed in the intervention. Although the BTPS is new, it has strong psychometric properties (Smith et al., 2016). Participants respond on a Likert Scale from 0 (Disagree Strongly) to 4 (Agree Strongly).

Depression, Anxiety and Stress Scales-21 (DASS-21; Lovibond & Lovibond, 1995). The DASS-21 is a 21-item self-report questionnaire consisting of three subscales: depression, anxiety and stress. Ratings are provided using Likert Scales ranging from 0 (“Did not apply to me at all”) to 3 (``Applied to me very much or most of the time”). The instructions on the DASS-21 were aligned with the timeline of the study; participants were asked to rate their symptoms ‘in the past week.’ The subscales and total score have good reliability and construct validity in both clinical and nonclinical samples (Antony, Bieling, Cox, Enns, & Swinson, 1998; Henry & Crawford, 2005). DASS-21 scores are highly correlated with other measures of depression and anxiety (Osman et al., 2012) and perfectionism (Aldahadha, 2018).

Additional Questions about Perfectionism and the Intervention. To assess the intervention’s ability to alter metacognitive self-appraisals about perfectionism, the following three questions were asked at baseline: “Do you consider yourself to be a perfectionist?,” “Does perfectionism get in the way of your happiness,” and “Does perfectionism slow progress toward your goals?” Participants also completed several questions about the acceptability, likability, and perceived helpfulness of the intervention, which were partially derived from the Program Feedback Scale (PFS), a valid and reliable measure used to assess acceptability and perceptions of internet based SSIs (Sung, Mumper, & Schleider, 2021). They were asked, “How much did you like the intervention?,” “How much do you feel like the intervention helped you understand perfectionism?,” and “How helpful do you think the intervention was in teaching you to combat your perfectionism?” In addition, perceived changes in perfectionism were assessed at 1-week follow-up with the question: “Do you think that completing the program helped you with your perfectionism in the past week?”

Procedure

Approval for this study was received from the Institutional Review Board (IRB) at Barnard College. Upon signing up for the study, participants were directed to a Qualtrics form, where they completed a consent form and agreed to study duties. Study duties required the participant to fill out pre, post, and 1-week follow-up measures, in addition to actively engaging with the intervention. After providing consent, participants responded to a series of demographic questions and measures assessing perfectionism, depression, anxiety, and stress. Immediately following the completion of the measures, participants were randomly assigned to either the experimental or stress-management control condition through the Qualtrics randomizer element. The experimental intervention took approximately 40 minutes and required participants to maintain active participation through video and open-ended responding, while the stress-management condition lasted for a similar amount of time and consisted of a video and articles about stress. Attention checks were included to ensure participants were paying attention, along with a direct question following the intervention which asks participants if they did their best. Following the completion of the intervention, participants immediately filled out the measures for a second time, except for the Depression, Anxiety, and Stress Scales. Participants had access to the follow-up assessment 168 hours (seven days) after they completed their first assessment and received a reminder email to take the assessment. If they did not take the assessment within 24 hours of receiving the email, they were sent additional reminders. The follow-up assessment consisted of questions about perceived changes in perfectionism in addition to the same measures used at baseline. Participants were debriefed at the end of the study. Statistical analyses were conducted in SPSS. Independent samples t-tests were run to assess for differences between groups at baseline and a series of repeated measures, ANOVAs, were utilized to examine the effect of the perfectionism intervention compared to the stress-management control on the various outcome measures. Descriptive statistics on the likability and acceptability of the intervention were also analyzed.

Perfectionism Intervention: The intervention was based on a treatment guide for clinical perfectionism (Egan et al., 2016), which has been used to develop prior perfectionism interventions (e.g. Rozental et al., 2017; Shafran et al., 2017). The intervention was delivered through Qualtrics and consists of colorful text, images, and video clips. Participants can easily follow the intervention by clicking the arrows to the next page when they appear and writing in the text boxes provided. More specifically, the intervention included 4 major modules that take around 40 minutes total to complete and last for approximately 10 minutes each: 1) psychoeducation about perfectionism and its negative consequences, 2) cognitive restructuring of self-critical perfectionistic thoughts, 3) self-compassion techniques, and 4) generalization and maintenance of information learned in the intervention.

Psychoeducation and Consequences of Perfectionism: Given that perfectionism is considered ego-syntonic and perfectionists believe their high standards improve their functioning, it is important educate participants about the counterproductive nature and negative effects of perfectionism (Overholser & Dimaggio, 2020). By increasing awareness about the problems associated with perfectionism, clients may feel more motivated (Yeh et al., 2017) and be able to more easily change the thoughts and behaviors contributing to these outcomes (Kutlesa & Arthur, 2008). Therefore, the program defines perfectionism and discusses its consequences in the form of telling fictional stories about characters who are perfectionists. After learning what perfectionism is, participants learn the various ways perfectionism can manifest in life and its multidimensional nature. Throughout the psychoeducation module, they apply this information to their own lives.

Cognitive Restructuring and Automatic Thoughts. Cognitive-behavioral therapy can reduce the tendency for unconstructive automatic thoughts about the self (Overholser & Dimaggio, 2020). Interventions can help perfectionists become aware of, challenge, and change their automatic self-critical thought processes (Besser et al., 2004; Overholser & Dimaggio, 2020). Therefore, this section of the intervention explains how to combat the self-critical thoughts related to perfectionism through utilizing the three-step process (catch, challenge, and change) of CBT. Challenging and changing is explained in the context of self-compassion. The primary way participants are instructed to challenge their thoughts is by thinking about what a loving friend would say about their thoughts. Participants will practice catching, challenging, and changing self-critical thoughts that both a fictional character and they themselves experience.

Self-Compassion: Because low self-compassion is often characteristic of perfectionism (Brown, 2008), and interventions rooted in self-compassion are effective in reducing perfectionism (Nadeau, 2020), participants will be taught how to be more compassionate, forgiving, and kind to themselves. Self-compassion principles (Raes et al., 2011) are fused throughout the CBT section (e.g., mindfulness versus over-identification) and are also explained in the self-compassion module. Self-kindness over self-judgement is taught by explaining the tendency to focus on mistakes in perfectionism, the importance of forgiveness, and affirmation exercises. Next, common humanity versus isolation is taught through a discussion and exercise on how everybody makes mistakes and no one is alone in their suffering.

Concluding Components: The intervention ends by discussing how reducing perfectionism can improve one's life. Neurogenesis is explained to show that change is possible, although change requires commitment. Participants are encouraged to apply the cognitive distortion and self-compassion practices to everyday life.

Stress Management Program: The stress-management program was designed as an active control, for comparison with the perfectionism intervention, and did not contain any perfectionism components. Although different in structure, the course content was similar to stress management conditions used in prior studies (Hoge et al., 2013). Like the perfectionism intervention, the stress-management program lasted for around 40 minutes and was delivered through Qualtrics. Participants were instructed to watch a video and read excerpts describing the effects of stress and the various ways to manage stress.. The content included details about how chronic stress affects the mind, body, and performance, and explored ways to reduce stress, including exercise, nutrition, mindfulness, positive experiences, time management, and sleep.

Results

Baseline Scores and Group Comparisons

Table 1 shows the demographic characteristics for each group and the group mean scores of each of the outcome measures. According to independent samples t-tests, there were no significant differences between the groups on any demographic or clinical variables at baseline. Scores on the Big Three Perfectionism Scale and the Dysfunctional Attitudes Perfectionism Scale were elevated in both groups (M=55.7 on the DAS-PA; See Table 1) compared to previously studied populations, including a clinically depressed adolescent sample (M=53.7 on the DAS-PA; Jacobs et al., 2009). Total scores on the DASS-21 indicated mild to moderate levels of depression, anxiety, and stress (Lovibond & Lovibond, 1995) (See Table 1). The majority of the sample (95.3%) considered themselves to be a perfectionist to some degree, to a considerable degree, or very much, while 74.7% acknowledged that perfectionism sometimes gets in the way of their happiness, and 72.9% said that perfectionism sometimes slows progress towards goals.

Intervention Acceptability and Likability

Participants reported liking the intervention; 22.2% liked it "a little bit," 42.2% liked it "somewhat." and 35.6% liked it "very much," while no one reported disliking the intervention. Participants also indicated that the intervention was helpful in increasing their understanding of perfectionism; in terms of how helpful it was, 2.2% said "not at all," 6.7% said "a little bit," 42.2% reported "somewhat," and 48.9% selected "very." Perceived helpfulness in teaching participants to combat their perfectionism was also assessed, with 4.4% of participants saying it was not at all helpful, 8.9% saying it was a little bit helpful, 60% saying it was somewhat helpful, and 26.7% saying it was very helpful. At the 1-week follow-up, participants were asked, "Do you think that completing the program helped you with your perfectionism in the past week?" and 18.2% said "not at all," 25% selected "very little," 52.3% said "somewhat," and 4.5% said "very much." When asked "How much did you try to implement the techniques learned into your daily life?," most people reported putting some effort in; 6.8% said "not at all," 22.7% selected "very little," 65.9% said "somewhat," and 4.5% said "very much." Twenty-two participants left open-ended feedback, and 19 of these comments were positive and emphasized the utility or likable features of the intervention.

Change in Outcome Measures

Repeated measures ANOVAs were utilized to test the hypothesis that the perfectionism intervention would reduce perfectionistic attitudes and depression, anxiety, and stress more than the stress-management control. Differences were assessed at baseline, immediate post-test, and 1-week follow-up, except for DASS-21, which was only assessed at baseline and 1-week follow-up. The first ANOVA was conducted on the Dysfunctional Attitudes Perfectionism Subscale, which violated Mauchly's Test of Sphericity X2 (2)=25.924, p<.05, thus the degrees of freedom was adjusted for by reporting the Greenhouse-Geisser results. The effect of time was significant (F (1.54, 115.782) = 6.145, p<.05, n2= .076); however, the time*condition interaction effect was not (F, 1.54, 115.782) = 1.935, p>.05, n2=.025) (See Figure 1). Similarly, Mauchly's Test of Sphericity indicated that the assumption of sphericity was violated for the changes in the Big Three Perfectionism Scale (X2 (2)=38.864, p<.05. After using the Greenhouse-Geiser correction, the effect was significant for time F(1.327, 74.335)=4.685, p<.05, n2=.077, but not for time*condition F(1.327, 74.335)=.072, p>.05, n2=.001) (See Figure 2). Finally a repeated measures ANOVA of DASS total scores at baseline and at the 1-week follow-up revealed that there were significant differences for time (F (1, 74) =185.872, p<.05, n2=.715), but no effect for condition (F (1,74)=188.943, p>.05, n2=.011). Subscale scores were not analyzed due to the lack of significant differences found for time and time*condition on the DASS total scores.

Discussion

The present study sought to examine the effectiveness of a single-session intervention in improving perfectionism and associated mental health characteristics in a college-student population. Although symptoms improved for all clinical measures following the perfectionism intervention, they similarly improved in the control condition. These findings are discussed in detail below.

There are several potential reasons for the lack of significant differences between the intervention and control conditions in the study. Small sample size may have contributed to the null findings. The decreases were fairly large for some measures, yet statistical significance was not achieved. For instance, the mean score dropped from 55.31 at baseline to 48.01 at 1-week follow-up on the Dysfunctional Attitudes Perfectionism Scale in the experimental group, which is almost a seven-point decrease and a decline of three points greater than the stress-management control. A post-hoc sensitivity analysis with the study's sample size (N=79) suggested we had adequate power (.80) to detect an effect size of .10 or larger, which is slightly larger than a medium effect (partial eta squared = .9). Therefore, our sample size was likely underpowered if the intervention had a small to medium effect. Although traditional 8-week perfectionism interventions have medium to large effect sizes, the extremely brief nature of the intervention makes a larger effect size less probable. Additionally, although there were 79 participants who passed attention checks and submitted each assessment, many participants skipped questions. This led to fewer participants' scores being analyzed in each

ANOVA; for example, only 58 participants' scores were analyzed for the repeated measures ANOVA for the BTPS. The majority of randomized controlled trials conducted on online single-session interventions have included hundreds of participants (e.g., Schleider & Weisz, 2017), and the in-person SSI conducted on perfectionism included 105 participants in their data analysis with no control group (LaSota et al., 2017). Future work with larger samples is needed to further compare these conditions. In addition, it is possible that the single dose of perfectionism intervention was not sufficient to demonstrate superiority over stress management. Most trials conducted on treatments for perfectionism have ranged from 8-16 sessions in length (LaSota et al., 2017), thus more practice and intervention time may be necessary to create score decreases greater than what the stress-management control produced. Relatedly, although one-week follow-ups have been used in studies of single session interventions in the past (e.g., Davidson, Malloch, & Humphris, 2018; Duan & Bu, 2019), a longer follow-up after the intervention would have provided participants with more time to implement practices they learned in their lives and utilize strategies more frequently. Therefore, future studies could include 3-, 6-, or 12-month follow-ups and also explore whether repeated sessions and practice post-treatment contribute to reductions in perfectionism symptoms. Although there is nothing in the student timeline or a widespread stressful event that would have accounted for a decline in perfectionism, depression, anxiety, and stress from baseline to one-week follow-up, including follow-up assessments at various time points in future research would allow for a more comprehensive understanding of changes in perfectionism over time.

It is also plausible that the stress-management control condition represented an active intervention, suggesting that it was not necessarily an appropriate placebo control, but may provide information about the required specificity of perfectionism interventions. Perfectionistic behavior is thought to be associated with psychopathology, in part due to its association with stress.. Perfectionists have higher levels of stress exposure and stress reactivity (i.e.i.e., maladaptive ways of reacting to stressors), and they may generate or exacerbate stress for themselves by engaging in self-critical thinking in response to daily stressors and failing to reach high standards. If stressful events are reduced, it is likely that perfectionistic behavior and automatic perfectionistic thoughts will subsequently decrease, and when perfectionistic attitudes decrease, stress levels will fall, indicating a bidirectional relationship (Hewitt & Flett, 2002). The results of the present study are similar to a study that had compared face-to-face CBT for perfectionism to a stress-management condition; this study investigated the effectiveness of an online 10-week CBT intervention for perfectionism in first-year psychology students and randomly assigned participants to either stress management, stress-management + CBT, or a control condition (Arpin-Cribbie et al., 2008). The techniques taught in their stress-management condition overlapped with some of the techniques presented in the stress-management control of the current intervention. In line with our results, participants in the stress-management condition had significant decreases in self-oriented perfectionism and concern over mistakes, although the perfectionism intervention and stress-management combined led to greater decreases in the perfectionism construct. The results from Arpin-Cribbie (2008) and the present study may indicate that stress management represents an active treatment for perfectionism and works well in combination with more targeted content. Thus, future research should further explore the complex relationship between perfectionism and stress, compare perfectionism interventions to other types of active controls, and work to elucidate the importance of specificity in perfectionism interventions.

The intervention should also be tested in different populations; one of the main limitations to the study is that the perfectionism intervention was designed for high levels of maladaptive perfectionism, yet it was tested in a typical nonclinical college-student population. Although participants had higher than average levels of perfectionism, prior research indicates that perfectionism interventions, including single-session interventions may not be effective for those with low perfectionism. LaSota and colleagues (2017) focused on psychoeducation, setting high standards, fear over mistakes, and reducing stress in their in-person single-session intervention, dividing participants into high, medium, and low perfectionists. They found that participants with low perfectionism showed no significant change over time, while the other two groups did. Many perfectionism treatment studies have also utilized clinical samples (e.g., eating disorder or depressed populations) with high levels of maladaptive perfectionism (e.g., Kearns, Forbes, & Gardiner, 2007). Additionally, most single-session interventions have been tested in adolescents (Schleider et al., 2020) and previous experience with psychoeducation about perfectionism or CBT for perfectionism could potentially limit its effectiveness. This was highlighted by one participant who wrote, “The intervention was likely not a good match for me as I have undergone CBT in the past for anxiety and perfectionism and already knew a lot of these techniques;” thus, those who have not undergone therapy in the past are younger and more capable of change—potentially making themselves better candidates. In addition to age, other demographic factors, such as gender, may interact with perfectionism (Gnilka & Novakovic, 2017) and impact treatment response. Considering that the present sample was almost 90% female and previous literature on perfectionism consists of majority female participants (Stoeber & Stoeber, 2009), future research could aim to recruit more males to better elucidate the impact of gender on effectiveness of perfectionism treatments. Further research with larger sample sizes, especially research that includes populations with clinical perfectionism, is warranted.

Despite the study’s limitations, many participants appear to have benefited from the intervention, as respondents reported that the intervention was highly acceptable, engaging, and useful. All participants reported liking the intervention “at least a little bit” and 86.7% of participants thought the program was somewhat or very helpful. Additionally, 57% of students thought it was somewhat or very helpful in combating their perfectionism in the week following the intervention and 93% of respondents attempted to implement the learned skills into their daily lives. To conclude, this study is the first to examine the effectiveness of an online single-session intervention in reducing perfectionism, depression, anxiety, and stress. Although prior literature was utilized to formulate hypotheses that the perfectionism intervention would be more effective than a stress-management control, the results only revealed a significant effect for time. The study was likely underpowered to detect differences between conditions, had limitations in that the sample contained many individuals with low perfectionism, and the stress-management condition was perhaps effective at reducing stress, thus indirectly impacting perfectionism. Future research with larger sample sizes, various types of control groups, and clinical populations is needed to more extensively test the effects of the intervention. Nevertheless, this study adds important information about the effectiveness of a perfectionism intervention when compared to a stress-management control, and it was rated as highly useful, likable, and acceptable by participants.

References

Aldahadha, B. (2018). The psychometric properties of perfectionism scale and its relation to depression and anxiety. *Cogent Psychology, 5(1),* 1524324.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.).

Andersson, G. (2016). Internet-delivered psychological treatments. *Annual Review of Clinical Psychology, 12(1),* 157–179. doi:https://doi.org/10.1146/annurev-clinpsy-021815-093006

Andrade, L. H., Alonso, J., Mneimneh, Z., Wells, J. E., Al-Hamzawi, A., Borges, G., ... & Florescu, S. (2014). Barriers to mental health treatment: results from the WHO World Mental Health (WMH) Surveys. *Psychological medicine, 44(6),* 1303. doi: https://doi.org/10.1017/s0033291713001943

Antony, M. M., Bieling, P. J., Cox, B. J., Enns, M. W., & Swinson, R. P. (1998). Psychometric properties of the 42-item and 21-item versions of the Depression Anxiety Stress Scales (DASS) in clinical groups and a community sample. P*sychological Assessment, 10,* 176–181. doi: 10.1037/1040-3590.10.2.176

Arpin-Cribbie, C.A., Irvine, J., & Ritvo, P (2008). Perfectionism and Psychological Distress: A Modeling Approach to Understanding their Therapeutic Relationship. *Journal of Rational-Emotive Cognitive-Behavioral Therapy. 26,* 151–167. doi: 10.1007/s10942-007-0065-2

Bedewy, D., & Gabriel, A. (2015). Examining perceptions of academic stress and its sources among university students: The Perception of Academic Stress Scale. *Health psychology open, 2(2).* doi: 10.1177/2055102915596714

Besser, A., Flett, G., & Hewitt, P. (2004). Perfectionism, cognition, and affect in response to performance failure vs. success. *Journal of Rational-Emotive and Cognitive-Behavior Therapy, 22(4),* 301–328.

doi:10.1023/B:JORE.0000047313.35872.5c

Besser, A., Flett, G., Sherry, S., & Hewitt, P. (2020). Are perfectionistic thoughts an antecedent or a consequence of depressive symptoms? A cross-lagged analysis of the Perfectionism Cognitions Inventory. *Journal of Psychoeducational Assessment, 38(1),* 99–111. doi: 10.1177/0734282919877764

Blatt, S. J., Zuroff, D. C., Bondi, C. M., Sanislow, C. A. III, & Pilkonis, P. A. (1998). When and how perfectionism impedes the brief treatment of depression: Further analyses of the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *Journal of Consulting and Clinical Psychology, 66(2),* 423–428. doi:10.1037//0022-006x.66.2.423

Brown, B. (2008). *I Thought It Was Just Me (But It Isn't): Telling the Truth about Perfectionism.*

Campbell, A. (2012). Single-session approaches to therapy: Time to review. *Australian and New Zealand Journal of Family Therapy, 33(1),* 15-26. doi:10.1017/aft.2012.3

Center for Collegiate Mental Health. (2016). 2015 a*nnual report*. https://ccmh.psu.edu/annual-reports

Crandall, C. S., Preisler, J. J., & Aussprung, J. (1992). Measuring life event stress in the lives of college students: The Undergraduate Stress Questionnaire (USQ). *Journal of behavioral medicine, 15(6),* 627-662. doi:10.1007/BF00844860

Curran, T., & Hill, A. P. (2019). Perfectionism is increasing over time: A meta-analysis of birth cohort differences from 1989 to 2016. *Psychological Bulletin, 145(4),* 410. doi: 10.1037/bul0000138

Davidson, J., Malloch, M., & Humphris, G. (2018). A single-session intervention (the Mini-AFTERc) for fear of cancer recurrence: A feasibility study. *Psycho-oncology, 27(11),* 2668–2670. https://doi.org/10.1002/pon.4724

De Graaf, L. E., Roelofs, J., & Huibers, M. J. (2009). Measuring dysfunctional attitudes in the general population: the dysfunctional attitude scale (form A) revised. *Cognitive therapy and research, 33(4),* 345. doi: 10.1007/s10608-009-9229-y

Dodd, D. R., Parsons, E. M., Clerkin, E. M., Forrest, L. N., Velkoff, E. A., Kunstman, J. W., & Smith, A. R. (2019). Perfectly imperfect: The use of cognitive bias modification to reduce perfectionism. *Journal of behavior therapy and experimental psychiatry, 64,* 167-174. doi: 10.1016/j.jbtep.2019.04.002

Dozois D.J., Seeds P.M., & Collins K.A. (2009). Transdiagnostic approaches to the prevention of depression and anxiety. *Journal of Cognitive Psychotherapy, 23(1),* 44-59. doi: 10.1891/0889-8391.23.1.44

Duan, W., & Bu, H. (2019). Randomized trial investigating of a single-session character-strength-based cognitive intervention on freshman's adaptability. *Research on Social Work Practice, 29(1),* 82-92. doi: 10.1177/1049731517699525

Egan, S. J., van Noort, E., Chee, A., Kane, R. T., Hoiles, K. J., Shafran, R., & Wade, T. D. (2014). A randomised controlled trial of face to face versus pure online self-help cognitive behavioural treatment for perfectionism. *Behaviour Research and Therapy, 63,* 107-113. doi: 10.1016/j.brat.2014.09.009

Egan, S. J., Wade, T. D., & Shafran, R. (2011). Perfectionism as a transdiagnostic process: A clinical review. *Clinical psychology review, 31(2),* 203-212. doi: 10.1016/j.cpr.2010.04.009

Egan, S. J., Wade, T. D., Shafran, R., & Antony, M. M. (2016). *Cognitive-behavioral treatment of perfectionism*. Guilford Publications.

Fairweather-Schmidt, A. K., & Wade, T. D. (2015). Piloting a perfectionism intervention for pre-adolescent children. *Behaviour research and therapy, 73,* 67–73. doi: 10.1016/j.brat.2015.07.004

Ferrari, J. R. (1995). Perfectionism cognitions with nonclinical and clinical samples. Journal of Social *Behavior & Personality, 10(1),* 143–156. https://psycnet.apa.org/record/1995-31523-001

Ferrari, M., Yap, K., Scott, N., Einstein, D. A., & Ciarrochi, J. (2018). Self-compassion moderates the perfectionism and depression link in both adolescence and adulthood. *PloS one, 13(2)*. doi: 10.1371/journal.pone.0192022

Fisak, B. J., Richard, D., & Mann, A. (2011). The prevention of child and adolescent anxiety: A meta-analytic review. *Prevention Science, 12(3),* 255-268. doi: 10.1007/s11121-011-0210-0

. *Morbidity and Mortality Weekly Report, 67(23),* 653–658. https://doi.org/10.15585/mmwr.mm6723a1

Flett, G. L., Hewitt, P. L., Blankstein, K. R., & Mosher, S. W. (1995). Perfectionism, life events, and depressive symptoms: A test of a diathesis-stress model. *Current Psychology, 14(2),* 112-137. doi: 10.1007/BF02686885

Frost, R. O., & Marten, P. A. (1990). Perfectionism and evaluative threat. *Cognitive Therapy and Research, 14,* 559-572. doi: 10.1007/BF01173364

Frost, R. O., Marten, P., Lahart, C., & Rosenblate, R. (1990). The dimensions of perfectionism. C*ognitive Therapy and Research,14,* 449–468. doi: 10.1007/BF01172967

Gardenswartz, C. A., & Craske, M. G. (2001). Prevention of panic disorder. *Behavior Therapy, 32(4),* 725-737. doi: 10.1016/S0005-7894(01)80017-4

Gnilka, P. B., Ashby, J. S., & Noble, C. M. (2012). Multidimensional perfectionism and anxiety: Differences among individuals with perfectionism and tests of a coping-mediation model. *Journal of Counseling & Development, 90(4),* 427–436. doi: 10.1002/j.1556-6676.2012.00054.x

Gnilka, P. B., & Novakovic, A. (2017). Gender differences in STEM students' perfectionism, career search self-efficacy, and perception of career barriers. *Journal of Counseling & Development, 95(1),* 56-66. doi: 10.1002/jcad.12117

Grzegorek, J. L., Slaney, R. B., Franze, S., & Rice, K. G. (2004). Self-Criticism, Dependency, Self-Esteem, and Grade Point Average Satisfaction Among Clusters of Perfectionists and Nonperfectionists. *Journal of Counseling Psychology, 51(2),* 192. doi: 10.1037/0022-0167.51.2.192

Harvey, A. G., Watkins, E., & Mansell, W. (2004). *Cognitive behavioural processes across psychological disorders: A transdiagnostic approach to research and treatment.* Oxford University Press, Oxford.

Henry, J. D., & Crawford, J. R. (2005). The short-form version of the Depression Anxiety Stress Scales (DASS-21): Construct validity and normative data in a large non-clinical sample. British *journal of clinical psychology, 44(2),* 227-239. doi: 10.1348/014466505X29657

Hewitt, P. L., & Flett, G. L. (1991). Dimensions of Perfectionism in Unipolar Depression. *Journal of Abnormal Psychology, 100(1),* 98–101. doi: 10.1037//0021-843x.100.1.98

. *Morbidity and Mortality Weekly Report, 67(23),* 653–658. https://doi.org/10.15585/mmwr.mm6723a1Hewitt, P. L., Flett, G. L., & Mikail, S. F. (2017). *Perfectionism: Conceptualization, assessment, and treatment.* New York, NY: Guilford.

Hewitt, P. L. (2020). Perfecting, belonging, and repairing: A dynamic-relational approach to perfectionism. *Canadian Psychology, 61(2),* 101. doi: 10.1037/cap0000209

Hoge, E. A., Bui, E., Marques, L., Metcalf, C. A., Morris, L. K., Robinaugh, D. J., Worthington, J. J., Pollack, M. H., & Simon, N. M. (2013). Randomized controlled trial of mindfulness meditation for generalized anxiety disorder: effects on anxiety and stress reactivity. T*he Journal of clinical psychiatry, 74(8)*, 786–792. doi: 10.4088/JCP.12m08083

James, K., Verplanken, B., & Rimes, K. A. (2015). Self-criticism as a mediator in the relationship between unhealthy perfectionism and distress. *Personality and Individual Differences, 79,* 123-128. doi: 10.1016/j.paid.2015.01.030

Jacobs, R. H., Silva, S. G., Reinecke, M. A., Curry, J. F., Ginsburg, G. S., Kratochvil, C. J., & March, J. S. (2009). Dysfunctional attitudes scale perfectionism: A predictor and partial mediator of acute treatment outcome among clinically depressed adolescents. *Journal of Clinical Child & Adolescent Psychology, 38(6),* 803-813. doi: 10.1080/15374410903259031

Kearns, H., Forbes, A., & Gardiner, M. (2007). A cognitive behavioural coaching intervention for the treatment of perfectionism and self-handicapping in a nonclinical population. *Behaviour Change, 24(3),* 157. doi: 10.1375/bech.24.3.157

Klibert, J., Lamis, D. A., Collins, W., Smalley, K. B., Warren, J. C., Yancey, C. T., & Winterowd, C. (2014). Resilience mediates the relations between perfectionism and college student distress. *Journal of Counseling & Development, 92(1),* 75-82. doi: 10.1002/j.1556-6676.2014.00132.x

Kutlesa, N., & Arthur, N. (2008). Overcoming negative aspects of perfectionism through group treatment. *Journal of Rational-Emotive and Cognitive-Behavior Therapy, 26,* 134–150. doi: 10.1007/s10942-007-0064-3

LaSota, M. T., Ross, E. H., & Kearney, C. A. (2017). A cognitive-behavioral-based workshop intervention for maladaptive perfectionism. *Journal of Rational-Emotive & Cognitive-Behavior Therapy, 35(3),* 314-328. doi: 10.1007/s10942-017-0261-7

Limburg, K., Watson, H. J., Hagger, M. S., & Egan, S. J. (2017). The relationship between perfectionism and psychopathology: A meta-analysis. *Journal of clinical psychology, 73(10),* 1301-1326. doi: 10.1002/jclp.22435

Lloyd, S., Schmidt, U., Khondoker, M., & Tchanturia, K. (2015). Can psychological interventions reduce perfectionism? A systematic review and meta-analysis. *Behavioural and cognitive psychotherapy, 43(6),* 705-731. doi: 10.1017/S1352465814000162

Lovibond, P. F., & Lovibond, S. H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour research and therapy, 33(3),* 335-343. doi: 10.1016/0005-7967(94)00075-u

Lundh, L. G., & Öst, L. G. (2001). Attentional bias, self-consciousness and perfectionism in social phobia before and after cognitive-behaviour therapy. *Scandinavian Journal of Behaviour Therapy, 30(1),* 4-16. doi: 10.1080/02845710117841

Molnar, D., Sirois, F., Flett, G., & Sadava, S. (2020). A person-oriented approach to multidimensional perfectionism: Perfectionism profiles in health and well-being. *Journal of Psychoeducational Assessment, 38(1),* 127–142. doi: 10.1177/0734282919877754

Nadeau, M. M., Caporale-Berkowitz, N. A., & Rochlen, A. B. (2020). Improving women's self-compassion through an online program: A randomized controlled trial. J*ournal of Counseling & Development, 99(1),* 47-59. doi: 10.1002/jcad.12353

Nehmy, T. J., & Wade, T. D. (2015). Reducing the onset of negative affect in adolescents: Evaluation of a perfectionism program in a universal prevention setting. *Behaviour Research and Therapy, 67,* 55-63. doi: 10.1016/j.brat.2015.02.007

Obsessive Compulsive Cognitions Working Group. (1997). Cognitive assessment of obsessive compulsive disorder. *Behaviour Research and Therapy, 35,* 7. doi:10.1016/S0005-7967(97)00017-X

O'Connor, R., & O'Connor, D. (2003). Predicting hopelessness and psychological distress: The role of perfectionism and coping. *Journal of Counseling Psychology, 50(3),* 362–372. doi:10.1037/0022-0167.50.3.362

O'Connor, R. (2007). The relations between perfectionism and suicidality: A systematic review. Suicide and Life-Threatening *Behavior, 37(6),* 698–714. doi: 10.1521/suli.2007.37.6.698

Olfson, M., Mojtabai, R., Sampson, N. A., Hwang, I., Druss, B., Wang, P. S., ... & Kessler, R. C. (2009). Dropout from outpatient mental health care in the United States. *Psychiatric Services, 60(7),* 898-907. doi: 10.1176/appi.ps.60.7.898

Osman, A., Wong, J. L., Bagge, C. L., Freedenthal, S., Gutierrez, P. M., & Lozano, G. (2012). The depression anxiety stress Scales—21 (DASS-21): further examination of dimensions, scale reliability, and correlates. *Journal of clinical psychology, 68(12),* 1322-1338. doi:10.1002/jclp.21908

Overholser, J., & Dimaggio, G. (2020). Struggling with perfectionism: When good enough is not good enough. *Journal of Clinical Psychology, 76(11),* 2019-2027. doi: 10.1002/jclp.23047

Pacewicz, C. E., Gotwals, J. K., & Blanton, J. E. (2018). Perfectionism, coping, and burnout among intercollegiate varsity athletes: A person-oriented investigation of group differences and mediation. *Psychology of Sport and Exercise, 35,* 207-217. doi: 10.1016/j.psychsport.2017.12.008

Parks AC, Williams A, Tugade M, Hokes K, Honomichl R, Zilca R. (2018). Testing a scalable web and smartphone based intervention to improve depression, anxiety, and resilience: A randomized controlled trial. International Journal of *Wellbeing, (2),* 22-67. https://internationaljournalofwellbeing.org/index.php/ijow/article/download/745/677/3641

Raes, F., Pommier, E., Neff, K. D., & Van Gucht, D. (2011). Construction and factorial validation of a short form of the Self-Compassion Scale. *Clinical Psychology & Psychotherapy. 18,* 250-255. doi: 10.1002/cpp.702

Rose, A., McIntyre, R., & Rimes, K. A. (2018). Compassion-focused intervention for highly self-critical individuals: Pilot study. *Behavioural and cognitive psychotherapy, 46(5),* 583-600. doi: 10.1017/S135246581800036X

Rozental, A., Shafran, R., Wade, T., Egan, S., Nordgren, L. B., Carlbring, P., ... Andersson, G. (2017). A randomized controlled trial of Internet-Based Cognitive Behavior Therapy for perfectionism including an investigation of outcome predictors. *Behaviour Research and Therapy, 95,* 79–86. doi: 10.1016/j.brat.2017.05.015

Samson, J. E., & Tanner-Smith, E. E. (2015). Single-session alcohol interventions for heavy drinking college students: A systematic review and meta-analysis. *Journal of Studies on Alcohol and Drugs, 76(4),* 530-543. doi: 10.15288/jsad.2015.76.530

Schleider, J. L., Dobias, M., Sung, J., Mumper, E., & Mullarkey, M. C. (2020). *Acceptability and Utility of an Open-Access, Online Single-Session Intervention Platform for Adolescent Mental Health.* doi: 10.2196/20513

Schleider JL, Dobias ML, Sung JY, Mullarkey MC. (2020). Future Directions in Single-Session Youth Mental Health Interventions. *Journal of Clinical and Child Adolescent Psychology, 49(2):*264–278. doi: 10.1080/15374416.2019.1683852

Schleider, J. L., & Weisz, J. R. (2016). Reducing risk for anxiety and depression in adolescents: Effects of a single-session intervention teaching that personality can change. *Behaviour research and therapy, 87,* 170-181. doi: 10.1016/j.brat.2016.09.011

Schleider, J. L., & Weisz, J. R. (2017). Little treatments, promising effects? Meta-analysis of single-session interventions for youth psychiatric problems. *Journal of the American Academy of Child & Adolescent Psychiatry, 56(2)*, 107-115. doi: 10.1016/j.jaac.2016.11.007

Shafran, R., Cooper, Z., & Fairburn, C. G. (2002). Clinical perfectionism: A cognitive–behavioural analysis. *Behaviour research and therapy, 40(7),* 773-791. doi: 10.1016/s0005-7967(01)00059-6

Shafran, R., & Mansell, W. (2001). Perfectionism and psychopathology: A review of research and treatment. *Clinical Psychology Review, 21(6),* 879–906. doi: 10.1016/S0272-7358(00)00072-6

Shafran, R., Wade, T. D., Egan, S. J., Kothari, R., Allcott-Watson, H., Carlbring, P., ... & Andersson, G. (2017). Is the devil in the detail? A randomised controlled trial of guided internet-based CBT for perfectionism. B*ehaviour Research and Therapy, 95,* 99-106. doi: 10.1016/j.brat.2017.05.014

Sirois, F. M. (2014). Procrastination and stress: Exploring the role of self-compassion. *Self and Identity, 13(2),* 128-145. doi: 10.1080/15298868.2013.763404

Smith, M. M., Saklofske, D. H., Stoeber, J., & Sherry, S. B. (2016). The big three perfectionism scale: A new measure of perfectionism. *Journal of Psychoeducational Assessment, 34(7),* 670-687. doi: 10.1177/0734282916651539

Steele, A. L., & Wade, T. D. (2008). A randomised trial investigating guided self-help to reduce perfectionism and its impact on bulimia nervosa: A pilot study. *Behaviour Research and Therapy, 46(12),* 1316-1323. doi: 10.1016/j.brat.2008.09.006

Stoeber, J., & Stoeber, F. S. (2009). Domains of perfectionism: Prevalence and relationships with perfectionism, gender, age, and satisfaction with life. *Personality and individual differences, 46(4),* 530-535. doi: 10.1016/j.paid.2008.12.006

Stober, J. (1998). The Frost Multidimensional Perfectionism Scale: More perfect with four (instead of six) dimensions. *Personality and Individual Differences, 24(4),* 481-491. https://core.ac.uk/download/63723.pdf

Suh, H., Sohn, H., Kim, T., & Lee, D. G. (2019). A review and meta-analysis of perfectionism interventions: Comparing face-to-face with online modalities. J*ournal of Counseling Psychology, 66(4),* 473. doi:10.1037/cou0000355

Sung, J. Y., Mumper, E., & Schleider, J. L. (2021). Empowering Anxious Parents to Manage Child Avoidance Behaviors: Randomized Control Trial of a Single-Session Intervention for Parental Accommodation. *JMIR mental health, 8(7),* e29538. doi:10.2196/29538

Tanner-Smith, E. E., Steinka-Fry, K. T., Hennessy, E. A., Lipsey, M. W., & Winters, K. C. (2015). Can brief alcohol interventions for youth also address concurrent illicit drug use? Results from a meta-analysis. *Journal of youth and adolescence, 44(5),* 1011-1023. doi: 10.1007/s10964-015-0252-x

Wasil, A. R., Taylor, M. E., Franzen, R. E., Steinberg, J. S., & DeRubeis, R. J. (2021). Promoting Graduate Student Mental Health During COVID-19: Acceptability, Feasibility, and Perceived Utility of an Online Single-Session Intervention. *Frontiers in psychology, 12,* 569785. doi: 10.3389/fpsyg.2021.569785

Weissman, A. N., & Beck, A. T. (1978). *Development and validation of the Dysfunctional Attitude Scale: A preliminary investigation.* https://eric.ed.gov/?id=ED167619

Weisz, J. R., Kuppens, S., Ng, M. Y., Eckshtain, D., Ugueto, A. M., Vaughn-Coaxum, R., ... & Weersing, V. R. (2017). What five decades of research tells us about the effects of youth psychological therapy: a multilevel meta-analysis and implications for science and practice. *American Psychologist, 72(2),* 79. doi: 10.1037/a0040360

Wheaton, M. G., Pinto, A. M., Cervoni, C., Crosby, J. M., Tifft, E. D., Mathes, B. M., ... & Pinto, A.

(2020). Perfectionism in Intensive Residential Treatment of Obsessive–Compulsive Disorder. *Cognitive Therapy and Research, 44(1),* 136-144. doi: 10.1007/s10608-019-10047-6

Yeh, M. Y., Tung, T. H., Horng, F. F., & Sung, S. C. (2017). Effectiveness of a psychoeducational programme in enhancing motivation to change alcohol-addictive behaviour. J*ournal of Clinical Nursing, 26(21–22)*, 3724–3733. doi: 10.1111/jocn.13744.

Zalta, A. K., & Shankman, S. A. (2016). Conducting psychopathology prevention research in the RDoC era. *Clinical Psychology: Science and Practice, 23(1),* 94-104. doi: 10.1111/cpsp.12144

Table 1.

Demographic and Mental Health Characteristics for the Perfectionism Intervention and Stress-Management Control Group

Variable	Intervention	Control	Test of the difference
Demographic characteristics			
1st or 2nd Year Student	36 (80%)	33 (80.5%)	p > .05
No. Female (%)	40 (88.9%)	37 (90.2%)	p > .05
No. non-Hispanic White (%)	18 (40%)	14 (34.1%)	p > .05
Mental Health characteristics			
DAS-Perfectionism	55.31 (18.41)	56.08 (17.69)	p >.05
DASS Total	40.86 (20.73)	46.32(22.46)	p > .05
BTPS Total	122.76(25.76)	123.73(25.98)	p > .05

Figure 1.

Changes in Dysfunctional Attitudes-Perfectionism Scores at Three Time Points

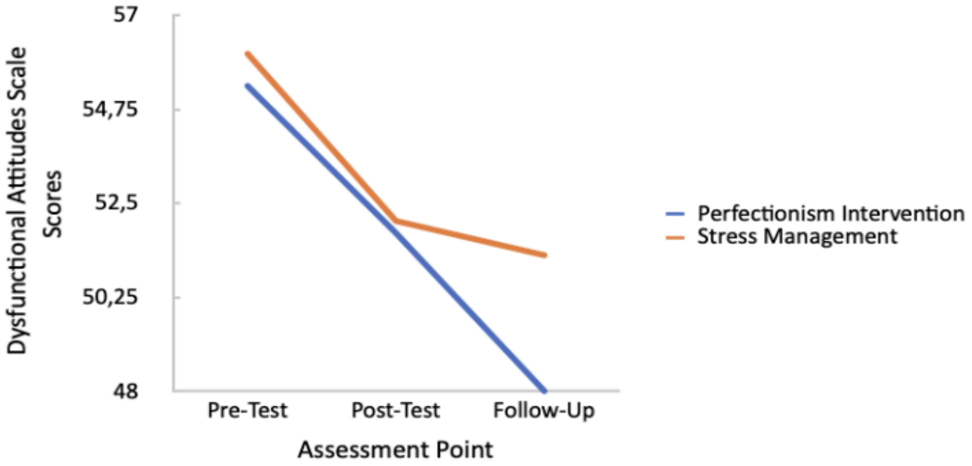
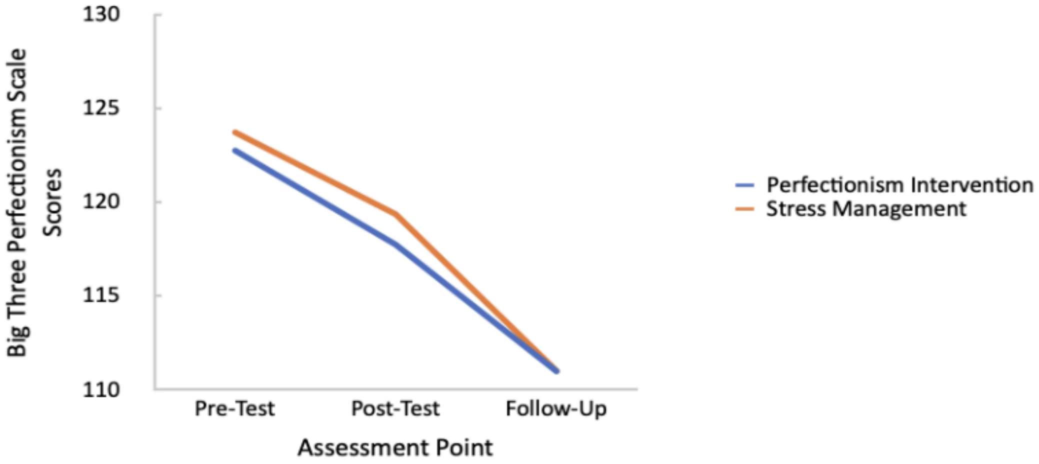


Figure 2

Changes in Big Three Perfectionism Scores at Three Time Points



Graduate Student Journal of Psychology
2022, Vol. 19

Evaluating the Sampling Precision of Social Identity Related Published Research

Cobe Wilson, David Trafimow, and Tony Wang,
New Mexico State University
Cong Wang,
University of Nebraska Omaha

Social identity theory states that a person's sense of who they are is based largely on their group membership(s). We categorize ourselves, identify with groups, and compare our groups with others, in the hopes that our self-esteem is maintained or boosted from this comparison. A prerequisite of scientific research, even regarding social identity, is that researchers need to be confident that the empirical facts really are factual; that is, that the sample statistics reported accurately estimate corresponding population parameters; this is known as sampling precision, or how precisely our sample corresponds to our population. By employing the recently invented a priori procedure, the present research assesses the sampling precision with which published experimental and correlational social identity research statistics, across three time periods, estimate corresponding population parameters. We hypothesized 1: The precision of research in the social identity area should be hopefully below the 0.10 level or at least the 0.20 level for true experimental designs and 2: Precision in the social identity area should be improving, with recent social identity research enjoying a precision advantage over less recent social identity research. A sample of 75 academic papers, across 46 different journals was collected for analysis. For experimental studies, the mean precision level was 0.51 and the median precision level was 0.50 (n = 39). For correlational studies, the mean precision level was 0.24 and the median precision level was 0.20 (n = 55). The main findings are pessimistic, but with the glimmer of light that precision is improving.

Keywords: precision, confidence, social identity theory, a priori procedure, estimation

The overarching idea of social identity theory is that the individual and the group are intertwined. People are born into social societies, they develop ties to social groups, they classify themselves as members of those social groups, and the unique combinations of those social classifications create unique individuals. According to Stets and Burke (2000), identity is reflexive in that it makes itself an object which can be categorized, classified, and named in relation to other social classifications.

Social identity is a person's knowledge that they belong to a social group or category (Hogg & Abrams, 1988). A social group consists of individuals who hold a set of common social identifications. These groups are further categorized as the ingroup, which consists of individuals who are like the self, and the outgroup, which consists of individuals who are unlike the self. This process of accentuating the similarities with the ingroup and differences with the outgroup leads to a second process called social comparison. The process of identifying and categorizing yourself is known as self-categorization. Social identity researchers commonly assume that by classifying social identities through self-categorization (and by extension accentuating similarities with the ingroup, etc.), individuals will enhance their own positive self-esteem by promoting the ingroup and denigrating the outgroup.

No matter what the theoretical criticisms of social identity happen to be, the theoretical basis of identity and social identity research is not the current question. Trafimow and Myüz (2019) demonstrated imprecision with respect to five areas of psychology (social, developmental, clinical, cognitive, and neuro) and Trafimow, Hyman, and Kostyk (2020) demonstrated imprecision with respect to marketing research specifically. Our goal now is to expand that line of inquiry and address a potential criticism that, to our knowledge, has not been made. Specifically, our goal is to address the issue of the sampling precision of social identity research. Sampling precision is a measure of how close a sample's descriptive statistics are to the corresponding population parameters. Usually, and in the present case, this is measured as a fraction of a standard deviation. If the sampling precision of that research is impressive, it would imply that, whatever the other deficiencies of social identity research, at least there would be good reason to trust that the sample statistics are a good estimate of the corresponding population parameters. In contrast, if the sampling precision of social identity research is poor, it would imply that the sample statistics are not good estimates of the corresponding population parameters. A further implication would be that the findings cannot be trusted, which would be a preliminary problem that future research would have to solve before addressing the more conceptual issues identified in the foregoing paragraph.

To summarize the present thinking, we believe that getting the empirical facts straight is vital both for theorizing and for evaluating theories. If a theory is unable to explain human behavior reliably and validly, then it is not a strong psychological theory. By ensuring that sample statistics are good estimators of population parameters, the reliability and validity of theoretical conclusions related to social identity theory can inspire greater confidence. Further, the generalization of these theories would also be improved. Thus, for the purpose of this article, the actual content of social identity theory is unimportant. What is most important, is the sampling practices of that research. To examine this issue, we use the a priori procedure.

The A Priori Procedure

The a priori procedure (APP) assumes the cruciality of obtaining sample statistics that are good estimators of corresponding population parameters (Trafimow et al., 2020). Lest the reader doubts this, imagine a fanciful scenario where Laplace's omniscient demon appears and informs us that there is no relationship between sample statistics and corresponding population parameters. In that case, no empirical reports would be trusted, nor would sample statistics be taken to provide good tests of hypotheses. Of course, there is no demon, but considering the fanciful scenario renders salient the importance of estimation.

Given that estimation is crucial, there are two related questions. First, there is the precision question: How close do we desire that sample statistics be to their corresponding population parameters? Second, there is the confidence question: What probability do we insist on of being that close? The basic idea of the APP is that the researcher makes specifications for precision and confidence, and then an APP equation provides the necessary sample size to meet the specifications. If the researcher collects the necessary sample size, or a larger one, then the researcher can be assured of meeting the specifications (Trafimow et al., 2020).

Consider an example. Suppose that a researcher intends to collect a single sample and is interested in finding out the sample size needed to be 95% confident of obtaining a sample mean within one-tenth of a standard deviation of the population mean. Using an equation by Trafimow (2017; 2019), $n = \left(\frac{z_{(1-c)/2}}{f}\right)^2$ where n is the sample size, f is the desired precision, and $z_{(1-c)/2}$ is the z-score that corresponds to the desired degree of confidence. Instantiating values into the equation indicates the following: $n=(1.96/0.1)^2 =384.16 \approx 385$. Thus, the researcher would have to collect 385 participants to meet specifications. Insisting on 385 participants may seem extreme compared to the much smaller sample sizes in most research, however, this provides a much greater level of precision than in typical psychology or marketing research (Trafimow & Myüz, 2019; Trafimow, Hyman, & Kostyk, 2020). If the researcher is willing to settle for less precision, such as precision at the 0.3 or 0.4 level typical in much social psychology research, the necessary sample size would drop dramatically. This precision can be treated much like confidence level, where we want to shoot for a 95% confidence level but may have to settle for 90%, for precision we can shoot for .1 but may have to settle for .3 or.4, depending on the circumstances of the research.

Although the APP was designed to be used pre-data, it can be used post-data too, which is a necessary condition for the present work. Remaining with the foregoing example, suppose the researcher wanted to collect 385 participants but only succeeded in obtaining 100 of them. Using 95% confidence as a conventional value, what is the precision entailed by 100 participants? To answer, the foregoing equation can be algebraically manipulated to yield f as opposed to

$$n: f = \frac{z_{(1-c)/2}}{\sqrt{n}} = \frac{1.96}{\sqrt{100}} = 0.196.$$

Thus, if the sample size is 100, rather than 385, the actual precision would be 0.196 rather than the desired value of 0.10 (Trafimow et al., 2018).

Researchers are usually interested in more complex cases such as differences between groups, correlation coefficients or functions of them (e.g., regression weights), and so on (Trafimow, 2019). In more complex cases, the mathematics can become extremely complex; but it is not necessary to address that complexity here because Hui et al. (2020) published links to programs for rendering the computations and these are free and user-friendly. However, those interested in the mathematics can consult Trafimow and MacDonald (2017) for multiple groups, Trafimow, Wang, and Wang (2020) for differences in means for independent samples or dependent samples, and Wang et al. (2021) for correlation coefficients.

To place into perspective the importance of precision values, we turn to Trafimow et al (2018). Within scientific inquiry, replication is a vital tool for the validity and reliability of a theory. If the results based on a theory cannot be replicated, then that theory lacks sufficient support. In a single-sample study, if we desire a precision value of .1, and we only have a sample size of 10, then the probability of replicating would only be .06! However, with 111 participants, that probability becomes .5, and so on as the sample size increases. The bottom line, then, is that the APP can be used, in a post-data fashion, to determine the sampling precision of varied published social identity research, where the sample sizes are reported. And this in turn can be used to draw conclusions about the possible replicability, validity, etc. of that published research.

One question that often comes up is why sampling precision matters at all? If we have a good effect size, then what is the point of the APP? Well, think about effect size. If we are looking for a specific experimental effect, and our effect size is large enough, no matter how small a sample we have, we will be able to see the effect. However, does that mean that the whole population of the world will also undergo that same effect? Realistically not. That is, no matter what the obtained effect size for our sample, we have no idea if that effect size is close to the population effect size unless there has been some sort of APP calculation. Therefore, there is no way to know how well the obtained effect supports or disconfirms the hypothesis or theory from which the hypothesis was derived. This is the strength of the APP and sampling precision. Sampling precision values, and the APP, can be used to provide increased confidence in our sample statistics, and thus increased confidence that the effect we are seeing is representative of the overall population, not just that the effect is present in our sample. Overall, the APP sampling precision values can be viewed as a closeness statistic similar to confidence levels in that the more stringent criteria we use, the more confident we can be in our inferences and conclusions.

In this study, we tested competing hypotheses from optimistic versus pessimistic viewpoints. From an optimistic viewpoint, the expectations would be as follows.

H1: The precision of research in the social identity area should be hopefully below the 0.10 level or at least the 0.20 level for true experimental designs (Trafimow, 2018).

H2: Precision in the social identity area should be improving, with recent social identity research enjoying a precision advantage over less recent social identity research.

Of course, from a pessimistic perspective, the foregoing hypotheses should not be supported. If the precision of social identity research is worse than 0.30 (that is larger), that would support pessimism (Trafimow, 2018). Further, it would be undesirable to find that precision is not improving, as an increase in precision is related to an increase in replicability and stronger generalization of theory, and as research methods have supposedly improved as time passes, we should hope to see an improvement in sampling precision as well. Finally, because the APP has never been used systematically to analyze correlational research, the relevant correlational analyses to be presented can be considered exploratory as a first step towards a more integral use of the APP in sampling analysis.

Method

The goal of this research was to investigate the precision (how close a sample statistic would be to its population parameter) of a sample of published social identity research. To investigate the question of whether social identity research findings are precise enough to be trusted, we applied the APP programs presented by Hui et al. (2020) to test the precision of the studies included in a sample of published social identity research. The present a posteriori use of the APP has been systematically employed by Trafimow and Myüz (2019) and Trafimow, Hyman, and Kostyk (2020) for experimental research but not for correlational research.

Procedure

A total sample of 75 academic papers (see Table 1) was collected, prior to data analysis, broken down into 25 papers per time period across three time periods: 2014 to 2021, 1995 to 2001, and 1975 to 1981. It was important to include a sample sufficiently diverse to be representative of the existing literature, while also ensuring that the sample broke into equal portions. The sample included articles from 46 different journals, randomly selected from the set of articles that met our criteria, from multiple academic disciplines (i.e., psychology, advertising, business, sociology, etc.). The primary criteria for selection were:

1. Must include social identity as a primary target of interest for investigation. This may be as

a theoretical concept, an applied concept, etc. This was checked via the mention of social identity in subject terms, titles, or the paper itself.

2. Must directly specify the methodology (correlational, between-subjects, etc.). This is important, as determining the method of analysis is vital to understanding the precision of particular samples using the APP equations.
3. Must have a publication date within one of the three date ranges (1975-1981, 1995-2001, 2014-2021). These date ranges were chosen as an exploratory (not exhaustive) representation of research across time.

The articles were picked using the institutional library search tool. Keywords used to filter results were as follows: social identity, social identity theory, SIT, experimental, correlational, and research. Articles were picked using a random method of 2 articles per results page starting from the top. If the article did not specify social identity, the research method, or was outside of the appropriate date ranges, then it was skipped, and the next article was picked. Once two articles on the page were selected, we moved to the next results page and began selection again. All studies picked in this sample were quantitative leaning (i.e., analysis was done quantitatively).

Once the articles were collected, the relevant information was cataloged via an Excel spreadsheet (data available from Wilson, Trafimow, Wang, & Wang, 2021 via the OSF open-access database). The information cataloged included the citation for the paper, journal, publication year, number of studies included in the paper, sample sizes, and number of conditions. While some of this information is not pertinent for the actual precision analysis, examining differences in the means and medians by publication year and methodology type is of interest for potential broader investigations and conclusions.

To determine the precision of each study, we used the APP programs listed in Hui et al. (2020). Given sample sizes reported in the articles, and assuming a conventional 95% confidence level, the programs provide the precision level, but with a complication. Specifically, because some studies were experimental whereas others were correlational, there are mathematical reasons why the precision of the correlational studies cannot be compared directly to the precision of the experimental studies (Wang et al., 2021)[1]. Consequently, results pertaining to experimental studies and results pertaining to correlational studies will be presented separately.

Results

There were too few mixed designs for analysis and so we focused on between-participants experimental studies and correlational studies. For experimental studies, the mean precision level was 0.51 (SD = .25) and the median precision level was 0.50 (n = 39). These findings support a pessimistic view of experimental social identity research. For correlational studies, the mean precision level was 0.24 (SD = .16) and the median precision level was 0.20 (n = 55). However, we reiterate that the precision of correlational research cannot be compared to the precision of experimental research. As correlations range from -1 to +1, and the precision value pertains to the fraction of a variance (not a fraction of a standard deviation), the degree to which this level of precision in correlational research is pessimistic or optimistic is a judgment call. Certain theories and paradigms may require more stringent cutoffs for what is considered optimistic or pessimistic, and since correlational APP equations use variance instead of standard deviation in analysis, the cutoffs are a judgment call of the researcher.

To evaluate the effect of the year of publication on precision, we performed multiple correlational analyses. First, we obtained mean and median precision levels across the three time periods for the experimental research. The mean precision levels were 0.59 (n = 17, SD = .20), 0.64 (n = 10, SD = .24), and 0.29 (n = 12, SD = .16) for the most distant to most recent time periods, respectively. The corresponding median precision levels were 0.59 (n = 17), 0.71 (n = 10), and 0.23 (n = 12). With respect to correlational research, the mean precision levels were 0.23 (n = 11, SD = .09), 0.27 (n = 24, SD = .19), and 0.21 (n = 20, SD = .15) for the most distant to most recent time periods, respectively. The corresponding median precision levels for correlational studies were 0.25 (n = 11), 0.23 (n = 24), and 0.18 (n = 20). Overall, both correlational and experimental precisions have shown improvement over time, with experimental studies showing much more improvement than correlational studies.

Second, we correlated precision with the year of

---

[1] These include that correlations are bounded (-1, +1) and that the precision level refers to a squared standard deviation (variance) as opposed to a standard deviation.

publication. Because smaller values indicate better precision, an optimistic perspective suggests a negative correlation whereas a pessimistic perspective suggests no correlation or a positive correlation. The correlation was -0.445 (n = 39, p = .004) for experimental research and -0.056 (n = 55, p = .687) for correlational research. Figure 1 contains a scatterplot pertaining to experimental research and Figure 2 contains a scatterplot pertaining to correlational research, both with the best fit trendline to examine the progression of the data.

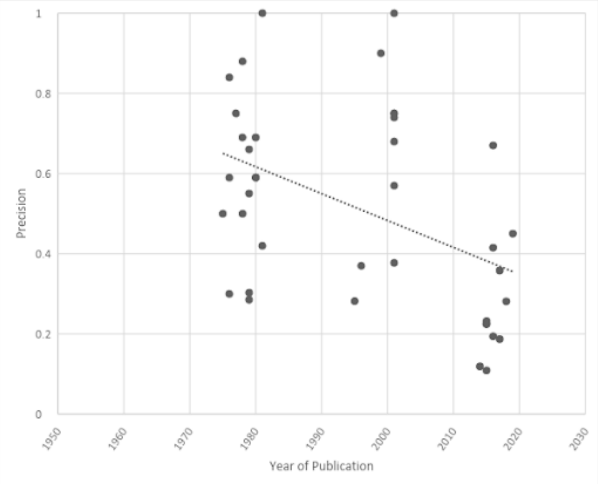


Figure 1. Scatterplot representing precision of experimental social identity research along the vertical axis as a function of year of publication along the horizontal axis.

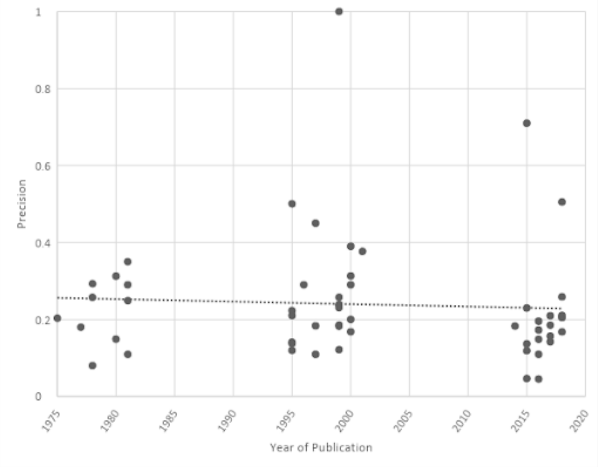


Figure 1. Scatterplot representing precision of correlational social identity research along the vertical axis as a function of year of publication along the horizontal axis.

From the data presented in Figure 1, we can see a clear negative trend across time. As the publication year becomes more recent, the precision level moves further towards zero, indicating a promising improvement in experimental precision. In contrast, Figure 2 does not show much of an effect of time on correlational precision. These time-period correlations should be taken as exploratory however due to the small sample sizes within each time period.

Discussion

Social identity research composes an important domain within social psychology that is relevant to other fields too such as marketing, sociology, and others. Hence, it is unsurprising that there is a large literature and many arguments both praising and criticizing social identity theory. The present work is agnostic about the larger social identity issues briefly mentioned earlier. This is not to say that these issues are unimportant; in contrast, we believe they are very important. However, we also believe that if empirical facts are going to be used either to test theories or to provide the foundation for theory formation, it is crucial to be clear that the empirical facts really are factual. That is, it is a prerequisite that researchers have good reason to believe that the sample statistics researchers report are reasonably good estimates of corresponding population parameters. Our goal was to test whether this prerequisite has been met in a sample of published research.

The picture is very clear with respect to experimental research. Sample sizes are simply insufficient, thereby resulting in precision levels that one would consider untrustworthy. It is true that there are many factors that can create these insufficient numbers such as funding constraints, time, feasibility, and more. However, this does not disqualify the conclusion of insufficient sample sizes, but only increases the need for more detailed and stringent research practices. It is important to note that quality research with smaller sample sizes is still very much alive and is in no way disqualified due to sampling precision. Instead, sampling precision should be taken as an additional tool to assist researchers, in an a priori fashion, to determine a sample size needed for sample statistics that meet their corresponding population parameters, in essence giving researchers a goal for their eventual work to shoot for.

There is some good news, which is that precision is improving, with research in the most recent period of research exhibiting more precision than research in previous periods. But even with the improvement, it would be difficult to argue that the level of precision is sufficient. However, Trafimow and Myüz (2019) demonstrated imprecision with respect to five areas of psychology (social, developmental, clinical, cognitive, and neuro), and Trafimow, Hyman, and Kostyk (2020) demonstrated imprecision with respect to marketing, there is no reason to be more pessimistic with respect to social identity research than other social science research. It is perhaps better to view social identity research as another domain in which researchers should devote increased research efforts in the direction of improved precision.

As the present work constitutes the first systematic application of the APP to correlational research, the correlational analyses are better considered exploratory than definitive. Correlational research offers both advantages and disadvantages relative to experimental research. From an APP perspective, an advantage of correlational research is that there is only one group, thereby rendering precision easier to obtain. Going beyond an APP perspective, correlational research can be argued to be more representative of reality because of a lack of potentially artificial laboratory-induced manipulation. In addition, it is easier to obtain diverse samples in correlational contexts, though the recent proliferation of web-performed experiments is reducing this difference.

But there are disadvantages too. From an APP perspective, a disadvantage is the lack of previous systematic APP applications to correlation coefficients, thereby rendering comparison difficult. This is not a fault of the correlational research itself, but rather due to the historical fact that APP equations applicable to experimental data were developed prior to APP equations applicable to correlational data, and thus the basic assumptions underlying the equations cannot be directly compared on an "A = B" comparison. More generally, it is well-known that correlational research provides a less convincing case for causal mechanisms than experimental research. On the other hand, however, given the present demonstration of the lack of precision of experimental social identity research, it is not clear that even the experimental findings provide strong causal evidence. If a sample mean difference cannot be trusted to reasonably estimate the population mean difference, a strong causal conclusion is contraindicated. Therefore, the disadvantage of correlational research paradigms relative to experimental research paradigms might be considered decreased in the context of small sample experimental research.

Major implications can be discussed surrounding sampling precision and how it might affect social identity research and theoretical findings of social identity theory. It is important, as mentioned previously, to have sample statistics that are good estimates of corresponding population parameters. Without this, basic assumptions of the applicability of theories and sample findings are not met. If sampling precision for research in social identity work is consistently low across the board, with no real attempt to improve, then the replicability of those findings is in jeopardy. Given the major push by scientific advocates for greater replicability in scientific research, especially in social sciences, this means that sampling precision is a vital step towards a more open and truthful scientific discipline. If sampling precision were to remain low, the findings of social research and the applicability of theories developed from those findings would be questionable at best.

As always with scientific research, there are some limitations. One limitation is that we only tested three time periods. This of course was done due to time constraints as well as the exploratory (rather than final) nature of the research. A second limitation is that the sample size of the studies included was limited (again due to time constraints). Thus, conclusions may be clear, but a more comprehensive analysis with a much larger sample size would provide a more detailed and stringent review of sampling precision in a posteriori fashion. Unfortunately, no APP techniques have yet been developed to estimate the number of studies that should be included in an analysis such as that conducted here, and so traditional APP techniques are invalid to determine the sample size for a meta-style analysis of this kind. A third limitation is that, even within social identity research, there are research categories not addressed here. For example, there is social identity research with a basic or applied focus, a focus on integrating other literature, and many others.

Of course, one potential avenue for future research is to address the foregoing limitations. A second potential avenue is to expand to domains that are not precisely about social identity but are related. These could include work in attachment, aggression, or stereotyping. A third avenue is to pursue non normal distributions. Because the researchers in the experimental papers all performed statistics based on the assumption of normality, we used that assumption too in the present analyses for the sake of consistency. However, this assumption is likely wrong, as most distributions are skewed (Blanca et al., 2013; Ho & Yu, 2015; Micceri, 1989). The usual counter to skewness arguments is that the Central Limit Theorem renders deviations from normality unimportant, but that depends on the goal of the research. For example, for the family of skew normal distributions, Trafimow, Wang, and Wang (2019; 2021) have demonstrated that there are important precision gains to be had for analyses analogous to those conducted here, even under low levels of skewness, provided that the researcher focuses on locations as opposed to means. Because locations are a parameter of skew normal distributions, whereas means are not, it makes sense to use locations rather than means anyway as locations are the more generally applicable parameter. And because the location equals the mean when there is normality, nothing is lost by including locations in APP analyses, even in the rare cases where the normality assumption is true. It is also important to note again that this research is not an examination of the accuracy of social identity theory itself and is not limited to "social identity researchers" in the traditional sense (that is those who test the theory of SIT). The goal of this research was to examine a broad and eclectic mix of social identity applications.

In conclusion, the notion that theory should be checked against reality is a staple of science, including the social sciences. But the reality in the social sciences tends to be characterized by summary statistics such as means, standard deviations, correlation coefficients, etcetera. Social scientists do not obtain such sample summary statistics as ends in themselves. Rather, social scientists obtain sample summary statistics because of the faith they have that these provide good estimates of corresponding population parameters. In contrast, the foregoing analyses demonstrate that the sample statistics on which research involving social identity depends cannot be trusted as accurate estimators of corresponding population parameters. We hope and expect that the present work will stimulate social identity research to focus on the preliminary issue of obtaining trustworthy empirical facts, as a prerequisite to resolving the important conceptual issues of consuming interest.

References

*Starred references indicate an article used in analysis. Non-starred references indicate articles used in the manuscript.

*Afshari, L., Young, S., Gibson, P. and Karimi, L. (2019), "Organizational commitment: exploring the role of identity", *Personnel Review*, Vol. 49 No. 3, pp. 774-790. https://doi.org/10.1108/PR-04-2019-0148

*Alsos, G., Clausen, T., Hytti, U., & Solvoll, S. (2016). Entrepreneurs' social identity and the preference of causal and effectual behaviours in start-up processes. *Entrepreneurship and Regional Development*, 28(3-4), 234–258. Doi: 10.1080/08985626.2016.1155742

*Arroyo, C. G., & Zigler, E. (1995). Racial identity, academic achievement, and the psychological well-being of economically disadvantaged adolescents. *Journal of Personality and Social Psychology,* 69(5), 903–914. doi: 10.1037/0022-3514.69.5.903

*Bat-Chava, Y. (2000). Diversity of Deaf Identities. *American Annals of the Deaf*, 145(5), 420–428. doi: 10.1353/aad.2012.0176

*Baumberg, B. (2016). The stigma of claiming benefits: A quantitative study. *Journal of Social Policy*, 45(2), 181-199. doi:10.1017/S0047279415000525

*Ben-Nun Bloom, P., Arikan, G., & Courtemanche, M. (2015). Religious Social Identity, Religious Belief, and Anti-Immigration Sentiment. *American Political Science Review,* 109(2), 203-221. doi:10.1017/S0003055415000143

*Bezdek, W. (1976). Sex Identity, Values and Balance Theory. *Sociometry*, 39(2), 142–153. doi: 10.2307/2786214

Blanca, M. J., Arnau, J., López-Montiel, D., Bono, R., & Bendayan, R. (2013). Skewness and kurtosis in real data samples. M*ethodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 9(2), 78-84. doi: 10.1027/1614-2241/a000057

*Blumstein, P. W. (1975). Identity bargaining and self-conception, *Social Forces*, 53(3), 476–485. doi: 10.2307/2576590

Brown, R. (2000). Social identity theory: Past achievements, current problems, and future challenges, *European Journal of Social Psychology*, 30, 745-778. doi: 10.1002/1099-0992(200011/12)30:63.0.CO;2-O

*Brown, R. J. & Turner, J. C. (1979) The criss-cross categorization effect in intergroup discrimination, *British Journal of Social and Clinical Psychology*, 18, 371-383. doi: 10.1111/j.2044-8260.1979.

tb00907.x

*Bruner, M., Boardley, I., Benson, A., Wilson, K., Root, Z., Turnnidge, J., Sutcliffe, J., & Côté, J. (2017). Disentangling the Relations between Social Identity and Prosocial and Antisocial Behavior in Competitive Youth Sport. *Journal of Youth and Adolescence*. 47. doi: 10.1007/s10964-017-0769-2.

*Burke, P., & Reitzes, D. (1981). The Link Between Identity and Role Performance. *Social Psychology Quarterly*, 44(2), 83–92. doi: 10.2307/3033704

*Burris, C., & Jackson, L. (2000). Social identity and the true believer: Responses to threatened self-stereotypes among the intrinsically religious. *British Journal of Social Psychology*, 39(2), 257–278. doi: 10.1348/014466600164462

*Cancian, F., & Davis, M. (1981). Social Structure and Situations: The effects of Roles, Actions, and Situations on Interpersonal Perception. *Sociological Inquiry*, 51(2), 105–111. doi: 10.1111/j.1475-682X.1981.tb01033.x

*Cardenas, D. & de la Sablonnière, R. (2019). Participating in a new group and the identification processes: The quest for a positive social identity. *British Journal of Social Psychology*. 59. doi: 10.1111/bjso.12340.

*Carmeli, A., Brammer, S., Gomes, E., & Tarba, S. (2017). An organizational ethic of care and employee involvement in sustainability-related behaviors: A social identity perspective: An Ethic of Care and Sustainability Behaviors. *Journal of Organizational Behavior*. doi: 10.1002/job.2185.

*Chadborn, D., & Reysen, S. (2018). Moved by the Masses: A Social Identity Perspective on Inspiration. *Current Psychology*. 37. 625-631. doi: 10.1007/s12144-016-9545-9.

*Chakravarty, S., Fonseca, M. A., Ghosh, S., & Marjit, S. (2016) Religious Fragmentation, Social Identity and Conflict: Evidence from an Artefactual Field Experiment in India. *PLoS ONE* 11(10). doi: 10.1371/journal.pone.0164708

*Cheng, C., Bartram, T., Karimi, L., & Leggat, S. (2016). Transformational leadership and social identity as predictors of team climate, perceived quality of care, burnout and turnover intention among nurses. *Personnel Review*, 45(6), 1200–1216. doi: 10.1108/pr-05-2015-0118

*Christian, J., Gadfield, N. J., Giles, H., & Taylor, D. M. (1976). The multidimensional and dynamic nature of ethnic identity. *International Journal of Psychology*, 11(4), 281. doi: 10.1080/00207597608247363

*Clark, A., Hocevar, D., & Dembo, M. H. (1980). The role of cognitive development in children's explanations and preferences for skin color. *Developmental Psychology*, 16(4), 332–339. doi: 10.1037/0012-1649.16.4.332

*Deseran, F. A., & Chung, C-S. (1979). Appearance, Role-Taking, and Reactions to Deviance: Some Experimental Findings. *Social Psychology Quarterly,* 42(4), 426–430. doi: 10.2307/3033815

*Devine, C. (2014). Ideological Social Identity: Psychological Attachment to Ideological In-Groups as a Political Phenomenon and a Behavioral Influence. *Political Behavior,* 37(3), 509–535. doi: 10.1007/s11109-014-9280-6

*Dietz-Uhler, B. (1996). The escalation of commitment in political decision-making groups: a social identity approach. *European Journal of Social Psychology*, 26(4), 611–629. doi: 10.1002/(SICI)1099-0992(199607)26:4<611::AID-EJSP781>3.0.CO;2-6

*Ely, R. J. (1995). The Power in Demography: Women's Social Constructions of Gender Identity at Work. *Academy of Management Journal*, 38(3), 589–634. doi: 10.2307/256740

*Forgas, J. P. (1981). Responsibility attribution by groups and individuals: the effects of the interaction episode. *European Journal of Social Psychology*, 11(1), 87–99. doi: 10.1002/ejsp.2420110106

*Fryrear, J. L., Nuell, L. R., & White, P. (1977). Enhancement of male juvenile delinquents' self-concepts through photographed social interactions. *Journal of Clinical Psychology*, 33(3), 833–838. doi: 10.1002/1097-4679(197707)33:3<833::AID-JCLP2270330347>3.0.CO;2-N

*Gagnon, A., & Bourhis, R. (1996). Discrimination in the Minimal Group Paradigm: Social Identity or Self-Interest? *Personality and Social Psychology Bulletin*, 22(12), 1289–1301. doi: 10.1177/01461672962212009

*Gangadharan, L., Jain, T., Maitra, P., & Vecci, J. (2016). Social identity and governance: The behavioral response to female leaders. *European*

Economic Review, 90, 302–325. doi: 10.1016/j.euroecorev.2016.01.003

*Genesee, F., Tucker, G. R., & Lambert, W. E.. (1978). The Development of Ethnic Identity and Ethnic Role Taking Skills in Children from Different School Settings. *International Journal of Psychology,* 13(1), 39. doi: 10.1080/00207597808246613

*Giles, H., Llado, N., McKirnan, D. J., & Taylor, D. M. (1979). Social Identity in Puerto Rico. *International Journal of Psychology*, 14(1–4), 185–201. doi: 10.1080/00207597908246730

*Goll, J.C., Charlesworth, G., Scior, K., & Stott, J. (2015) Barriers to Social Participation among Lonely Older Adults: The Influence of Social Fears and Identity. *PLoS ONE* 10(2), 1-17. doi: 10.1371/ journal.pone.0116664

*Grant, P. R., & Brown, R. (1995). From Ethnocentrism to Collective Protest: Responses to Relative Deprivation and Threats to Social Identity. *Social Psychology Quarterly,* 58(3), 195–212. Doi: 10.2307/2787042

*Grant, P., Abrams, D., Robertson, D., & Garay, J. (2014). Predicting Protests by Disadvantaged Skilled Immigrants: A Test of an Integrated Social Identity, Relative Deprivation, Collective Efficacy (SIRDE) Model. *Social Justice Research*, 28(1), 76–101. doi: 10.1007/s11211-014-0229-z

*Haslam, S., & Platow, M. (2001). The Link between Leadership and Followership: How Affirming Social Identity Translates Vision into Action. *Personality & Social Psychology Bulletin*, 27(11), 1469–1479. doi: 10.1177/01461672012711008

*Haslam, S., Oakes, P., Reynolds, K., & Turner, J. (1999). Social Identity Salience and the Emergence of Stereotype Consensus. *Personality & Social Psychology Bulletin,* 25(7), 809–818. doi: 10.1177/0146167299025007004

*Hennessy, J., & West, M. (1999). Intergroup Behavior in Organizations: A Field Test of Social Identity Theory. *Small Group Research,* 30(3), 361–382. doi: 10.1177/104649649903000305

Ho, A. D., & Yu, C. C. (2015). Descriptive statistics for modern test score distributions: Skewness, kurtosis, discreteness, and ceiling effects. *Educational and Psychological Measurement,* 75(3), 365-388. doi: 10.1177/0013164414548576

*Hofman, J. E., & Rouhana, N. (1976). Young Arabs in Israel: Some Aspects of a Conflicted Social Identity. *Journal of Social Psychology*, 99(1), 75. doi: 10.1080/00224545.1976.9924750

Hogg, M. A. & Abrams, D. (1988). *Social Identifications: A Social Psychology of Intergroup Relations and Group Processes.* London: Routledge.

Hogg, M. A., & Turner, J. C. (1985). Interpersonal attraction, social identification, and psychological group formation. *European Journal of Social Psychology*, 15, 51-66. doi: 10.1002/ejsp.2420150105

*Hopkins, N., Reicher, S., Khan, S., Tewari, S., Srinivasan, N., & Stevenson, Clifford. (2015). Explaining effervescence: Investigating the relationship between shared social identity and positive experience in crowds. *Cognition & emotion*. 30. 1-13. doi: 10.1080/02699931.2015.1015969.

*Hughes, M., Kiecolt, K., Keith, V., & Demo, D. (2015). Racial Identity and Well-Being among African Americans. *Social Psychology Quarterly,* 78(1), 25-48. doi: 10.2307/44075387

Hui., L., Trafimow, D., Wang, T., Wang, C., & Hu, L. (2020). User-friendly computer programs so econometricians can run the a priori procedure. *Frontiers in Management and Business*. 1(1), 2-6. doi: 10.25082/FMB.2020.01.002

*Jackson, J. W., & Smith, E. R. (1999). Conceptualizing Social Identity: A New Framework and Evidence for the Impact of Different Dimensions. *Personality and Social Psychology Bulletin*, 25(1), 120–135. doi: 10.1177/0146167299025001010

*Jetten, J., Dane, S., Williams, E., Liu, S., Haslam, C., Gallois, C., & McDonald, V. (2018). Ageing well in a foreign land as a process of successful social identity change. *International Journal of Qualitative Studies on Health and Well-Being*, 13(1), 1508198–.doi:10.1080/17482631.2018.1508198

*Khan, S., Hopkins, N., Reicher, S., Tewari, S., Srinivasan, N., & Stevenson, C. (2015). How Collective Participation Impacts Social Identity: A Longitudinal Study from India. *Political Psychology*. 37. doi: 10.1111/pops.12260.

*Kim, J., & Kim, J. (2018). Effects of social identity and schadenfreude on attitude toward brand sponsoring an instant replay review: The moderating role of rivalry and suspense. *Journal of Consumer Behaviour*. doi: 10.1002/cb.1738.

*Kirk, L., & Burton, M. (1977). Meaning and Context: A Study of Contextual Shifts in Meaning of Maasai Personality Descriptors. *Ameri-*

can Ethnologist, 4(4), 734–761. doi: 10.1525/ae.1977.4.4.02a00090

*Klein, O., & Azzi, A. (2001). Do High-Status Groups Discriminate More? Differentiating between Social Identity and Equity Concerns. *Social Behavior & Personality: An International Journal,* 29(3), 209–221. doi: 10.2224/sbp.2001.29.3.209

Korte, R. F. (2007). A review of social identity theory with implications for training and development, *Journal of European Industrial Training*, 31(4), 166-180. doi: 10.1108/03090590710739250

*Kugihara, N. (2001). Effects of aggressive behaviour and group size on collective escape in an emergency: A test between a social identity model and deindividuation theory. *British Journal of Social Psychology*, 40(4), 575. doi: 10.1348/014466601164993

*Laverie, D., & Arnett, D. (2000). Factors Affecting Fan Attendance: The Influence of Identity Salience and Satisfaction. *Journal of Leisure Research*, 32(2), 225–246. doi: 10.1080/00222216.2000.11949915

Long, K.M., Spears, R., & Manstead, A.S.R. (1994). The influence of personal and collective self-esteem on strategies of social differentiation, *British Journal of Social Psychology* 33, 313-329. doi: 10.1111/j.2044-8309.1994.tb01028.

*McGowan, M., Shiu, E., & Hassan L. M. (2017). The Influence of Social Identity on Value Perceptions and Intention. *Journal of consumer behaviour*.16(3), 242–253. doi: 10.1002/cb.1627

*McKirnan, D. (1980). The identification of deviance: A conceptualization and initial test of a model of social norms. *European Journal of Social Psychology,* 10(1), 75–93. doi: 10.1002/ejsp.2420100106

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological. Bulletin*, 105(1), 156-166. doi: 10.1037/0033-2909.105.1.156

*Millward, L. J. (1995). Contextualizing social identity in considerations of what is means to be a nurse. *European Journal of Social Psychology, 25*(3), 303–324. doi: 10.1002/ejsp.2420250305

*Moore, C. W., & Pride, W. M. (1980). Selected Effects of Qualifying Statements Regarding Technical Performance Claims: An Experimental Investigation. *Journal of Advertising,* 9(2), 20–44. doi: 10.1080/00913367.1980.10673315

*Morton, T., & Duck, J. (2000). Social Identity and Media Dependency in the Gay Community: The Prediction of Safe Sex Attitudes. *Communication Research*, 27(4), 438–460. doi: 10.1177/009365000027004002

*Mummendey, A., Kessler, T., Klink, A., & Mielke, R. (1999). Strategies to Cope With Negative Social Identity: Predictions by Social Identity Theory and Relative Deprivation Theory. *Journal of Personality and Social Psychology*, 76(2), 229–245. doi: 10.1037/0022-3514.76.2.229

*Murnieks, C., McMullen, J., & Cardon, M. (2017). Does Congruence with an Entrepreneur Social Identity Encourage Positive Emotion Under Environmental Dynamism?. *Journal of Small Business Management*. doi: 10.1111/jsbm.12335.

*Myaskovsky, L., & Wittig, M. (1997). Predictors of feminist social identity among college women. *Sex Roles,* 37(11), 861–883. doi: 10.1007/BF02936344

*Nauroth, P., Gollwitzer, M., Bender, J., & Rothmund, T. (2015) Social Identity Threat Motivates Science-Discrediting Online Comments. *PLoS ONE*, 10(2). doi: 10.1371/journal.pone.0117476

*Ng, S. (1978). Minimal Social Categorization, Political Categorization, and Power Change. *Human Relations*, 31(9), 765–779. doi: 10.1177/001872677803100902

*Oakes, P. J., & Turner, J. C. (1980). Social categorization and intergroup behaviour: does minimal intergroup discrimination make social identity more positive?, *European Journal of Social Psychology,* 10(3), 295–301. doi: 10.1002/ejsp.2420100307

*Platow, M. J., Durante, M., Williams, N., Garrett, M., Walshe, J., Cincotta, S., Lianos, G., & Barutchu, A. (1999). The contribution of sport fan social identity to the production of prosocial behavior. *Group Dynamics: Theory, Research, and Practice*, 3(2), 161–169. doi: 10.1037/1089-2699.3.2.161

*Platow, M., & van Knippenberg, D. (2001). A Social Identity Analysis of Leadership Endorsement: The Effects of Leader Ingroup Prototypicality and Distributive Intergroup Fairness. *Personality and Social Psychology Bulletin*, 27(11), 1508–1519. doi: 10.1177/01461672012711011

Pratt, M.G. (2003), "Disentangling collective identities", in Polzer, J.T. (Ed.), *Identity Issues in Groups*, Elsevier Science, Amsterdam, pp. 161-188. doi: 10.1016/s1534-0856(02)05007-7

*Reitzes, D. C. (1980). College Student Identities and Behaviors, *Sociological Focus*, 13(2), 113–124. Doi: 10.1080/00380237.1980.10570365

*Rotondi Jr., T. (1975). Organizational Identification and Group Involvement. *Academy of Management Journal,* 18(4), 892–897. doi: 10.2307/255389

Rubin, M., & Hewstone, M. (1998). Social Identity Theory's self-esteem hypothesis: a review and some suggestions for clarification, *Review of Personality and Social Psychology,* 2, 40-62. doi: 10.1207/s15327957pspr0201_3

*Scott, C. R. (1999). The Impact of Physical and Discursive Anonymity on Group Members' Multiple Identifications During.. *Western Journal of Communication*, 63(4), 456. doi: 10.1080/10570319909374654

*Sheikh, H., Gómez, A., & Atran, S. (2016). Empirical Evidence for the Devoted Actor Model. *Current Anthropology*. 57. S204-S209. doi: 10.1086/686221.

*Shih, M., Pittinsky, T., & Ambady, N. (2016). Stereotype Susceptibility: Identity Salience and Shifts in Quantitative Performance. *Psychological Science,* 10(1), 80–83. doi: 10.1111/1467-9280.00111

*Simon, R. (1997). The Meanings Individuals Attach to Role Identities and Their Implications for Mental Health. *Journal of Health and Social Behavior*, 38(3), 256–274. doi: 10.2307/2955370

*Skevington, S. M. (1981). Intergroup relations and nursing. *European Journal of Social Psychology,* 11(1), 43–59. Doi: 10.1002/ejsp.2420110103

*Smith-Lovin, L. (1979). Behavior Settings and Impressions Formed from Social Scenarios. *Social Psychology Quarterly*, 42(1), 31–43. doi: 10.2307/3033871

Stets, J. E., & Burke, P. J. (2000). Identity Theory and Social Identity Theory. *Social Psychology Quarterly*, 63(3), 224–237. doi: 10.2307/2695870

Tajfel, H. (1978). Social categorization, social identity and social comparison. In H. Tajfel (Ed.), Differentiation between social groups: *Studies in the social psychology of intergroup relations* (pp. 61–76). London, UK: Academic Press.

*Taylor, D. M., & Guimond, S. (1978). The Belief Theory of Prejudice in an Intergroup Context. *Journal of Social Psychology*, 105(1), 11. doi: 10.1080/00224545.1978.9924085

*Terry, D., Hogg, M., & White, K. (1999). The theory of planned behaviour: Self-identity, social identity and group norms. *British Journal of Social Psychology*, 38(3), 225–244. doi: 10.1348/014466699164149

*Thoits, P. (1995). Identity-Relevant Events and Psychological Symptoms: A Cautionary Tale. *Journal of Health and Social Behavior,* 36(1), 72–82. doi: 10.2307/2137288

Trafimow, D. (2017). Using the coefficient of confidence to make the philosophical switch from a posteriori to a priori inferential statistics. *Educational and Psychological Measurement,* 77(5), 831 – 854. doi: 10.1177/0013164416667977

Trafimow, D. (2018). An a priori solution to the replication crisis. *Philosophical Psychology,* 31(8), 1188-1214. doi: 10.1080/09515089.2018.1490707

Trafimow, D. (2019). A frequentist alternative to significance testing, p-values, and confidence intervals. *Econometrics*, 7(2), 1-14. doi: 10.3390/econometrics7020026

Trafimow, D. T., Hyman, M. R., & Kostyk, A. (2020). The (im)precision of scholarly consumer behavior research. *Journal of Business Research*, 114, 93-101. doi: 10.1016/j.jbusres.2020.04.008

Trafimow, D., & MacDonald, J. A. (2017). Performing inferential statistics prior to data collection. *Educational and Psychological Measurement,* 77(2), 204-219. doi:10.1177/0013164416659745

Trafimow, D., & Myüz, H. A. (2019). The sampling precision of research in five major areas of psychology. *Behavior Research Methods*, 51, 2039-2058, doi: 10.3758/s13428018-1173-x.

Trafimow, D., Wang, T., & Wang, C. (2019). From a sampling precision perspective, skewness is a friend and not an enemy! *Educational and Psychological Measurement,* 79(1), 129-150. doi: 10.1177/0013164418764801

Trafimow, D., Wang, C., & Wang, T. (2020). Making the a priori procedure (APP) work for differences between means. *Educational and Psychological Measurement*, 80(1), 186-198. doi:10.1177/0013164419847509

Trafimow, D., Wang, T., & Wang, C. (2021). A viable approach to mitigating irreproducibility. *Stats*, 4, 205–215. doi: 10.3390/stats4010015

*van Knippenberg, A., Wilke, H., & De Vries, N. K. (1981). Social comparison on two dimensions. *European Journal of Social Psychology*, 11(3), 267–283. doi: 10.1002/ejsp.2420110303

*Vaughan, G. M. (1978). Social change and intergroup preferences in New Zealand. *European Journal of Social Psychology,* 8(3), 297–314. doi: 10.1002/ejsp.2420080304

*Wenzel, M. (2001). A social categorization approach to distributive justice: social identity as the link between relevance of inputs and need for justice. *British Journal of Social Psychology*, 40(3), 315–335. doi: 10.1348/014466601164858

Wang C., Wang T., Trafimow D., Li H., Hu L., Rodriguez A. (2021) Extending the A Priori Procedure (APP) to Address Correlation Coefficients. In: Ngoc Thach N., Kreinovich V., Trung N.D. (Eds.), *Data Science for Financial Econometrics: Studies in Computational Intelligence*, vol 898. Springer, Cham. doi: 10.1007/978-3-030-48853-6_10

Wilson, C., Trafimow, D., Wang, T., & Wang, C. (2021, April 7). Evaluating the Precision of Published Research on Social Identity Theory. Retrieved from https://osf.io/na75s/?view_only=2ff5655a16a240aba8c867780e75a6c6

Yuki, M. (2003). Intergroup Comparison Versus Intragroup Relationships: A Cross-Cultural Examination of Social Identity Theory in North American and East Asian Cultural Contexts, *Social Psychology Quarterly,* 66(2), 166-183. doi: 10.2307/1519846

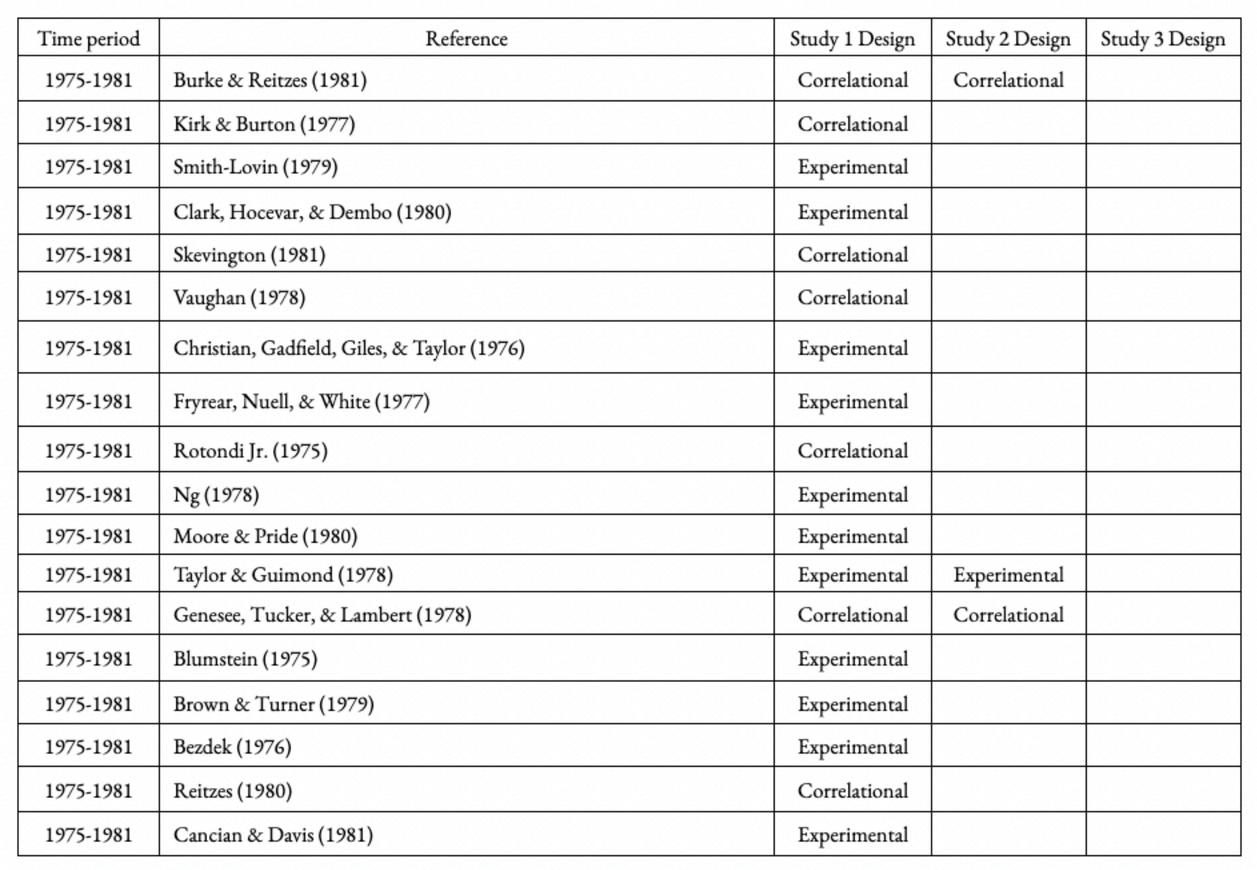
*Zwettler, C., Reiss, N., Rohrmann, S., Warnecke, I., Luka-Krausgrill, U., & van Dick, R. (2018). The relation between social identity and test anxiety in university students. *Health Psychology Open,* 5(2), doi: 10.1177/2055102918785415

Table 1

Citations of Included Studies, Broken Down by Whether they are Experimental or Correlational and by Time Period.
**For data please contact first author*

Time period	Reference	Study 1 Design	Study 2 Design	Study 3 Design
2014-2021	Gangadharan, Jain, Maitra, & Vecci, (2016)	Mixed Methods		
2014-2021	Afshari, Young, Gibson, & Karimi, L. (2019)	Mixed Methods		
2014-2021	Cardenas & de la Sablonnière (2019)	Experimental		
2014-2021	Khan, Hopkins, Reicher, Tewari, Srinivasan, & Stevenson (2015)	Correlational		
2014-2021	Sheikh, Gómez, & Atran (2016)	Correlational	Correlational	
2014-2021	Murnieks, McMullen, & Cardon (2017)	Correlational		
2014-2021	Kim & Kim (2018)	Experimental		
2014-2021	Bruner, Boardley, Benson, Wilson, Root, Turnnidge, Sutcliffe, & Côté (2017)	Correlational		
2014-2021	Hopkins, Reicher, Khan, Tewari, Srinivasan, & Stevenson (2015)	Correlational		
2014-2021	Carmeli, Brammer, Gomes, & Tarba (2017)	Experimental	Experimental	
2014-2021	Goll, Charlesworth, Scior, & Stott (2015)	Correlational		
2014-2021	Chakravarty, Fonseca, Ghosh, & Marjit, (2016)	Experimental		
2014-2021	Chadborn, & Reysen, (2018)	Correlational	Correlational	Correlational
2014-2021	Nauroth, Gollwitzer, Bender, & Rothmund, (2015)	Correlational	Experimental	Experimental
2014-2021	Hughes, Kiecolt, Keith, & Demo, (2015)	Correlational		
2014-2021	Ben-Nun Bloom, Arikan, & Courtemanche, (2015)	Experimental	Experimental	
2014-2021	Baumberg (2016)	Correlational		
2014-2021	McGowan, Shiu, & Hassan (2017)	Correlational	Correlational	
2014-2021	Devine (2014)	Experimental		
2014-2021	Zwettler, Reiss, Rohrmann, Warnecke, Luka-Krausgrill, & van Dick (2018)	Correlational		
2014-2021	Jetten, Dane, Williams, Liu, Haslam, Gallois, & McDonald (2018)	Correlational		
2014-2021	Shih, Pittinsky, & Ambady, (2016)	Experimental	Experimental	
2014-2021	Grant, Abrams, Robertson, & Garay, (2014)	Correlational		
2014-2021	Cheng, Bartram, Karimi, & Leggat, (2016)	Correlational		
2014-2021	Alsos, Clausen, Hytti, & Solvoll, (2016)	Correlational		

Time period	Reference	Study 1 Design	Study 2 Design	Study 3 Design
1995-2001	Jackson & Smith (1999)	Correlational	Correlational	
1995-2001	Ely (1995)	Correlational		
1995-2001	Simon (1997)	Correlational		
1995-2001	Millward (1995)	Correlational		
1995-2001	Arroyo & Zigler (1995)	Correlational	Correlational	Correlational
1995-2001	Bat-Chava (2000)	Correlational	Correlational	
1995-2001	Thoits (1995)	Correlational		
1995-2001	Scott (1999)	Experimental		
1995-2001	Klein & Azzi (2001)	Experimental		
1995-2001	Laverie & Arnett (2000)	Correlational		
1995-2001	Morton & Duck (2000)	Correlational		
1995-2001	Grant & Brown (1995)	Experimental		
1995-2001	Gagnon & Bourhis (1996)	Experimental		
1995-2001	Myaskovsky & Wittig (1997)	Correlational		
1995-2001	Wenzel (2001)	Experimental	Correlational	
1995-2001	Hennessy & West (1999)	Correlational		
1995-2001	Burris & Jackson (2000)	Correlational		
1995-2001	Kugihara (2001)	Experimental	Experimental	
1995-2001	Platow, Durante, Williams, Garrett, Walshe, Cincotta, Lianos, & Barutchu (1999).	Correlational		
1995-2001	Dietz-Uhler (1996)	Correlational	Mixed Methods	
1995-2001	Terry, Hogg, & White (1999)	Correlational		
1995-2001	Haslam & Platow (2001)	Experimental	Experimental	
1995-2001	Haslam, Oakes, Reynolds, & Turner (1999)	Correlational	Correlational	
1995-2001	Platow, & van Knippenberg, (2001)	Experimental		
1995-2001	Mummendey, Kessler, Klink, & Mielke (1999)	Correlational		
1975-1981	Oakes & Turner (1980)	Experimental		
1975-1981	Deseran & Chung (1979)	Experimental		
1975-1981	Forgas (1981)	Correlational		
1975-1981	McKirnan (1980)	Correlational		
1975-1981	van Knippenberg, Wilke, & De Vries (1981)	Experimental		
1975-1981	Giles, Llado, McKirnan, & Taylor (1979)	Experimental		
1975-1981	Hofman & Rouhana (1976)	Experimental		



Time period	Reference	Study 1 Design	Study 2 Design	Study 3 Design
1975-1981	Burke & Reitzes (1981)	Correlational	Correlational	
1975-1981	Kirk & Burton (1977)	Correlational		
1975-1981	Smith-Lovin (1979)	Experimental		
1975-1981	Clark, Hocevar, & Dembo (1980)	Experimental		
1975-1981	Skevington (1981)	Correlational		
1975-1981	Vaughan (1978)	Correlational		
1975-1981	Christian, Gadfield, Giles, & Taylor (1976)	Experimental		
1975-1981	Fryrear, Nuell, & White (1977)	Experimental		
1975-1981	Rotondi Jr. (1975)	Correlational		
1975-1981	Ng (1978)	Experimental		
1975-1981	Moore & Pride (1980)	Experimental		
1975-1981	Taylor & Guimond (1978)	Experimental	Experimental	
1975-1981	Genesee, Tucker, & Lambert (1978)	Correlational	Correlational	
1975-1981	Blumstein (1975)	Experimental		
1975-1981	Brown & Turner (1979)	Experimental		
1975-1981	Bezdek (1976)	Experimental		
1975-1981	Reitzes (1980)	Correlational		
1975-1981	Cancian & Davis (1981)	Experimental		

Department of Counseling and Clinical Psychology
Teachers College, Columbia University
525 West 120th Street
New York, NY 10027
www.tc.edu/publications/gsjp