Assessment Reactivity within the Context of a Web-Based Brief Intervention for Alcohol Use

Tera Leigh Fazzino
University of Vermont, terafazz@gmail.com

Follow this and additional works at: http://scholarworks.uvm.edu/graddis
Part of the Psychology Commons

Recommended Citation
ASSESSMENT REACTIVITY WITHIN THE CONTEXT OF A WEB-BASED BRIEF INTERVENTION FOR ALCOHOL USE

A Dissertation Presented

by

Tera Fazzino

to

The Faculty of the Graduate College

of

The University of Vermont

In Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy
Specializing in Experimental Psychology

October, 2014

Defense Date: September 5, 2014
Dissertation Examination Committee:

Gail Rose, Ph.D., Advisor
John Helzer, M.D., Advisor
Charles Maclean, M.D., Chairperson
John Hughes, M.D.
Keith Burt, Ph.D.
Valerie Harder, Ph.D.
Cynthia J. Forehand, Ph.D., Dean of the Graduate College
ABSTRACT

Introduction: Unhealthy alcohol use is a substantial problem among college students and can lead to a variety of negative consequences. Commercially available web-based brief alcohol intervention (WBI) programs have demonstrated efficacy in a range of student groups and have been widely disseminated to colleges to address this issue. However, the majority of published WBI studies required participants to complete baseline research assessments (RA) about their alcohol use before the WBI and reactivity to the RA may have inflated WBI efficacy estimates in these studies. The present study tested whether there was an additive effect of RA administered online plus a WBI on alcohol consumption, alcohol consequences, and protective behaviors related to alcohol used in the past month compared to the effects of only a WBI. It was hypothesized that participants randomized to the RA+WBI condition would have significantly lower alcohol consumption in the past month, fewer alcohol-related problems, and use more protective behaviors related to alcohol consumption in the past month than participants randomized to the WBI only condition.

Methods: Undergraduate students (n= 856) from universities in the United States and Canada were recruited for this online study. Seventy percent of the sample was female and 82% were Caucasian. The sample had a mean age was 20.0. Sixty four percent (n= 547) of participants who were randomized completed the WBI. Sixty-eight percent completed the one month follow up questionnaire.

Results: Multiple regression analyses using 20 multiply imputed datasets revealed that there were no significant differences in groups at follow up on alcohol use measures, alcohol related problems, or protective behaviors used when controlling for variables with theoretical and statistical relevance to the models. A repeated measures analysis of covariance indicated that there was a significant decrease in peak estimated blood alcohol concentration from baseline to follow up, but no differential effect by randomization group. The results suggested there was a moderate effect of the WBI consistent with studies of WBI efficacy in the literature and that there were no substantial assessment reactivity effects.

Discussion: The current study contributes to the literature by identifying an experimental condition under which assessment reactivity may not be present and does not appear to cloud the detection of WBI efficacy when measured within subjects. The results indicate that WBI researchers may be justified in conducting brief pretreatment research assessments online to collect information about participant alcohol use without biasing within subjects estimates of WBI efficacy. Universities using these programs may likely observe similar effect sizes to those reported in the literature, however effectiveness studies are warranted.
DEDICATION

Dad, this is for you.
TABLE OF CONTENTS

DEDICATION ........................................................................................................................................... ii

LIST OF TABLES ...................................................................................................................................... vii

LIST OF FIGURES ................................................................................................................................... ix

CHAPTER 1: INTRODUCTION ...................................................................................................................... 1

1.1. Overview ........................................................................................................................................... 1

1.2. Alcohol Use among College Students ............................................................................................. 2

1.3. Consequences of Unhealthy Alcohol Consumption ......................................................................... 6

1.4. Gender Differences in Alcohol Consumption among College Students ........................................ 8

1.5. Brief Alcohol Interventions ............................................................................................................. 10

1.6. Web-Based Brief Interventions ......................................................................................................... 12

1.7. Assessment Reactivity ....................................................................................................................... 15

1.7.1 Assessment Reactivity Theory ....................................................................................................... 16

1.7.2 Assessment Reactivity Therapeutic Benefit ................................................................................. 17

1.7.3 Assessment Reactivity History ....................................................................................................... 19

1.7.4 Experimental Studies of Assessment Reactivity .......................................................................... 22

1.7.5 Conditions Under Which Assessment Reactivity May be Observed ......................................... 26
1.7.6 Factors that Influence Assessment Reactivity Effects ........................................28
1.7.7 Aspects of Assessment Reactivity Currently Unknown ........................................30
1.8. Study Purpose and Relevance of the Findings ..................................................32
1.9. Specific Aims ........................................................................................................33

CHAPTER 2: METHODS .................................................................................................36
2.1. Study Design and Justification ............................................................................36
2.2. Randomization Procedure ................................................................................37
2.3. Incentives ..............................................................................................................38
2.4. Study Structure ....................................................................................................39
2.5. Programming the Study Website and Protection of Confidential Information

2.6. Recruitment ...........................................................................................................43
2.7. Communication with Participants and Participant Tracking .............................45
2.8. Measures ..............................................................................................................46
2.9. Web-Based Intervention Used in the Current Study ..........................................53
2.10. Data Analysis ......................................................................................................54
2.10.1 Outcome Variables ..........................................................................................54
2.10.2 Multiple Regression...........................................................................................................56
2.10.3 Confounding Variables.......................................................................................................59
2.10.4 Moderating Variables .........................................................................................................62
2.10.5 Validation of Web-Based Brief Intervention Efficacy ......................................................63
2.10.6 Examination of Missing Data and Multiple Imputation ....................................................63
2.10.7 Model Building ................................................................................................................68
2.10.8 Final Models for Each Outcome Variable ........................................................................68
2.10.9 Examination of Baseline Differences in Randomization Groups ..................................69
2.10.10 Power Calculation ...........................................................................................................70

CHAPTER 3: RESULTS....................................................................................................................73
3.1. Recruitment ..........................................................................................................................73
3.2. Descriptive Statistics ...........................................................................................................76
3.3. Tests for Baseline Differences between Randomized Groups ............................................78
3.4. Comparison of Participants who Completed the Web-Based Brief Intervention Versus Those Who Did Not ........................................................................................................78
3.5. Multiple Regression with Multiply Imputed Datasets..........................................................79
  3.5.1 Results for Specific Aim 1................................................................................................79
  3.5.2 Results for Specific Aim 2................................................................................................84
  3.5.3 Results for Specific Aim 3................................................................................................85
3.6. Moderation Analyses: Results for Specific Aim 4.................................87

3.7. Validation of WBI Efficacy: Change in Peak Estimated Blood Alcohol
Content from Baseline to Follow up ..........................................................93

CHAPTER 4: DISCUSSION........................................................................94

4.1. Main Findings: Specific Aims 1-3 .......................................................94

4.2. Moderating Effects: Specific Aim 4.....................................................99

4.3. Future Studies......................................................................................101

4.4. Limitations.........................................................................................103

4.5. Strengths............................................................................................105

4.6. Implications........................................................................................106

References...............................................................................................108
LIST OF TABLES

Table 1: Participant Demographic Characteristics .................................................................77

Table 2: Comparison of Participants Who Completed the Web-Based Brief Intervention to Those Who Did Not.................................................................79

Table 3: Main effect of randomization assignment on the total number of drinks consumed in the past month at follow up (TOT).........................................................80

Table 4: Main effect of randomization assignment on peak estimated blood alcohol content in the past month at follow up (PBAC) ..........................................................82

Table 5: Main effect of randomization assignment on the number of times drank past NIAAA guidelines in the past month at follow up (DP).............................................83

Table 6: Main effect of randomization assignment on the number of alcohol-related problems in the past month at follow up (RAPI score) ..............................................85

Table 7: Main effect of randomization assignment on protective behaviors used in the past month at follow up (PBSS score).................................................................86

Table 8: Moderating effects of baseline AUDIT on total number of drinks consumed in the past month at follow up (TOT).................................................................88

Table 9: Moderating effects of gender on total number of drinks consumed in the past month at follow up (TOT).................................................................89

Table 10: Moderating effects of importance rating on total number of drinks consumed in the past month at follow up (TOT).................................................................91
Table 11: Moderating effects of confidence rating on total number of drinks consumed in the past month at follow up (TOT)
### LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1: Study Design</td>
<td>39</td>
</tr>
<tr>
<td>Figure 2: Power Calculation</td>
<td>70</td>
</tr>
<tr>
<td>Figure 3: Recruitment of Study Participants and Participant Flow</td>
<td>74</td>
</tr>
</tbody>
</table>
CHAPTER 1: INTRODUCTION

1.1. Overview

Unhealthy alcohol use is a substantial problem among college students and can lead to a variety of negative consequences. Commercially available web-based brief alcohol intervention (WBI) programs have been developed to address this problem and have demonstrated efficacy in a wide range of student groups. However, a methodological issue that is prevalent in this area of research may have confounded efficacy findings. Research has demonstrated that asking participants questions about their alcohol use can lead to significant decreases in subsequent alcohol consumption and alcohol-related consequences (Clifford & Davis, 2012; Clifford & Maisto, 2000; McCambridge & Kypri, 2011), a phenomenon that has been termed assessment reactivity. In the majority of WBI studies conducted to date, participants were required to complete baseline research assessments (RA) about their alcohol use and related problems in addition to the WBI. It is possible that findings of WBI efficacy may be inflated by reactivity to RA in these studies.

It is currently unknown whether the combination of web-based RA and WBI leads to lower alcohol consumption and related problems at follow up compared to the effects of a WBI alone. The present study tested whether there was an additive effect of RA administered online plus a WBI on alcohol consumption, alcohol consequences, and protective behaviors related to alcohol used in the past month compared to the effects of only a WBI. Results from the study were expected to either reveal a bias in WBI efficacy estimates or to validate WBI efficacy studies that have been conducted to date. The present study is important within the context of the college student drinking...
literature and the assessment reactivity literature. This document provides a detailed description of both areas of the literature, an explanation of the current study design and findings, and a discussion of the findings within the context of the literature.

1.2. Alcohol Use among College Students

Alcohol use is prevalent among college students. The National Institute on Alcohol Abuse and Alcoholism (NIAAA) defines unhealthy alcohol consumption as drinking 4[5] or more drinks for women [men] on one occasion (NIAAA, 2005). Multiple nationally representative studies have demonstrated that unhealthy alcohol consumption among college students is high and has remained stable for the past two decades. The Harvard School of Public Health College Alcohol Survey (CAS) was a large scale endeavor that evaluated the prevalence and rates of alcohol use among a nationally representative sample of 120 four-year colleges in the United States (U.S.) at four time points between 1993 and 2001 (Wechsler & Nelson, 2008). The first survey in 1993 found that almost half (44%) of college students reported drinking in excess of NIAAA guidelines in the past two weeks (Wechsler, Davenport, Dowdall, Moeykens, & Castillo, 1994). The rate of unhealthy alcohol consumption remained consistent over all four CAS administrations and thus revealed a persistent problem (Wechsler, Dowdall, Maenner, Gledhill-Hoyt, & Lee, 1998; Wechsler, Lee, Kuo, & Lee, 2000; Wechsler et al., 2002).

Three additional nationally representative surveys of college student alcohol use have also been conducted in the past two decades and reported similar estimates. In the National College Health Risk Behavior Survey, researchers found that 35% of students
reported drinking 5 or more drinks in a single day in the past 30 days (Douglas et al., 1997). The estimate of unhealthy alcohol consumption among college students from the National Survey on Drug Use and Health was 40% in 2012, which had decreased from 44% in 2002 (Substance Abuse and Mental Health Services Administration, 2013). A virtually identical estimate of change in unhealthy alcohol use from 44% to 40% was observed in the Monitoring the Future Study, in which researchers enrolled high school students and surveyed them over a 13 year period. Although estimates of unhealthy alcohol consumption have been reported to be around 40%, Presley, Meilman, and Leichliter (2002) found that estimates of unhealthy alcohol use vary dramatically between U.S. universities and reported unhealthy alcohol use rates between 9-71%.

Many factors influence college student alcohol consumption. A description of each of the factors that have the most direct relevance to the current study is presented below.

Membership in a Greek fraternity or sorority has been found in many studies to be strongly associated with alcohol consumption. Capone, Wood, Borsari, and Laird (2007) found that Greek membership predicted an increase in alcohol consumption from freshman to sophomore year. Bartholow, Sher, and Krull (2003) also found that greater exposure to the Greek system led to increased heavy drinking in the college years. Researchers have also found that students during the 4 years of college who changed from being Greek members to non-members had significant decreases in risky alcohol consumption and students who changed from nonmembers to members had a significant increase in risky drinking (Park, Sher, & Krull, 2008). In a review of a decade of Greek research on alcohol consumption, Borsari, Hustad, and Capone (2009)
confirmed that students with Greek membership had substantially higher alcohol consumption than their peers who did not belong to Greek organizations.

There is some evidence in the literature that residence on or off campus influences level of alcohol consumption. Harford, Wechsler, and Seibring (2002) found that college residence on or off campus was related to differential exposure to alcohol settings and that heavy drinking was associated with off campus locations at house parties and bars. In this regard, alcohol consumption has been found to be prevalent at off campus parties, but also in university residence halls (Paschall & Saltz; 2007). Results of a thorough investigation into alcohol consumption in residence halls revealed that students living in same-sex dorms had significantly lower alcohol consumption compared to students living off campus, however students living in coed dorms on campus had significantly more alcohol-related problems than students living in same-sex dormitories (Harford, Wechsler, & Muthén, 2002).

Intercollegiate athletes have been identified as a student group that is at increased risk for risky alcohol use (Yusko, Buckman, White, & Pandina, 2008). Researchers have found that college athletes consumed more alcohol, engaged in more frequent episodes of drinking past NIAAA guidelines, and experienced more negative alcohol-related consequences as compared with nonathletes (Leichliter, Meilman, Presley, & Cashin, 1998; Nelson & Wechsler, 2001; Wechsler, Davenport, Dowdall, Grossman, & Zanakos, 1997). These findings have been further supported by a meta-analysis of the college athlete drinking literature (Martens, Dams-O’Connor, & Beck, 2006).
Alcohol use among college students has been shown to vary depending on the time point within the semester and during university breaks periods. Student alcohol and tobacco use has been demonstrated to be higher in the beginning (Del Boca, Darkes, Greenbaum, & Goldman, 2004; Dierker et al., 2008) and end of semesters (Dierker et al., 2008) than in the middle. Clapp, Johnson, Shillington, Lange, and Voas (2008) found evidence of a seasonal drinking pattern among college students and reported that breath alcohol concentration samples taken from students in the winter and spring were higher than samples taken in the fall, which mirror alcohol consumption patterns in the general population. Further, evidence indicates that alcohol use peaks at holidays and during spring break (Del Boca et al., 2004) and is lower near academic deadlines and exam periods (Del Boca et al., 2004; Greenbaum, Del Boca, Darkes, Wang, & Goldman, 2005). Finally, research suggests that students drink heavily the summer before they transition to college (Aaron White & Swartzwelder, 2009). Researchers have yet to examine the pattern of student alcohol use during summer breaks once students are attending college.

Initial evidence suggests that there are some similarities and differences between students’ drinking habits in the U.S. and Canada. Both the U.S. and Canada define a standard drink with the same quantity of alcohol (although with different metric systems) however, U.S. students are not legally allowed to drink until age 21, whereas Canadian students are of legal drinking age at 18 years old. Evidence from 119 nationally representative U.S. and Canadian colleges suggested that both U.S. and Canadian students had relatively high instances of heavy drinking, however Canadian students drank more frequently than U.S. students, and U.S. students had more episodes
of heavy drinking than Canadian students (Kuo et al., 2002). More research on the comparison between U.S. and Canadian student alcohol use is sorely needed.

There is evidence that college student age differences are related to alcohol consumption. The majority of research on age and alcohol use in college students has been on those of legal drinking age versus underage drinkers in the U.S. (e.g., Wechsler, Kuo, Lee, & Dowdall, 2000). However, one study evaluated age as a continuous measure and found in a large sample of university students that older students drank more frequently than their younger counterparts (Svenson, Jarvis, & Campbell, 1994).

Wechsler and Nelson (2008) described the general drinking style of many college students as “one of excess and intoxication” (p. 483). In the 4th administration of the CAS, the researchers reported that among college student drinkers, 48% responded that getting drunk is an important reason for drinking (Wechsler et al., 2002). Researchers have found that this approach to drinking is often accompanied by a general lack of acknowledgement or concern from the students themselves that they are heavy drinkers or experiencing alcohol-related problems (Dowdall, 2008; Wechsler & Nelson, 2008). In addition, the CAS survey revealed that the percentage of students who sought help for their drinking significantly decreased from 2.4% in 1993 to 1.5% in 2001 (Wechsler et al., 2002).

1.3 Consequences of Unhealthy Alcohol Use

Longitudinal studies of alcohol consumption patterns among college students have found that unhealthy alcohol use significantly declines post-college (Dawson,
Grant, Stinson, & Chou, 2004; O’Neill, Parra, & Sher, 2001) and that a small percentage of those with unhealthy alcohol consumption during college develop alcohol use disorders after college (Donovan, Jessor, & Jessor, 1983). Furthermore, in a national longitudinal survey of over 40,000 people in the U.S., attending college significantly decreased the odds ratio of developing past year and lifetime DSM-IV alcohol dependence compared to those who achieved a high school education or less (Grant, 1997).

However, unhealthy alcohol consumption during college has been shown to be associated with a host of negative consequences including legal involvement on and off campus (Wechsler et al., 2002), health problems, including injuries sustained while intoxicated (Hingson et al., 2002), academic problems, such as missing class and falling behind with school work (Wechsler et al., 2002), social problems, including risky sexual behavior while intoxicated (Cashell-Smith, Connor, & Kypri, 2007), sexual assault (Hingson, Heeren, Winter, & Wechsler, 2005), and death from motor vehicle accidents (Hingson et al., 2005). Further, Knight et al. (2002) found that over 40% of the CAS sample reported one or more symptom of DSM-IV abuse or dependence. A number of studies have shown that current college students, when compared to individuals of the same age who were not attending college, exhibited higher levels of unhealthy alcohol consumption (Dawson et al., 2004; Johnston, O’Malley, Bachman, Abuse, & Research, 1999; O’Malley & Johnston, 2002; Slutske et al., 2004; Timberlake et al., 2007) and reported greater levels of clinically significant alcohol use disorders (Slutske, 2005). Thus, although individuals who attend college have a decreased long-term risk of experiencing alcohol problems post-college than peers who
do not attend college, current college students are at increased risk for acute alcohol related consequences.

1.4 Gender Differences in Alcohol Consumption among College Students

Substantial gender differences have been found in relation to alcohol consumption patterns and alcohol-related consequences during college. The CAS revealed that college males had higher rates of unhealthy alcohol consumption in the last two weeks compared to females in college (49% versus 40%; Wechsler et al., 1998). In a nationally representative, longitudinal sample of U.S. college students, Johnston et al. (1999) reported that males had higher rates of daily drinking during college (5.8%) compared to females in the sample (2.7%).

However, rates of female heavy alcohol use have significantly increased and in recent decades the magnitude of the difference between male and female alcohol consumption has narrowed (Kelly-Weeder, 2008; O’Malley & Johnston, 2002). Wechsler et al. (2002) reported that although unhealthy alcohol consumption rates remained similar between 1993 to 2001, there was a significant (almost 4%) increase in females who reported frequently drinking unhealthy amounts of alcohol and in females who were consuming very high levels of alcohol.

Patterns of drinking over the four years of college differ for males and females. Females have been found to drink more in their first year in college, whereas males have been found to drink more in their third and fourth years of college (McCabe, 2002). Gross (1993) examined collegiate drinking as it relates to legal drinking age and
found that men of legal drinking age consumed more alcohol than their underage peers, whereas underage women drank more alcohol than their legal drinking age peers.

Gender differences are also present in the experience of alcohol-related consequences. Wechsler, Dowdall, Davenport, and Rimm (1995) utilized data from the first wave of the CAS and evaluated the likelihood of males and female college students experiencing twelve different alcohol-related consequences. The researchers found that whereas males were likely to experience alcohol consequences after consuming five or more drinks in one episode, females were likely to experience the same consequences when consuming four or more drinks. Because of these differences, the researchers recommended using differential thresholds to define unhealthy alcohol consumption for male and female college students.

There are also biological differences in alcohol effects on males and females. Alcohol metabolism differs in males and females; females experience higher blood alcohol levels than males after consuming equivalent quantities of alcohol because women have less gastric oxidation of alcohol than males (Frezza et al., 1990). This metabolic difference may leave women at increased risk for acute and chronic effects of alcohol (Frezza et al., 1990).

Given the high and stable rates of unhealthy alcohol consumption among males and female college students and the host of negative consequences that can accompany alcohol consumption, an important task for the alcohol research field has been to develop interventions to reduce college student alcohol consumption. This endeavor has been challenging because of the common belief among college students that being
drunk is an important part of consuming alcohol and the general reluctance to seek help to modify drinking behavior (Wechsler & Nelson, 2008; Wechsler et al., 2002).

1.5 Brief Alcohol Interventions

A brief treatment option that can be provided to individuals who consume unhealthy amounts of alcohol but who either do not require or are unwilling to engage in more intensive treatment, is a brief alcohol intervention (BI). BI consists of assessing alcohol consumption, providing personalized feedback related to alcohol consumption and related risks, and establishing a goal of abstinence or reduced alcohol use. BI has typically been conducted in-person by primary care physicians (National Institute on Alcohol Abuse and Alcoholism, 2005; Ockene, Wheeler, Adams, Hurley, & Hebert, 1997), however the use of BI has been expanded to hospital emergency departments (D’Onofrio & Degutis, 2002; Gentilello et al., 1999) and college student health centers (Schaus et al., 2009). The advantages of conducting BIs are substantial. For example, a large, randomized controlled trial of individuals who screened positive for unhealthy alcohol use in primary care found significant (15–30%) sustained reductions in binge drinking episodes, mean drinks per week, and percent of heavy drinkers over a 48-month period (Fleming et al., 2002). These changes were associated with significantly different health care utilization, including 20% fewer emergency department visits and 37% fewer days of hospitalization. Further, the control group in the study incurred 46% more arrests for legal events, including arrests for controlled substance and liquor violations. In addition, the researchers estimated medical care savings of $712 per patient resulted from differences in emergency department visits and hospitalizations.
between groups. The largest benefit was related to the difference in motor vehicle events, where estimated savings amounted to $7,171 per patient in the experimental group. Thus, these findings were both clinically and financially important and indicated that BIs have significant benefits for individuals as well as to the health care system.

Despite impressive benefits from brief interventions, researchers have found that rates of physician administered BI is very low. For example, in a survey of primary care patients who recently attended a primary care appointment, 11% of patients reported that they were asked about or advised on their alcohol use (Aalto & Seppä, 2004). Physicians estimated they only conduct alcohol screenings in approximately one third of patients during annual exams (Spandorfer, Israel, & Turner, 1999). In a review of chart data, Funk et al. (2005) found that alcohol screening and BI was only conducted in 1% of patients that were seen for appointments.

There are multiple reasons that likely contribute to low BI administration rates in primary care. First, alcohol screening and BI take time to conduct and time constraints during appointments and high patient care loads may present a significant barrier to BI implementation (Aalto, Pekuri, & Seppä, 2003). In this regard, Beich, Gannik, and Malterud (2002) found that the physicians trained to conduct BI in the study reported that BI administration was a burden and that they could not recommend provider initiated alcohol screening and BI, even though they agreed that alcohol counseling was important.

There is also evidence in the literature that physician discomfort in discussing alcohol use is a barrier to BI administration. McCormick et al. (2006) conducted a qualitative study of audio taped alcohol discussions and evaluated how primary care
providers discuss alcohol use with their patients. They found that although providers generally believed that addressing alcohol misuse was important clinically, physicians frequently avoided discussions of alcohol use, generally provided vague and ambiguous advice on decreasing alcohol use, and exhibited discomfort in discussing alcohol use, with observed behaviors that included hesitation and stuttering, inappropriate laughter. Further, while a majority of patients indicate they were open to a discussion of their alcohol use with their physician (Isaacson, Fleming, Kraus, Kahn, & Mundt, 2000; Marcell, Halpern-Felsher, Coriell, & Millstein, 2002; Miller, Sheppard, Colenda, & Magen, 2001), Spandorfer et al. (1999) found that 72% of the physicians sampled indicated they preferred not to counsel patients about alcohol use.

1.6 Web-Based Brief Alcohol Interventions

Many of the barriers to BI administration can be overcome by automating the process. Web-based brief intervention (WBI) programs that can be completed online may be particularly well suited to college students. They provide students the intervention in an efficient, convenient, and technologically adept format online, do not rely on a physician to administer the intervention and do not require in-person treatment attendance for students. In addition, they are cost effective and scalable to universities nationally and internationally.

A substantial number of commercially available WBI programs have been developed and disseminated to colleges in an effort to decrease levels of unhealthy alcohol consumption. The WBIs that have gained the most popularity are comprised of an assessment of student drinking followed by the presentation of personalized
feedback to students with information about their quantity/frequency of alcohol use, drinking-related consequences, risk factors for alcohol problems, and personalized normative feedback (PFN) that compares a student’s drinking level to campus or national drinking norms (Lewis & Neighbors, 2006). PNF is based on the social norms approach which indicates that college students overestimate how much their peers drink and that correcting this overestimation can reduce student drinking and alcohol-related consequences (Perkins & Berkowitz, 1986; Wesley Perkins, 2002; Henry Wechsler, Seibring, Liu, & Ahl, 2004). The first WBI programs that were developed for college students were created by Kypri et al. (2004) in Australia and by Neighbors, Larimer, and Lewis (2004) in the U.S. Kypri et al (2004) demonstrated that students randomized to complete the WBI had significant reductions in total alcohol consumption, heavy drinking episodes, and alcohol-related negative consequences at 6 week follow up. Similarly, Neighbors et al. (2004) also found that the WBI decreased the mean number of drinks participants consumed in the past three months, decreased the maximum number of drinks consumed in a single episode in the past month, and decreased the number of alcohol-related problems. The authors reported that these decreases were a function of a change in perceived peer drinking norms.

Efficacy of Web-Based Brief Alcohol Intervention

WBI programs have demonstrated efficacy in a wide range of student populations, including college freshmen (Hustad, Barnett, Borsari, & Jackson, 2010; Doumas & Andersen, 2009; Kazemi, Sun, Nies, Dmochowski, & Walford, 2011; Saitz et al., 2007), heavy drinkers (Ekman et al., 2011; Walters, Vader, & Harris, 2007;
Carey, Carey, Maisto, & Henson, 2006; Chiauzzi, Green, Lord, Thum, & Goldstein, 2010; Kypri et al., 2004; Linke, Brown, & Wallace, 2004; Spijkerman et al., 2010; Voogt, Poelen, Kleinjan, Lemmers, & Engels, 2011) college athletes (Doumas, Haustveit, & Coll, 2010), students mandated to treatment (Doumas, Workman, Smith, & Navarro, 2011; Mastroleo, Murphy, Colby, Monti, & Barnett, 2011), and college students with alcohol-related risky sexual behavior (Lewis et al., 2014). Cohen’s d effect sizes in these studies have generally been in the small to moderate range (.23 to .59), however some studies have reported large effects (.68-.75; Hester, Delaney, & Campbell, 2012; Hustad et al., 2010). Although effects on alcohol consumption have been detected up to one year following WBI administration, WBI effects have been shown to be most prominent in the first month post-WBI (Chiauzzi et al., 2010; Hester et al., 2012).

Several groups of researchers have conducted systematic reviews of the WBI literature and confirmed that mean effect sizes for WBI studies have been small to medium. White et al. (2010) reported that of 17 WBI studies evaluated, the differential mean effect size for treatment groups versus assessment only controls was a Cohen’s d of 0.42 and the mean pre-post treatment effect size for alcohol consumption was 0.56. Khadjesari, Murray, Hewitt, Hartley, and Godfrey (2011) found that the mean difference in drinks per week between treatment and control groups at follow up was equivalent to 9 ounces of alcohol. In a large meta-analysis using a Bayesian multilevel modeling approach to evaluate WBI efficacy, Huh et al (2014) found that effect on drinks per week was small (OR = 0.75, RR = 0.94). Finally, Riper et al (2009) reported
an average Cohen’s d effect size of 0.22 for WBI studies conducted before 2009 (Riper et al., 2009).

In addition to establishing efficacy, researchers have also compared the efficacy of WBIs to BIs administered in-person. In a large systematic review of WBIs and in-person delivered BIs, Khadjesari et al. (2011) found there was no significant differences in efficacy.

The commercially available WBI programs that have been frequently researched have been the Neighbors et al. (2004) Brief Alcohol Screening and Intervention for College Students program (BASICS), the College Drinker’s Check Up (CDCU; Hester et al., 2012), Alcohol EDU (Outside the Classroom, 2010) and the Electronic Check Up to Go (ECHUG; San Diego State University Research Foundation, 2009). The programs range from 20 minutes to 3 hours in duration and are designed to be completed in either a single session (ECHUG, BASICS) or multiple brief sessions (Alcohol Edu, CDCU). ECHUG may be the most widely distributed WBI program. Over 600 universities in the U.S., Canada and the United Kingdom use ECHUG, according to the San Diego State University Foundation.

1.7 Assessment Reactivity

Despite evidence supporting WBI efficacy, a methodological issue with the vast majority of WBI studies may have confounded these findings. Most of the WBI efficacy studies in the literature (e.g., Bingham et al., 2010; Butler & Correia, 2009; Chiauzzi et al., 2010; Doumas & Andersen, 2009; Doumas et al., 2010; Doumas, Workman, et al., 2011; Ekman et al., 2011; Hustad et al., 2010; Kulesza, Apperson, Larimer, & Copeland, 2010; Neighbors, Lee, Lewis, Fossos, & Walter, 2009; Voogt et
al., 2011; Walters et al., 2007) required participants to complete baseline RA about their alcohol use and related problems prior to completing the WBI. Research has demonstrated that simply asking participants questions about their alcohol use can lead to subsequent decreases in alcohol consumption and alcohol-related problems (Clifford & Davis, 2012; McCambridge & Kypri, 2011; Schrimsher & Filtz, 2011). Of the published WBI studies, only Kypri et al. (2004) did not conduct baseline RA and reported they did so to avoid any assessment reactivity effects. With WBI studies that included baseline RA in their research design, it is unclear whether the effect sizes in these studies reflect the effect of the WBI only, or whether the effect sizes were padded by change in alcohol use due to the RA.

1.7.1 Assessment Reactivity Theory

Assessment reactivity occurs when a behavior is assessed or monitored and this assessment independently affects the participant’s behavior, regardless of other study interventions or manipulations (Clifford & Maisto, 2000). Assessment reactivity has been hypothesized to occur because assessments draw attention to the targeted behaviors, and this allows participants the opportunity to reflect on their behavior, its consequences, and reasons they may want to change (Clifford & Maisto, 2000). Moos (2008) described how assessment may increase an individual’s awareness of and feelings of responsibility for their behavior which may lead to the initiation of self-monitoring and changing the behavior. Researchers have suggested that research assessments contain similar elements to BIs themselves: identifying information regarding risk of behavior, providing a menu of options for change, highlighting
personal responsibility for the behavior, and increasing self-efficacy (Clifford & Davis, 2012; Schrimsher & Filtz, 2011). For example, the Protective Behaviors Strategy Scale (PBSS) evaluates risk reduction and self-control behaviors related to drinking (Martens et al., 2005) and is frequently used in WBI efficacy studies. The PBSS provides students a list of risk reduction behaviors related to drinking and asks to what extent they utilize each one. Thus, the PBSS provides a menu for behaviors to decrease drinking which may educate students about effective strategies and/or increase self-efficacy by providing concrete examples of how to change behavior. The Timeline Follow Back (TLFB; Sobell, Sobell, Leo, & Cancilla, 1988), used almost universally in alcohol research, identifies detailed drinking patterns and may highlight personal responsibility for both alcohol consumption and for changing alcohol consumption. Even the Sobells themselves pointed out that the TLFB could decrease control group drinking (Sobell & Sobell, 1981). The Rutgers Alcohol Problems Index (RAPI) (White & Labouvie, 1989) asks participants to report how frequently they experienced alcohol-related problems in a given time period. In reviewing the extent of alcohol consequences students have experienced, students may be prompted to consider changes in their drinking to avoid future problems.

1.7.2 Assessment Reactivity Therapeutic Benefit

Assessment reactivity is a double-edged sword, because it can be a confounder in intervention efficacy trials, but also used as a valuable therapeutic tool. For example, in WBIs, assessments are built into the programs and likely contribute to intervention effectiveness. In intensive alcohol treatment studies, follow up assessments have been
used as continuing care for therapeutic benefit (e.g., Gallen, 1974). This type of reactivity can be considered therapeutic assessment reactivity (TAR).

However, if research assessment reactivity (RAR) occurs during an intervention efficacy trial, detection of actual treatment effect sizes could be hampered. In research studies, reactivity can occur to baseline RA conducted pretreatment and/or to follow up assessments conducted post-treatment. For example, studies that employ intensive follow up assessments may observe improvements in control group drinking which could result in a lack of significant differences in alcohol consumption at follow up between the treatment and control groups (Bernstein, Bernstein, & Heeren, 2010). Such a lack of difference between groups may be interpreted as the treatment being ineffective and lead to decreased use of the intervention for clinical purposes. However, within subjects change in alcohol consumption from baseline to follow up may reveal decreases in consumption (Hester et al., 2012) and actually suggest that both groups experienced therapeutic interactions that helped decrease their alcohol consumption.

In study designs with a single follow up assessment point, reactivity to baseline RA could inflate estimates of treatment efficacy. There may be additive effects of RA plus a BI that could lead to the detection of greater within subjects estimates of efficacy than would occur with the administration of only a BI (e.g., Carey et al., 2006). This is a problem because when WBIs are disseminated to universities and used as general services for students, research assessments are not included and consequently universities may see lower effectiveness than was observed in efficacy trials.

Assessment reactivity can also be observed when comparing an experimental RA group (not combined with an intervention) to a no RA, no treatment control group
(e.g., McCambridge & Day, 2008). In studies that have utilized this study paradigm, the assumption is that the magnitude of reactivity to RA when not combined with an intervention may be similar to the effects of RA when combined with a WBI; however this assertion has not previously been tested.

1.7.3 Assessment Reactivity History

Early reports of behavior change in response to observation were from a program evaluating reasons for increased worker productivity in the Western Electrical Company’s Hawthorne Works in the 1920’s and 1930’s in Chicago, IL (Roethlisberger, Dickson, Wright, Pforzheimer, & Western Electric Company, 1939). The investigation suggested that the workers may have increased their productivity in response to changes in the work environment that indicated their productivity was being observed and measured, a phenomenon that was later termed the Hawthorne Effect. Although the actual reason for the increased productivity at the company was subsequently questioned (Franke & Kaul, 1978), the potential for people to change their behavior as a result of being observed or assessed had implications for clinical research in medicine and psychology and has been a topic of investigation and discussion since that time.

There have been many observational reports of assessment reactivity effects published in the literature that were designed to investigate intensive treatments for alcohol dependence. Multiple large scale studies surprisingly resulted in no differences between the treatment and control groups at follow up and the authors of these studies hypothesized that assessment reactivity to the follow up assessments may have led to improved outcomes in the control group, thus masking any treatment differences. For
example, Edwards et al. (1977) conducted a randomized controlled trial in which participants were assigned to receive brief advice to decrease drinking or intensive outpatient treatment with stepped-up inpatient treatment if necessary. Both groups completed monthly follow-up assessments. At the 12 month follow up, a comparison of both groups’ alcohol consumption during the past 12 months revealed that the groups did not significantly differ. The researchers hypothesized that assessment reactivity may have influenced the drinking levels of the control group participants.

In the multisite study Project Matching Alcoholism Treatments to Client Heterogeneity (Project MATCH Research Group, 1997a) in which participants were assigned to receive cognitive behavioral coping skills training, motivational enhancement, or 12 step facilitation, there were no substantial differences between the groups on subsequent alcohol consumption and the researchers suspected that intensive follow-up procedures may have clouded their ability to detect differences between the groups (Project MATCH Research Group, 1997b).

Investigators on several other randomized alcohol treatment studies have described similar experiences (Chick, Ritson, Connaughton, Stewart, & Chick, 1988; Elvy, Wells, & Baird, 1988; Zweben, Pearlman, & Li, 1988). In reviews of the literature, researchers have discussed the difficulty of detecting treatment effects while utilizing an intensive and lengthy follow-up procedure (Bien, Miller, & Tonigan, 1993; Breslin, O’Keeffe, Burrell, Ratliff-Crain, & Baum, March). To this point, a systematic review of the mechanisms of change in control group drinking in BI studies revealed that in 15 out of the 16 studies analyzed, control group drinking rates decreased by 11-
46% (Bernstein et al., 2010). Improvements in control group drinking could lead to a bias toward the null when comparing treatment to control groups (Kypri, 2007).

In addition to follow up procedures, researchers have also called attention to observations of assessment reactivity to RA administered before treatment in clinical trials, an area of the literature termed pretreatment change (Clifford & Maisto, 2000). Epstein et al. (2005) evaluated change in alcohol use at multiple time points before and after treatment began during a 6 month trial of individual or couples cognitive behavioral therapy for alcohol dependence. The researchers found that there was a significant reduction in alcohol consumption following the telephone screen and the intensive baseline RA. They also found that the degree of participants’ reactivity to the RA predicted level of alcohol consumption after treatment.

Similarly, in an analysis of pretreatment change within a clinical trial of cognitive behavioral therapy for alcohol dependence, Stasiewicz, Schlauch, Bradizza, Bole, and Coffey (2013) found significant increases in percent days abstinent and decreases in drinks per drinking day during the pretreatment phase of the study from the telephone screen to the baseline RA.

Although motivation may influence pretreatment changes in drinking particularly among those who recently enrolled in an alcohol treatment study, Morgenstern et al. (2007) found evidence of assessment reactivity even among participants who declined to participate in a treatment study. The researchers recruited participants for a study evaluating two alcohol use disorder treatments for men who have sex with men who were at risk for HIV transmission. There was a large group of individuals (n=102) who completed the baseline RA but subsequently declined
treatment. Treatment-declining group agreed to participate in follow up assessments, however, and the researchers identified a significant decrease in alcohol consumption following the baseline RA in this group.

Researchers evaluating substance use treatment for adolescents have also suspected that there is therapeutic efficacy to administering baseline RA (Kaminer & Godley, 2010). In an evaluation of cognitive behavioral therapy treatment for adolescents alcohol and drug use, Kaminer, Burleson, and Burke (2008) found a significant change to non-use of alcohol and other substances from the intake assessment to the first therapy session. The researchers also found that assessment reactivity predicted treatment response and asserted that assessment reactivity is a real construct.

1.7.4 Experimental Studies of Assessment Reactivity

Tests of Assessment Reactivity within Intensive Alcohol Treatment Studies

The majority of experimental tests of assessment reactivity have been conducted using college student samples; however there were two early studies that were designed to examine reactivity to follow up procedures in intensive alcohol treatment studies. Two initial tests of assessment reactivity did not observe reactivity effects; however both had substantial methodological problems, including low power to detect small differences between groups. Ogborne and Annis (1988) designed a study to test whether differential assessment reactivity effects would be observed with four different follow up procedures: 1) minimal follow up assessment at two points in 12 months, 2)
intensive follow up assessment, comprised of assessment every two weeks for the first three months and monthly follow up assessments through 12 months, 3) minimal follow up plus daily self-monitoring of alcohol use, and 4) intensive follow up plus self-monitoring. Participants for the study were those who were noncompliant with an alcohol treatment efficacy trial; once participants missed two sessions of the alcohol treatment, they were randomly assigned to one of the four follow up conditions. Perhaps not surprisingly, compliance with the follow up procedures and self-monitoring was very low. The researchers reported no significant differences between the four groups in alcohol consumption at the final follow up. This study was later criticized for being severely underpowered (40% power to detect effects) and for not maintaining experimental manipulations (Clifford & Davis, 2012).

The second pertinent study was conducted by Maisto, Sobell, Sobell, and Sanders (1985) and was designed to investigate whether there were differential outcomes between groups who completed alcohol treatment and then were randomized to either a basic follow up assessment condition or an extended follow up assessment condition. Like Ogborne and Annis (1988), the researchers did not detect any differences in alcohol consumption at 18 months post treatment. Clifford and Davis (2012) also criticized this study and reported that with 48 subjects total, the study had 25% power to detect small effects.

Experimental Tests of Assessment Reactivity in College Student Samples

Experimental tests of reactivity in college student samples have examined assessment reactivity effects alone or in combination with a brief treatment. Several
groups of researchers have tested the effects of assessment reactivity in isolation by comparing a group that completed a brief assessment battery to a control group that did not complete assessments. Groups were compared at follow up. Walters, Vader, Harris, and Jouriles (2009) found that college students who completed an assessment battery consisting of a 7 day drinking calendar, the Alcohol Use Disorders Identification Test (AUDIT), and the PBSS compared to delayed assessment controls, had significantly decreased risky drinking (lower peak blood alcohol content estimation scores and AUDIT scores) with Cohen’s d effect sizes of .34 and .37 respectively and significantly increased risk reduction behaviors based on the PBSS, with Cohen’s d effect size .35.

A similar study was conducted by Kypri, Langley, Saunders, and Cashell-Smith (2007) in which participants were randomized to receive a leaflet about drinking or a leaflet plus a 10 minute assessment battery that included a 14 day drinking calendar, the Alcohol Problems Scale, and the Academic Role Expectations Alcohol Scale. The researchers found that the students who completed the assessment battery had significantly lower overall alcohol consumption (geometric means ratio (GMR) = 0.82) and fewer heavy drinking episodes (GMR=0.66) compared to the control group at the 12 month follow up, but not at the 6 month follow up. The authors speculated that multiple assessment administrations may have been necessary to observe a reactivity effect.

Researchers have also found evidence of reactivity from the administration of a single questionnaire. McCambridge and Day (2008) randomized college students to an experimental group that completed the AUDIT or a control group that did not complete any assessments. The experimental and control groups were compared at 12 month
follow up, and the experimental group had significantly lower AUDIT scores (Cohen’s d= .23) and scores on the Leeds Dependence Questionnaire (Cohen’s d = .29) than the control group.

Parsing the effects of assessments in combination with an intervention can be more complicated than testing assessment reactivity effects in isolation; however several researchers have investigated this topic. Reactivity to assessments may either lead to overestimations or underestimations of intervention efficacy, depending on the study design. Some designs entail a comparison of two groups who receive intervention with or without RA, whereas other designs compare groups who both receive RA with or without intervention. In a study examining the additive effects of an in-person administration of the TLFB and live BI, participants who received a BI plus TLFB significantly reduced drinking and related problems at one month follow up compared to the BI only group, with Cohen’s d effect sizes ranging from .32 to .52 (Carey et al., 2006). Thus, in this study reactivity inflated the efficacy findings in the TLFB plus BI group.

Using a different study design, Hester et al. (2012) found that reactivity in their RA-only control group led to underestimations of WBI efficacy. The researchers used a design of RA+WBI compared to RA only and found that their RA only control group had significant reductions in alcohol use and related problems at the 1 month and 12 month follow ups, which resulted in small differences in alcohol consumption at follow up between the treatment and control groups. This led to a second study in which an intervention group was compared to a delayed assessment control condition. In the delayed assessment condition, participants did not complete baseline RA about
their alcohol use pretreatment, but provided information about their pretreatment alcohol use at the one month follow up assessment. In this study, the researchers found greater differences between the WBI group and the control group (Cohen’s $d = .82$) compared to the first study that observed an effect size of .35. In addition, a comparison of the RA control group from the first study to the delayed RA control from the second study revealed that at the 1 month follow up, the delayed assessment group had significantly greater drinks per week and peak estimated blood alcohol content. The Hester et al (2012) and Carey et al (2006) studies are prime examples of the difficulties that assessment reactivity can present in measuring accurate treatment effects.

The majority of reactivity studies in the literature have found evidence of reactivity, however the effect sizes in these studies have varied. In a meta-analysis of assessment reactivity studies conducted with college students and in hospital emergency departments, McCambridge and Kypri (2011) found that the differences between assessment and control conditions were equivalent to about one standard drink ($z= 1.94, p = .053$). When analyses were restricted to studies of college students, the pooled effect was stronger ($z= 2.46, p = .014$).

1.7.5 Conditions Under Which Assessment Reactivity May be Observed

There are several conditions that may influence whether assessment reactivity is observed, and the strength of assessment reactivity effects. First, the length and frequency of assessments can influence reactivity. In an experimental test of reactivity to follow up assessments in an alcohol treatment study, Clifford, Maisto, and Davis (2007) tested the influence of frequency of assessment and comprehensiveness (or
length) of assessments on alcohol use outcomes. Participants were randomized to one of four follow up groups: 1) frequent-comprehensive, 2) frequent-brief, 3) infrequent-comprehensive, or 4) infrequent-brief. The authors found that participants who were exposed to the frequent and comprehensive assessments had greater reactivity to the assessments, as demonstrated by significantly greater proportion of days abstinent, drinks per drinking day, and proportion of heavy drinking days than the other groups. The infrequent-brief assessment group had the highest levels of alcohol use among the four groups and thus exhibited the lowest levels of reactivity.

Initial evidence also suggests that there may be temporal effects of reactivity. Worden, McCrady, and Epstein (2008) investigated assessment reactivity to follow up assessments during a trial of cognitive behavioral therapy for alcohol dependent women. Follow up assessments were conducted in three month increments over an 18 month period. In secondary analyses of their TLFB follow up data, the researchers evaluated whether there were differences in alcohol use two weeks before the assessment, two weeks post assessment, and four weeks post assessment. Results indicated that participants had lower mean drinks per day in the two weeks immediately following the 12 month in-person follow up, but did not observe this effect at the other three telephone-based follow ups.

The time course in which assessment reactivity effects have been observed in research studies has varied substantially. For example, some researchers found evidence of reactivity up to one year post-RA administration (Walters, Vader, Harris, & Jouriles, 2009), but Carey et al. (2006) found significant reactivity post assessment but observed a diminution of the intensity of assessment reactivity effects at the three
month follow up intervention. The assessment reactivity studies to date have differed in the time points in which they conducted follow up assessments and the time course and conditions under which assessment reactivity effects are detectable require further inquiry.

Another factor that may contribute to assessment reactivity effects is the method of assessment administration. For example, Heather, Whitton, and Robertson (1986) conducted a study of a mailed self-help manual for heavy drinkers with follow up assessments administered either live via telephone or through mailed paper or pencil assessments. The researchers reported a greater reduction in alcohol consumption and alcohol-related problems in a group that had live, phone-based follow up assessments compared to a group that received the mailed assessment. The researchers questioned whether there may have been reactivity effects to the live interviews. No studies have directly compared whether there are differential reactivity effects from in-person versus web-based assessment administration, however investigations are warranted.

1.7.6 Factors that Influence Assessment Reactivity Effects

Factors that contribute to assessment reactivity are largely unknown. Clifford and Davis (2012) highlighted in their review of the assessment reactivity literature how the field’s understanding of assessment reactivity is “rudimentary” and called for more studies to evaluate factors that contribute to assessment reactivity and factors that moderate the relationship between assessment and behavior.

In one of the few studies in which researchers evaluated assessment reactivity mechanisms, Magill, Kahler, Monti, and Barnett (2012) examined whether in-person
assessment after a negative alcohol event led to decreases in alcohol consumption in college students compared to a no assessment control group and investigated whether gender and precollege AUDIT score moderated the relationship. Analyses revealed a three-way interaction effect between assessment condition, gender, and AUDIT score; however the magnitude of the effects was very small. The researchers found that the slope for change in the number of drinking days in the assessment group was steeper for females ($b = -0.03, p < .05$) and for participants who had high precollege AUDIT scores ($b = -0.003, p < .05$). For heavy drinking days, there was a negative slope for females assigned to the assessment compared to females assigned to the control group ($b = -0.03, p < .05$). The researchers speculated that females may have engaged in greater self-exploration after the interpersonal interaction with the assessment, which may have led the females to change their consumption more than the males. They also suggested that participants with higher baseline AUDIT scores may have experienced more alcohol related consequences and been more motivated to change their alcohol consumption than those who had less alcohol consequences.

Murray, Swan, Kiryluk, and Clarke (1988) measured smoking behavior in adolescents longitudinally and found significant differences in smoking between assessed and non-assessed girls, but not boys. Specifically, girls who completed the assessments significantly reduced their smoking, but boys did not. The researchers remarked that the effect could have been related to differences in girls’ and boys’ comfort in reporting smoking behaviors or may have been due to a bias in the group of participants who responded to the follow up assessments.
In a third study, Epstein et al. (2005) evaluated pre-treatment change in drinking among alcohol dependent women. They investigated whether stated goal of abstinence, stage of change, or interviewer were associated with assessment reactivity, but found no significant associations. With the exception of the Epstein et al (2005) study, there has been surprisingly little examination of the relationship between motivational factors and reactivity. Although Epstein et al (2005) did not find evidence that motivational factors were related to assessment reactivity effects, this area of research warrants additional examination because it is commonly assumed that assessment reactivity results from assessments increasing participants’ motivation to change or self-efficacy (Clifford & Davis, 2012; Schrimsher & Filtz, 2011).

1.7.7. Aspects of Assessment Reactivity Currently Unknown

Although assessment reactivity has been studied using various research paradigms and in different types of samples, there are still many unanswered questions. First, it is currently unknown whether there are additive effects of web-administered RA to a WBI on alcohol use at follow up. An experiment that tests alcohol use at follow up between an RA+WBI and a WBI-only group could answer this question. Further, understanding of the factors that contribute to reactivity are still largely unknown and an examination of the moderation effects of factors such as gender, baseline AUDIT indication of hazardous alcohol use, and motivation to change alcohol use is needed.

Testing the additive effects of RA within a treatment context could be conducted using a two or four arm design. Solomon (1949) proposed a four condition
experimental design to test assessment reactivity effects alone and in combination with a treatment. In what has been termed the Solomon four group design, study participants are randomized to an assessment + treatment group, treatment only group, assessment only group, or a no assessment no intervention group. With this design, alcohol use in the assessment + treatment group could be compared with the treatment only group to determine whether the groups significantly differed at follow up. With this design, researchers could also test whether comparing a treatment group to an assessment only control group results in similar or differential efficacy estimates than a comparison of a treatment group to a no assessment control group.

The Solomon four group studies that have been conducted to date have spanned a wide range of topics, including health screening for cervical cancer in Native American women (Dignan et al., 1998; Dignan et al., 1996), use of hearing protection among construction workers exposed to high noise levels (Lusk et al., 1999), brief advice or treatment for physically inactive adults (van Sluijs, van Poppel, Twisk, & van Mechelen, 2006), substance misuse in middle school (Campanelli, Dielman, Shope, Butchart, & Renner, 1989) and sex education (Kvalem, Sundet, Rivø, Eilertsen, & Bakketeig, 1996; Træen, 2003). The Campanelli et al. (1989) study revealed that students in the baseline RA condition reported significantly fewer peer problems related to their alcohol use at follow up compared to the no baseline assessment group. There were decreases in alcohol misuse in the RA group; however the effects did not reach significance.

No studies have conducted a Solomon four group design with college student samples using BIs or WBIs. While a Solomon design would allow for a comprehensive
examination of assessment reactivity effects within an intervention and in isolation, this
design requires substantial resources financially and a very large subject pool. An
alternative to answer the additive effect question is a two-arm design that tests the
effects of an RA+WBI group compared to a WBI-only group, as was used in this
dissertation study.

1.8 Study Purpose and Relevance of the Findings

This study filled a gap in the literature by testing the additive effects of web-
administered RA plus a WBI for heavy drinking college students in comparison to only
a WBI. Evaluating RAR in this context was important, because WBIs have been widely
disseminated to universities as general services outside the research context. If RAR did
significantly inflate efficacy results, recommendations could be made to universities of
how to obtain comparable effects with WBIs outside of a research context (e.g., suggest
using assessments plus the WBI). If RAR did not significantly inflate efficacy results,
findings of the current study would validate existing efficacy reports. Results would
also inform the field about how standard research methodology may influence
participant behavior and efficacy findings.
1.9 Specific Aims

There were four specific aims for the current study.

Specific Aim 1:

To test the additive effect of web-based research assessments to an existing web-based intervention on college student alcohol consumption, as measured by:

1) the total number of drinks consumed in the past month
2) peak estimated blood alcohol content in the past month
3) the number of times participants exceeded NIAAA guidelines in the past month

Hypothesis 1

Among college students who exceed NIAAA low-risk drinking guidelines, those randomized to the RA+WBI group will at one month follow up 1) report consuming fewer total alcoholic drinks (Hypothesis 1.1), 2) have lower peak estimated blood alcohol content (Hypothesis 1.2), and 3) exceed NIAAA guidelines fewer times (Hypothesis 1.3) compared with participants who received only the WBI.

Specific Aim 2:

To test the additive effect of web-based research assessments to an existing WBI on alcohol-related consequences among college students.

Hypothesis 2

Among college students who exceed NIAAA low-risk drinking guidelines, those randomized to the RA+WBI group will have significantly lower scores on the
Rutgers Alcohol Problem Index (White & Labouvie, 1989) at one month follow up compared with participants who received only a WBI.

Specific Aim 3:

To test the additive effect of web-based research assessments to an existing WBI on the use of protective self-control behaviors related to limiting and avoiding alcohol consumption among college students.

Hypothesis 3

Among college students who exceed NIAAA low-risk drinking guidelines, those randomized to the RA+WBI group will have higher scores on the Protective Behaviors Strategies Scale (Martens et al., 2005) at one month follow up compared with participants who received only the WBI.

Specific Aim 4:

To test factors that moderate the relationship between assessments and alcohol consumption, alcohol consequences, and protective behaviors use: gender, baseline AUDIT indication of hazardous alcohol use pattern, and motivation to change drinking behavior.

Hypothesis 4:

Gender: Degree of assessment reactivity observed will depend on level of gender; female participants will report lower levels of alcohol consumption and related
problems, and higher use of protective behaviors compared to their male counterparts in the RA+WBI condition at follow up. The differences on outcomes between male and females will be greater in the RA+WBI group compared to the WBI only group.

AUDIT score: Degree of assessment reactivity observed will depend on whether participants have hazardous alcohol use patterns at baseline as determined by a baseline AUDIT score of eight or higher. The differences in alcohol consumption, related problems, and protective behavior use between participants with hazardous alcohol use in the RA+WBI group compared to those of with hazardous use in the WBI-only condition will be greater than the differences between the RA+WBI and WBI only conditions in those with no hazardous alcohol use at baseline.

Importance of changing drinking: Degree of assessment reactivity exhibited will depend on the level of importance in changing drinking behavior that participants rate on a 10 point scale during the WBI. The differences in alcohol consumption, related problems, and protective behavior use between participants with high importance ratings in the RA+WBI compared to those of equivalent importance ratings in the WBI-only condition will be greater than the differences between the RA+WBI and WBI only conditions in those with lower importance ratings.

Confidence in changing drinking: Degree of assessment reactivity exhibited will depend on the level of confidence in changing drinking behavior that participants rate on a 10 point scale during the WBI. The differences in alcohol consumption, related
problems, and protective behavior use between participants with high confidence ratings in the RA+WBI compared to those of equivalent confidence ratings in the WBI-only condition will be greater than the differences between the RA+WBI and WBI only conditions in those with lower confidence ratings.

CHAPTER 2: METHODS

2.1. Study Design and Justification

The study was conducted entirely online, allowing students to participate at any time from any computer or mobile device with Internet access. Following approval from the University of Vermont’s Institution Review Board, undergraduate students attending universities in the United States and Canada were recruited to participate. Interested individuals accessed the study website through the study web address provided in the recruitment advertisements. At the study website, participants could read information about the study, provide their informed consent, and complete a screening questionnaire to determine their study eligibility. Eligible participants were randomized to one of two conditions: a research assessment plus web-based intervention condition (RA+WBI) or a WBI only condition.

Participants were blinded to the study hypothesis about assessment reactivity, because their awareness about the hypothesis may have altered their behavior during the investigation. The study purpose was presented in the consent form as a test of a new web-based alcohol assessment and feedback program for university students. However, students were told that if they were eligible, they would be randomized to
either a group that completes assessments plus the WBI, or a group that completes only the WBI.

One month following WBI completion, participants were emailed a request to complete a follow up RA online. A one-month follow up period to test for assessment reactivity effects was selected for several reasons. It was important to observe potential reactivity effects at their strongest and Carey et al. (2006) found evidence of significant diminution of intensity of RAR effects three months post-intervention. Additionally, accurate recall of a past event generally decreases as the time since the event increases (Ebbinghaus, 1913), which could lead to inaccuracies in reports of alcohol consumption. Finally, effects from WBIs have been shown to be most prominent in the first month post-WBI (Chiauzzi et al., 2010; Hester et al., 2012). Therefore, it was expected that any assessment reactivity effects should be sufficiently strong within the first month to be reliably detectable.

2.2 Randomization Procedure

Participants were randomized to the RA+WBI or WBI only condition and the randomization procedure was blocked on gender so that a similar amount of males and females would be assigned to each study group. Because of gender differences in rates of alcohol consumption and related consequences among college students (Kelly-Weeder, 2008; O’Malley & Johnston, 2002; Wechsler et al., 1995), it was important to distribute gender equally across groups so that gender did not confound the study results.
2.3. Incentives

Incentives were provided for completing the WBI and the follow up questionnaire. The incentive structure encouraged early study enrollment. For signing up for the study and completing the WBI, participants were entered into up to three drawings for iPad minis; the earlier participants completed the WBI relative to overall study recruitment, the more drawings they were entered into. The first 150 participants who completed the WBI were entered into three drawings, the second 150 participants were entered into two drawings, and the last 150 participants were entered into one drawing. For completing the follow up questionnaire, participants were entered into one of four drawings for a $500 Amazon.com gift card. UVM students who were enrolled in a UVM introductory psychology course at the time they enrolled in the study were offered an option to receive 5 bonus points in their psychology course for completing the WBI instead of being entered into the iPad mini drawings. A small minority of participants (n= 20) selected this option. These students received the same incentive for completing the follow up questionnaire as the other participants. The bonus point incentive option was tested in the first semester of recruitment but was not continued into the second and third semesters of recruitment due to low utilization of this option in the first semester.
2.4. Study Structure

Figure 1: Study Design

When participants accessed the study website, they first reached the study information page which included general information about the study purpose, the time requirement for participation, and incentives for participating. From the study information page, participants could click a button to access the informed consent webpage. The informed consent was a four page online document that provided detailed information about the study purpose, possible benefits and risks of participating, alternatives to participating, and provided details about requirements for participating in the study. Students were informed that they could discontinue their participation at any time. If students agreed to participate, they checked a box at the end
of the consent form confirming that were 18+ years old and then selected the button that said they were consenting to participate.

Pressing the consent button took participants to the screening webpage that was used to determine study eligibility. Eligibility criteria for the study consisted of the following: 1) current undergraduate student, 2) attending college/university in the U.S. or Canada, 3) self-reported alcohol consumption in excess of NIAAA guidelines for low-risk drinking: 4[5] or more drinks for women [men] at least one time in the past month, and 4) age 18-26. The screening form asked participants for the following information: gender, age, whether they were a current undergraduate student in the U.S. (yes/no), whether they were a current undergraduate student in Canada (yes/no), and the highest number of drinks they consumed in one day in the past month, with beer, wine, mixed drinks, and shots asked separately. The amount of alcohol that was considered to be one standard drink was listed in ounces and milliliters for U.S. and Canadian students on the screening form where participants were asked to report the number of each type of drink they consumed.

The website determined study eligibility and routed eligible participants to the incentive selection page, where participants selected either the drawing entry or bonus points options for completing the WBI. Following incentive selection, participants were sent to the demographics page. Individuals who were not eligible were informed they were not eligible for the study but were allowed to complete the WBI if they wished. Eligible participants were assigned a 4-digit Study ID and instructed to provide the following demographic information: year in school, race, ethnicity, school name, and the date and means through which they learned about the study. Participants provided
their university email address to confirm they were current students and their telephone number as a secondary contact method. Names were not collected for participants who chose the drawing incentives in order to limit the amount of personally identifying information that was collected. Participants who chose to receive bonus points in their psychology course for participating provided their names and filled out an online release of information form that allowed the study to release their name to the Psychology Department to confirm their participation in the study in order to receive the bonus points.

Following completion of the demographics form, participants were randomized. WBI-only participants were directed immediately to the WBI. For RA+WBI participants, a link first directed them to the RA and once they completed the RA, a link appeared directing participants to the WBI. Participation in the RA+WBI condition was expected to take 20-30 minutes. In the WBI-only participation was expected to take about 15-20 minutes. Participants identified themselves with their assigned Study ID for the RA and WBI.

One month following WBI completion, participants were sent an email with a link to the follow up questionnaire to be completed online.

2.5 Programming the Study Website and Protection of Confidential Information Stored Online

Since the research study was conducted online, strict protections were used to maintain data security and participant confidentiality. Data was collected at three different websites: the research study website, the WBI website, and UVM’s Lime
Survey Site that was used for baseline and follow up RA. Only the research study website collected participant identifying information. No identifying information was collected on the WBI website or the UVM Lime Survey website and participants identified themselves on these websites with a 6 digit numerical ID they were assigned when they enrolled in the study.

The study website stored participant consents, screening information, demographic information, and randomization assignments. The only alcohol use information that was stored on the site was the highest number of drinks consumed in a day in the past month. It was necessary to collect this information because it was needed to determine participant study eligibility. A professional web developer with a specialization in creating secure and stable web applications for business organizations and science and technology research volunteered to program the study website. The study website data were protected by 128 bit encryption with Secured Socket Layer (SSL) protection. An ID and password was required to access the data and were also encrypted as they were entered into the login site. The web programmer also secured the data stored by limiting access to the administrative portion of the study website by allowing only computers used by the PI or web programmer to access the data. Only the study Principal Investigator and the web programmer had access to the study website data.

Paper copies of the data were not used, as it was not necessary and would have increase the amount of places where participant identifying information was stored. Following study completion, a copy of the dataset was obtained from the study website and the web programmer wiped the website clean of study data.
The websites for the WBI program and baseline and follow up RA were separate, commercially available programs and thus stored their data on separate servers than the study website.

Data from the WBI program were stored on a secured server maintained by the San Diego State University Foundation (the creators of the WBI). The San Diego State University Foundation has a program designed for researchers who wish to use the WBI for research purposes, and thus facilitates the collection and storage of the WBI data on a secured web server. All WBI transactions were conducted across an encrypted Secure Sockets Layer (Transport Layer Security). WBI data was accessed only by the study PI and download after providing a secured username and password.

Data from the baseline and follow up RA were collected using UVM’s Lime Survey site that facilitated survey creation, management, and stored survey data. Data from the site could be accessed only by the study PI by providing her UVM ID and password.

Once data were downloaded from their respective websites, they were kept on the secured University of Vermont Medical School system network.

Data from all three websites were matched by participant study ID for subsequent data analysis.

**2.6. Recruitment**

Participants were recruited using both traditional recruitment techniques and through social media. The UVM Institution Review Board approved all recruitment techniques. Recruitment was conducted during the spring 2013, fall 2013, and spring 2014 university semesters. Social media recruitment was conducted by posting free
messages to social media platforms Facebook and Twitter. IRB-approved recruitment messages were posted on Facebook in U.S. and Canadian student groups and university pages. The IRB approved a 140 character version of recruitment message for Twitter and this message was “tweeted” to the Twitter community from individual accounts. Traditional recruitment consisted of posting flyers at UVM in places that were frequented by undergraduate students, including the cafeteria, student health center, residence halls, and the athletic center. The study was also listed one time in the weekly news email sent to all UVM undergraduates. No paid advertisements were used for recruitment.

Seasonal Recruitment

Recruitment was conducted during three consecutive university semesters. Recruiting participants during the summer was also tested for two weeks, however was discontinued after the PI observed a 20% larger drop in the rate of WBI completion compared to the completion rate during the school year and observed substantially lower alcohol use levels among the participants who enrolled during the summer.
2.7 Communication with Participants and Participant Tracking

Participants were contacted only via messages sent to their university email accounts and by telephone if they did not respond to the email. Participants were not contacted through social media in order to protect participant confidentiality. If participants contacted the study using a private Facebook message, a response to their query was sent to their university email address.

If participants were randomized but did not complete the RA and/or the WBI, they were emailed a maximum of three times a request to complete these components. Contact attempts were made in 24-hour increments. Each request to complete the remaining study components included a link that the participant could click that would start them at the closest point to where they left off. It was possible to start participants at different parts of the RA and at the beginning of the WBI, however the WBI did not record where participants exited the program if they did so prematurely, therefore they had to start at the beginning.

Participants were emailed an initial request to complete the follow up questionnaire. They were emailed a maximum of two additional times and called once if they did not respond to the initial request by completing the questionnaire. Each contact attempt was made 24 hours following the previous attempt. Assessment reliability was not expected to be compromised for those who finished the questionnaire several hours to days later; Miller et al. (2002) demonstrated that questionnaire reliability was retained for the drinking calendar and RAPI despite breaks
of between 3-48 hours in between when participants in the study started and finished the questionnaires.

### 2.8 Measures

Instruments to measure alcohol use, alcohol-related consequences, and protective behaviors related to alcohol consumption were chosen to reflect those commonly used in WBI studies to allow for generalizability of the study findings. The majority of WBI efficacy studies report using between 3-7 questionnaires for their baseline assessments. Although many studies do not report time length for completing the RA, studies that did mention time duration reported that their RA took 15 – 45 minutes (mean = 30 minutes) to complete (Hustad et al., 2010; Lee et al., 2014; Lewis et al., 2014; Paschall, Antin, Ringwalt, & Saltz, 2011). In these studies, the RA were reported to have taken between 25-100% as long as the actual WBI (Hustad et al., 2010; Lee et al., 2014; Lewis et al., 2014; Paschall et al., 2011). The RA in the current study consisted of 3 questionnaires and were expected to take approximately half as long as the WBI in order to maintain consistency with other studies in the literature. The RA were expected to take about 10-15 minutes to complete.

Both the baseline RA and the follow up questionnaire contained the same measurements.

The Daily Drinking Questionnaire (DDQ) is a calendar measure of frequency and quantity of alcohol consumption and asks participants to identify the number of drinks they consumed each day during a specified period in time (Collins, Parks, &
Marlatt, 1985). For the current study, participants were asked to report their alcohol consumption during the past 30 days. The DDQ is similar to the well-established Time Line Follow Back calendar-based interview that is considered the gold-standard in alcohol research. During the TLFB interview, the researcher gathers retrospective reports of a participant’s alcohol consumption by discussing alcohol consumption each day and reconstructs the period in time by discussing memorable events that occurred in order to facilitate recall (Sobell & Sobell, 1992). The TLFB has been used for many years, with reliability in clinical alcohol dependent samples (Sobell, Maisto, Sobell, & Cooper, 1979; Sobell et al., 1988; Sobell, Sobell, Klajner, & Pavan, 1986; Sobell et al., 1986), and the general population (Williams, Aitken, & Malin, 1985; Wolber, Carne, & Alexander, 1990). TLFB validity has been established when compared to collateral reports and biological samples via breathalyzer (Maisto et al., 1985; Polich, 1982).

The DDQ drinking calendar method is similar to the TLFB because it gathers reports of alcohol use but is more streamlined because it does not inquire about memorable events that occurred. An advantage of the DDQ is that it can be self-administered on a computer, which is ideal for web-based research studies. Researchers have examined test-retest reliability and concurrent validity for the self-administered DDQ. Miller et al. (2002) evaluated the test-retest reliability of the DDQ when administered both on paper and online in a large sample of college undergraduates. The researchers examined test-retest reliability using Pearson’s product momentary coefficients to assess association and intraclass correlations and found that test-retest reliabilities for both paper and web-based DDQ administration ranged from .59-.93, with the majority of associations being at or above .70, indicating sufficiently reliability
for research purposes. Rueger, Trela, Palmeri, and King (2012) also found evidence of sufficient test-retest reliability between a web-based DDQ and phone TLFB interview (all r’s .83-.93) and concurrent validity with the AUDIT in a sample of college students (r=.32, p<.001). Pedersen, Grow, Duncan, Neighbors, and Larimer (2012) also established concurrent validity of the web-based DDQ and an in-person TLFB administration in young adults. Pedersen et al (2012) found that there was no significant difference in alcohol consumption as measured in the in-person TLFB versus the online DDQ. The authors also computed several commonly used alcohol consumption measures from the calendar data (e.g. number of drinks consumed in the past 30 days) and found that correlations between the online and in-person administrations ranged from an r of .87-.95.

The DDQ has been used in most WBI efficacy trials and was included in the current study to assess alcohol use outcomes and to maintain consistency with the literature.

The Rutgers Alcohol Problem Index (White & Labouvie, 1989) is a 23-item tool that measures the number of alcohol-related consequences experienced in a given time period in adolescents and college students. The RAPI assesses both traditional physical consequences (e.g., tolerance, withdrawal symptoms, physical dependency) and consequences presumed to occur at higher rates in a college or university student population (e.g., missing school, not doing homework, going to school drunk). A one month timeframe was used for the questionnaire, consistent with previous research (Borsari & Carey, 2000; Carey & Correia, 1997). Participants are provided a list of
alcohol-related problems and asked to rate how frequently each occurred in the past month with the following response options: none, 1-2 times, 3-4 times, 5 or more times.

In the original RAPI article publication, the RAPI was intended to be scored as a one-dimensional construct as the count of the number of times an individual experienced each negative alcohol-related consequence in a given time period (White & Labouvie, 1989). However, other researchers have dichotomized each item to count the number of negative alcohol consequences individuals experienced (Larimer et al., 2001; Marlatt et al., 1998) and this strategy has also been established as a reliable and valid way to score the RAPI (Martens, Neighbors, Dams-O’Connor, Lee, & Larimer, 2007). In the current study, the RAPI was scored as the authors originally intended to gather a count of the number of times participants experienced each consequence in the past month. Although the dichotomous approach has been validated, dichotomizing a continuous variable can result in the loss of information measured; therefore the RAPI was scored as a count measure.

The RAPI has been found to have good internal consistency, with Chronbach’s alphas of .81 (Neal & Carey, 2004), .85 (Borsari & Carey, 2000), and .92 (White & Labouvie, 1989) in college student samples. The RAPI has also been demonstrated to have strong test-retest reliability in a large college student sample assessed at baseline, one month, six months, and 12 months with Pearson’s $r$’s ranging from .78 to .94 (Miller et al., 2002).

The RAPI was found to have predictive validity for alcohol use intensity as measured at multiple time points in the original publication article of the RAPI (White & Labouvie, 1989), however no additional validation studies have been conducted on
the validity of the RAPI as a continuous measure in other samples, despite this measure being used frequently in the literature. The convergent validity of the RAPI as scored by dichotomizing each response in the questionnaire has been demonstrated in a college student sample (Martens et al., 2007).

The Protective Behavioral Strategies Scale is a fifteen item questionnaire that evaluates alcohol-related risk reduction and self-control behaviors on three factors: limiting alcohol consumption, manner of drinking, and reduction of potentially dangerous alcohol related consequences (Martens et al., 2005). The scale's stemming question asks, “Please indicate the degree to which you engage in the following behaviors when using alcohol or 'partying’” on a 5-point Likert-type scale with the following response options: Never, Rarely, Sometimes, Pretty Often, and Always.

Researchers have established concurrent validity of all three PBSS subscales with alcohol consumption as measured by the DDQ (Pearson, Kite, & Henson, 2012). Although originally proposed with a three factor structure, the PBSS has been shown to have good reliability as a two factor model for limiting alcohol consumption (Chronbach’s alpha = 0.83) and avoiding alcohol consumption (Chronbach’s alpha = 0.84) (Novik & Boekeloo, 2011). Because the PBSS is a newer measure in alcohol research relative to other commonly used questionnaires, evaluating the reliability and validity of this measure is still a work in progress.

In addition to the study assessments, the WBI itself collected alcohol use information from participants. This information was collected from all participants and thus could be used to examine differences in randomization groups at baseline and be included for data analysis. The WBI administered the Alcohol Use Disorders
Identification Test (Babor, de la Fuente, Saunders, & Grant, 1992; Saunders, Aasland, Babor, & de la Fuente, 1993) and in addition the WBI asked participants for the following information: the highest number of drinks consumed in the past month in a single day, number of hours drank during the largest drinking episode in the past month, weight (used for peak blood alcohol content computation) and confidence and importance of changing alcohol use after the feedback was provided. In addition, the WBI collected demographic information about whether students were active in Greek fraternities/sororities, whether they were student athletes, and whether they resided on or off campus.

The Alcohol Use Disorders Identification Test (Babor et al., 1992; Babor & Grant, 1989) is a 10-item self-report screening measure developed by the World Health Organization to identify individuals with hazardous or harmful alcohol use (Babor et al., 1992; Babor & Grant, 1989). The AUDIT covers topics of alcohol consumption, drinking behavior, and alcohol-related problems. The AUDIT score range is 0-40, with higher scores indicating more harmful alcohol use.

Saunders et al. (1993) recommended AUDIT cut offs between 8-10 for determining hazardous alcohol use, thus AUDIT reliability and validity was evaluated using the cut points of 8-10. The researchers examined the sensitivity and specificity of the three AUDIT scales (alcohol consumption, drinking behavior, and alcohol-related problems) by comparing AUDIT scores to corresponding diagnostic information about hazardous alcohol use as collected from a structured clinical interview. With AUDIT cut points of 8, sensitivity and specificity for detecting hazardous alcohol use and related problems ranged from 91-100% and at 10, sensitivity and specificity was
between 88-98%. The validity of the AUDIT was established by comparing the AUDIT cut point of 8 to external reference groups of alcohol-dependent patients and non-drinkers. Of the alcohol-dependent individuals, 99% had an AUDIT score of 8 or more and of the non-drinkers, only 0.5% had a score of 8 or more (Saunders et al., 1993).

A meta-analysis of research on the AUDIT revealed that indices of AUDIT internal consistency (Chronbach’s alpha, item-total correlations) have generally been reported to be in the .80’s and that AUDIT sensitivity and specificity for identifying current hazardous alcohol use is high (Allen, Litten, Fertig, & Babor, 1997).

Motivation to change drinking behavior was briefly assessed during the WBI. After participants received their personalized feedback, the WBI asked participants to rate how important it was for them to change their drinking and how confident they were that they could change their drinking, each assessed on a separate 10-point scale from 0-9. Single-item ratings of importance and confidence in changing drinking behavior have been shown to predict change in alcohol consumption in men following a brief intervention (Bertholet, Gaume, Faouzi, Gmel, & Daeppen, 2012) and maintenance of sobriety in young adults following substance abuse treatment (Hoeppner, Kelly, Urbanoski, & Slaymaker, 2011). However, psychometric properties of such single item motivation measures have not been established. As mentioned previously, the WBI was a commercially available program and these proxy measures of motivation came pre-programmed into the WBI. Thus, evaluation of these variables was opportunistic in nature.
2.9 Web-Based Brief Intervention used in the Current Study

The WBI used for the current study was the Electronic Check Up to Go (ECHUG), a commercially available brief prevention intervention program for alcohol use managed by the non-profit San Diego State University Foundation that has demonstrated efficacy with a wide range of college students (Hustad et al., 2010; Doumas, Kane, Navarro, & Roman, 2011; Walters, Vader, Harris, Field, & Jouriles, 2009; Doumas & Andersen, 2009; Doumas et al., 2010; Walters et al., 2007). ECHUG collects information about student drinking and immediately presents a personalized feedback report which includes the following information: 1) quantity/frequency of drinking summary, 2) comparison to the U.S. and college drinking norms, 3) estimated level of risk for future alcohol use disorder (AUDIT score, genetic risk of alcoholism and tolerance), 4) amount of money spent per year on alcohol, and 5) advice and referral information. The program takes about 20 minutes to complete.
2.10 Data Analysis

2.10.1. Outcome Variables

Outcome variables consisted of measures of alcohol consumption in the past month, alcohol-related problems in the past month, and protective behaviors related to alcohol consumption used in the past month. Outcome variables were chosen in accordance with previous studies of WBI efficacy and assessment reactivity to allow for the current study results to be compared with the existing literature.

Alcohol Consumption

Outcome variables used to measure alcohol consumption in the past month included: 1) total number of drinks consumed in the past month (TOT), 2) peak estimated blood alcohol content in the past month (PBAC), and 3) the number of times participants drank past NIAAA guidelines in the past month (DP).

The TOT variable was constructed using the 30-day drinking calendar by summing the total number of drinks of beer, wine, mixed drinks, and shots participants reported consuming in the past month.

The DP variable was derived from the 30-day drinking calendar and was computed by summing the total number of times participants exceeded the NIAAA alcohol consumption guidelines for their gender (4+ drinks for females, 5+ drinks for males) in the past month.

The following information was used to compute the PBAC variable: the number of drinks consumed in a day, the number of hours over which participants drank that
day, and participant weight. The formula for PBAC was derived from (Matthews & Miller, 1979):

\[
BAC = \left(\frac{\text{consumption}}{2} \times (\text{GC/weight})\right) - (0.016 \times \text{hours})
\]

In this calculation, consumption equals the number of drinks in the drinking session, GC equals gender constant that accounts for differences in alcohol metabolism between genders: 9.0 for women and 7.5 for men (Borsari, Neal, Collins, & Carey, 2001; Matthews & Miller, 1979), weight equals weight in pounds, and hours equals the number of hours over which drinking occurred. To estimate BAC is necessary to account for the amount of alcohol metabolized since consumption. Metabolism rates show substantial individual differences (Kalant, 1971; Lelbach, 1974) and an average figure of 16 mg% per hour has been used and considered standard in the literature (Borsari et al., 2001; Matthews & Miller, 1979; Turner, Bauerle, & Shu, 2004).

BAC calculations have been found to be comparable in accuracy to in vivo breath tests of alcohol concentration (Carey & Hustad, 2002) and to correlate strongly with alcohol use and alcohol-related consequences in college student samples (Borsari et al., 2001; Turner et al., 2004).

The highest estimated BAC in the past month was chosen as the measure of PBAC.

**Alcohol-Related Problems**

Alcohol-related problems in the past month were measured by score on the RAPI. The RAPI score was a count of the number of times participants reported
experiencing alcohol-related problems in the past month, with higher scores indicating a higher frequency of alcohol-related problems.

**Protective Behaviors Used**

Score on the Protective Behaviors Strategies Scale was used to measure use of protective behaviors related to alcohol consumption in the past month. The PBSS variable was an ordinal scale, with higher scores indicating greater use of protective behaviors.

2.10.2. Multiple Regression

Multiple regression was used to determine whether the RA+WBI and the WBI only group significantly differed on three measures of alcohol use at follow up (TOT, PBAC, DP; Specific Aim 1), alcohol related problems (Specific Aim 2), and protective behaviors used in the past month (Specific Aim 3), while controlling for statistically and theoretically important confounding variables. Analyses were conducted with each dependent variable in a separate regression model. Potential confounding variables with theoretical and statistical significance were evaluated for inclusion in each regression model. Moderating effects of gender, AUDIT hazardous alcohol use, confidence rating, and importance ratings were evaluated for each dependent variable in separate analyses (Specific Aim 4). (When each variable was included as a moderating variable, it was not included as a confounding variable in the same model.)
Data Characteristics and Multiple Regression Assumptions

Prior to data analysis, data were examined for distribution characteristics, out of range values, and missing values. The PBSS variable was found to have a normal distribution based on skew and kurtosis (skew = 0.07, kurtosis = -0.05). In contrast, alcohol consumption variables TOT, DP, and PBAC and the RAPI variable all had substantial skew and kurtosis (TOT skew = 3.12, TOT kurtosis = 15.66; DP skew = 2.09, DP kurtosis = 6.03; PBAC skew = 1, PBAC kurtosis = 0.87; RAPI skew = 2.86, RAPI kurtosis = 11.3). The DP and RAPI variables were zero-inflated. Regarding out of range values, the follow up questionnaire was programmed to only accept values within an appropriate range for each corresponding variable in an effort to avoid this problem. However, frequencies on each variable were examined to confirm that the programming was effective in preventing out of range values. Participants were not able to skip questions during the questionnaire; therefore follow up information was complete for each section of the questionnaire that participants filled out. Details regarding missing data due to noncompletion of the follow up questionnaire are provided in the missing data section (below).

Multiple regression models were all initially run to check whether each model met the multiple regression assumptions of normal distribution of error (residual) variance and homogeneity of variance. If a model’s residual mean equaled zero, the residuals were all within two standard deviations from the mean, and observation of a scatterplot revealed no fanning out across the predicted ranges of Y values, a model was considered to meet these assumptions. Examination of the PBSS variable revealed that the residual variance parameters did not violate multiple regression assumptions.
This information indicated that it was not necessary to transform the PBSS variable or to use a regression method that specified a non-normal distribution. In contrast, the TOT, DP, PBAC, and RAPI variables had residual means close to 0, however the residuals reached beyond two standard deviations of their respective means and the residual variances did not appear to be homoscedastic upon examination of the scatterplots of standardized residual values and predicted values.

Based on evaluations of multiple regression assumptions for the models, multiple regression with least squares was determined to be appropriate for the normally distributed continuous variable PBSS score, however not for the other variables. Multiple regression with generalized linear models (GLM) specifying a negative binomial distribution was used for non-normally distributed count variables TOT, DP, and RAPI, because GLM allows for regression with variables that have non-normally distributed error variances and allows the magnitude of the variance of each measure to be function of its predicted value via a log link function (Nelder & Wedderburn, 1972; Hilbe, 2011). Because of the log link function, coefficients from a negative binomial regression analysis are presented as the log count of the dependent variable. Coefficients can be exponentiated to provide an Incidence Rate Ratio (IRR) for interpretation (Hilbe, 2011). DP and RAPI were both zero-inflated and analysis using zero-inflated negative binomial regression models would have been ideal for these variables. However R statistical software (R Development Core Team, 2012) was uses for data analysis and the R package that supports zero-inflated negative binomial regression analyses with multiple imputation is currently still in the beta testing phase.
Therefore, negative binomial regression using a more established R package was chosen for these variables as the second best option.

Because PBAC was a positively skewed continuous variable, a square root transformation was applied and then a least squares regression model was run with the transformed variable as the dependent variable. Both square root and log transformations can be used to transform positively skewed variables, however the square root transformation was chosen over the log transformation because there were zeros in the PBAC variable from participants who did not drink in the past month.

2.10.3. Confounding Variables

Variables with theoretical relevance and that have been demonstrated in the literature to have a relationship with the outcome variables were evaluated as potential confounding variables to be included in the analytic models. These variables consisted of the following: membership in a Greek fraternity or sorority (member/not a member), residence status on or off campus, student athlete status (student athlete/not), whether students were attending school in the U.S. or Canada, age, AUDIT hazardous alcohol use, highest number of drinks consumed in the past month in a single day as measured at baseline, age, class status (freshman, sophomore/junior, senior) importance in changing drinking, and confidence in changing drinking.

In addition, the time period in which participants enrolled in the study may also have had an influence on student drinking levels at follow up. Because participants were recruited over a 12 month period and a minority of participants (n=35, 9%) signed up for the study during the summer, it was important to consider temporal changes that
could relate to alcohol consumption. Three time variables were each evaluated as a representation of time during the study. Because these variables were likely to be highly correlated, the variable with the most influence on the beta and p value of the main outcome variable was included as the confounding variable representing time in the model. The first variable evaluated was day of the study year that participants enrolled in the study (1-365), with day 1 representing April 12, 2013, the first day of study recruitment. The second time variable evaluated was the time of year that participants signed up for the study, with the following time points: spring 2013, summer 2013, fall 2013, spring 2014.

In each of the four time points in the study, one time point contained participants who had levels of an outcome variable that substantially differed from the other time points. These natural breaks in the data allowed for the construction of a third variable with a dichotomous structure, with those in the outlying group compared to the others. Participants during the summer had substantially lower alcohol consumption (TOT) in the past month at follow up compared to those enrolled during the university semesters (mean 51.3 drinks vs 22.5 drinks; \( t(397) = 2.20, p = .03 \)). Participants who enrolled in the summer also had significantly fewer mean DP days, (2.97 versus 0.90; \( t(397) = 2.52, p = .001 \)). For alcohol-related problems, participants recruited in the spring 2013 semester had significantly lower mean RAPI scores at follow up compared to the other time points (5.7 versus 4.0, \( t(370) = 2.26, p = .02 \)). The differences in protective behavior use were less substantial, however students recruited during the spring 2014 semester had slightly higher PBSS scores compared to the participants from the other time points (44.8 versus 46.2, \( t(369) = -1.03, p = .30 \)).
Therefore, dichotomous variables were created for each outcome measure to compare scores from the time period with the specified difference in scores to the rest of the sample.

Two measures of alcohol use and related problems have been found to be strongly associated with subsequent alcohol consumption and consequences. First, the AUDIT has been shown extensively to predict alcohol consumption and alcohol related problems (e.g., Allen et al., 1997; Conigrave, Saunders, & Reznik, 1995; Dawson, Grant, Stinson, & Zhou, 2005). The majority of studies used the established binary measure of a hazardous alcohol use pattern with a score of 8 or higher indicating hazardous alcohol use; the reliability and validity of the AUDIT with a score of 8 or higher as an indicator of hazardous alcohol use pattern has previously been demonstrated (Saunders et al., 1993). In addition, previous alcohol consumption level has been found to predict future drinking levels (Tucker, Gavornik, Vuchinich, Rudd, & Harris, 1989). Researchers often control for baseline AUDIT hazardous alcohol use and baseline drinking status in their analyses of drinking and alcohol problems in order to account for variance in analytic models that is accounted for by these variables. In the current study, AUDIT and highest number of drinks consumed in one day in the past month were each evaluated as potential confounding variables in order to account for variance in alcohol consumption and alcohol problems at follow up that could be explained by these baseline measures. Baseline AUDIT was included in the models as a dichotomous variable with a score of 8 or higher indicating hazardous alcohol use.

Finally, motivation to change has been demonstrated to be related to alcohol consumption among college students. Researchers have found that in college student
samples, readiness to change predicted reductions in drinking outcomes (Carey, Henson, Carey, & Maisto, 2007; Fromme & Corbin, 2004; Kaysen, Lee, LaBrie, & Tollison, 2009) and that readiness to change moderated the effects between a WBI and alcohol consumption at follow up (Capone & Wood, 2009). Single-item ratings of importance and confidence in changing drinking behavior have been shown to predict change in alcohol consumption in men following a brief intervention (Bertholet et al., 2012) and maintenance of sobriety in young adults following substance abuse treatment (Hoeppner et al., 2011). It was plausible that confidence and importance in changing drinking as rated during the WBI may have influenced alcohol consumption in participants in the current study. Importance and confidence ratings in changing drinking were each evaluated as potential confounding variables in the analytic models.

Gender was not evaluated as a potential confounding variable to be included in the analytic models, because the randomization procedure was blocked on gender. A similar proportion of males and females were randomized to each group.

2.10.4. Moderating variables

The moderating effects of four variables: baseline AUDIT, gender, importance rating for changing drinking, and confidence rating for changing drinking were each examined in separate analytic models for each of the dependent variables. For each moderating variable, an interaction term between the variable of interest and the randomization group variable was added to each of the five models. (When variables were included in models as moderating variables, they were not included as covariates in the same models.)
2.10.5. Validation of WBI Efficacy

Although only one group completed baseline RA about their alcohol use before the WBI, all participants answered the screening question about the highest number of drinks they consumed in one day in the past month. In the WBI, participants provided the number of hours over which they drank during the highest drinking episode in the past month. Using this information, PBAC in the past month at baseline was calculated for all participants who completed the WBI. Change in PBAC from baseline to follow up was modeled to determine whether the effect size of the WBI was consistent with the literature. A repeated measures analysis of covariance (ANCOVA) model was constructed with baseline PBAC predicting PBAC at follow up. The same potential confounding variables were evaluated for this model as were for the main analyses. The final model included the following potential confounding variables: randomization group, baseline AUDIT, and importance rating in changing drinking.

2.10.6. Examination of Missing Data and Multiple Imputation

Twenty-seven percent of participants did not provide alcohol use data at follow up and 32% did not provide data for the RAPI and PBSS at follow up. Due to the substantial amount of missing follow up data, methods for dealing with missing data during data analysis were considered. Rubin (1976) presented a classification system for missing data with three missing data mechanisms: missing at random (MAR), missing completely at random (MCAR), and missing not at random (MNAR). The difference between the three missing data mechanisms is in the nature of the
relationship between the variable on which there is missing data, and the available data. Data are considered to be MAR when the probability of missing data on a variable Y is related to some other measured variable (or variables) in the analysis but not to the actual values of Y itself. Thus, the probability of missingness is related to the observed portion of the data. A limitation of this mechanism is that there is no practical way to establish whether the missing data are only a function of the other variables measured in the model and thus, there is no way to test data to determine whether they satisfy the MAR mechanism assumption (Enders, 2010). Data are considered to be MCAR when they are have no relationship with the other measured variables in a model and are unrelated to the values of the variable itself, and are thus missing completely at random. MNAR data occur when the missing data on a variable are related to the values of the variable itself, after controlling for other variables.

Given the three mechanisms, it was determined that the missing data for the current study should be considered MAR, because some relationship between alcohol use and other variables in the analysis models was expected to exist in relation to the missing data. Data were not expected to be missing completely at random and thus the MCAR mechanism was not considered to be a plausible representation of the missing data in the current dataset. While it was expected that alcohol use may have had a relationship with other variables that were related to the missingness, it was not expected that alcohol use was directly related to the missing data, therefore the MNAR mechanism was considered too restrictive in describing missingness for the current data.
Once the likely mechanism of missingness was determined, it was necessary to select a data analytic strategy to deal with the missing data. While listwise and pairwise deletion of missing data are very common approaches to handling missing data in social and behavioral sciences, these approaches assume MCAR and can produce biased parameter estimates if this assumption does not hold (Peugh & Enders, 2004). In addition, these approaches can reduce statistical power and are generally not recommended unless the amount of missing data is very small (Enders, 2010).

Multiple imputation is a robust statistical technique that produces unbiased parameter estimates with MAR data (Enders, 2010) and was selected for analysis with missing data for the current study. Multiple imputation consists of three stages: a data imputation phase, an analysis phase, and a pooling phase (Enders, 2010). In the imputation phase, estimates of the mean vector and covariance matrix are used to create a regression equation that predicts the missing values in the variable based on the observed values of the variable and related variables in the dataset. For each new imputed dataset, filled-in data from the previous imputed dataset are used to estimate the mean vector and covariance matrix for the next dataset, and a random residual term is selected from the normal distribution. Following imputation of N datasets, multiple regression analyses are run on each dataset and results from all analyses are pooled to provide final parameter estimates.

All analyses were conducted using R statistical software (R Development Core Team, 2012) with the Amelia() package (Honaker, King, & Blackwell, 2011) for data imputation and the Zelig () package (Imai, King, & Lau, 2007, 2008) for multiple regression with least squares and GLM and for pooling regression results.
The number of imputed datasets can have a substantial impact on statistical power. Although imputation with 10 datasets has previously been considered sufficient, Graham, Olchowski, and Gilreath (2007) conducted computer simulation studies that demonstrated that using 20 datasets improves power beyond what imputation with 10 provides. Therefore, 20 datasets were imputed for the current data.

Because data were considered MAR, it was important to determine what variables may have had a relationship to the dependent variables and may have been related to missingness. These variables would need to be included in the datasets to be used for imputation so that the dependent variables relationship with other variables could be taken into account during imputation. It is generally recommended that at least all of the variables to be included in the statistical analysis should be included, however other variables may also be included if they are suspected to be related to the missing data (Enders, 2010).

Variables that were provided to the Amelia () package for data imputation consisted of the following: baseline AUDIT score, highest number of drinks consumed in the past month at baseline, gender, age, residence status, athletic status, Greek status, class status, country in which participants were attending college, confidence in changing alcohol use, and importance of changing alcohol use. Total number of drinks consumed in the past month as measured at follow up was also included as a variable in the datasets for imputing RAPI and PBSS score. The drinking calendar data was the first section of the follow up survey and 5% of participants completed the first section but did not complete the RAPI and PBSS because they were in the last part of the questionnaire. Therefore, past month alcohol use at follow up was included in the
datasets provided to the *Amelia* () package to aid in imputation of RAPI and PBSS scores.

The *Amelia* () package also provides options for transforming variables during the imputation process to aid in replicating the distribution of the imputed variables. These transformations are conducted during imputation only and are returned to the imputed datasets in their original untransformed structure. Multiple imputation was first conducted for all dependent variables without transformations and the imputed variable distributions were inspected to determine whether they adequately replicated the distribution of the variable from the original dataset. It was observed that the following imputed variables did not adequately replicate the distribution of the dependent variables from the original dataset: PBAC, TOT, RAPI, and DP. This was likely because *Amelia* () assumes normal distribution of residual variance during imputation. Square root transformations were applied to these variables (which did not have normally distributed residual variances) during the imputation process (and were subsequently returned to the dataset in their original structure). These within-imputation transformations produced imputed variables that more appropriately represented the distributions of the variables in the original dataset.

Separate sets of 20 datasets were imputed for each of the 5 dependent variables. Following the generation of each of the 20 datasets, the imputed variables were checked to confirm there were no out of range values and in the case of ordinal variables, that only full integers were imputed.
2.10.7. Model building

For analyses with each dependent variable, potential confounding variables were examined to determine whether they influenced the beta or p values of the dependent variable and to evaluate the beta of the confounder and whether it had a significant p value. Each of 12 potential confounders were included in separate multiple regression equations for each dependent variable. In the interest of presenting the most parsimonious models, the final model for each dependent variable included only potential confounding variables that influenced the beta or p values of the dependent variable and/or significantly predicted the outcome.

Moderating variables were evaluated in separate analysis for each dependent variable. Basic models for each dependent variable were run with the moderating variable and without the confounding variables. Second, the moderating variable was added to each final model to determine whether the moderating effects remained significant following the addition of potential confounding variables.

2.10.8. Final Models for Each Outcome Variable

For each of the three separate alcohol use models in which randomization group predicted alcohol outcome (TOT, PBAC, or DP) the following confounding variables were selected for inclusion into each of the models: total number of drinks consumed in a single day in the past month as reported at baseline, baseline AUDIT, athletic status, summer enrollment, and importance in changing drinking. Other variables - confidence in changing drinking, age, residence status on or off campus, Greek membership, attending college in the U.S. or Canada, class status, day of the study year, and time
period of study enrollment - were evaluated as potential confounding variables, however they were not included in the final models because they were not found to influence beta or p values of the main outcome, and/or did not significantly predict the outcome.

For the regression model with randomization group predicting RAPI score, the following confounding variables were selected for inclusion into the model: total number of drinks consumed in a single day in the past month as reported at baseline, baseline AUDIT, importance rating, and spring 2013 enrollment. In order to present the most parsimonious model, the other aforementioned variables were not included in the models because they did not influence the beta or p values of the outcome variable and/or because they did not have a significant effect on the outcome variable.

For the regression model with randomization group predicting PBSS score, the following confounding variables were selected for inclusion into the model: total number of drinks consumed in a single day in the past month as reported at baseline, baseline AUDIT, attending college in the U.S./Canada, importance in changing drinking, and spring 2014 enrollment. The other aforementioned variables were not included in the final model because they were determined to not influence the relationship between randomization group and PBSS score, and/or did not significantly predict the outcome.

2.10.9. Examination of Baseline Differences in Randomization Groups

Descriptive statistics, two-tailed independent samples t-tests and Pearson’s chi-square tests of independence were used to determine whether the RA+WBI and WBI
only groups significantly differed on demographic or on alcohol-related measures at baseline. Independent samples t-tests were used to compare randomized groups on continuous variables and chi-square tests of independence were used to compare randomized groups on dichotomous variables. T-tests were conducted on the following continuous measures: highest number of drinks consumed in the past month, age, confidence in changing drinking behavior, and importance of changing drinking behavior. Chi-square tests of independence were conducted with the following dichotomous variables: gender, residence status, baseline AUDIT, college athlete, Greek membership, class, country attending college (U.S./Canada), completion during each of the 4 study time points (yes/no for each time point).

2.10.10. Power Calculation

Figure 2: Power Calculation

*Note: $f^2$ effect size of 0.0277 is equivalent to a Cohen’s $d$ effect size of .33

Determination of an adequate study N and follow up rate were based on the literature and power calculations using G*POWER (Faul, 2010). The majority of
studies in the literature on college students have found assessment reactivity effect sizes to be at or above a Cohen’s d of 0.30. For example, Walters, Vader, Harris, and Jouriles (2009) found evidence of significant reactivity effects on PBAC with effect size \( d = 0.373 \), lower alcohol consequences scores (\( d = 0.352 \)) and use of protective behaviors related to alcohol consumption (\( d = 0.352 \)) at follow up. In another study, researchers found that participants who received a live BI plus live TLFB significantly reduced drinking and related problems at one month follow up compared to the live BI only (no TLFB), with Cohen’s d effect sizes ranging from \( .32 \) to \( .52 \) (Carey et al., 2006).

McCambridge and Day (2008) found evidence of lower alcohol consequences at follow up, with a Cohen’s d = \( .23 \). Finally, Hester et al. (2012) found evidence of assessment reactivity in their assessment-only control group, which had decreased PBAC and alcohol consequences at follow up, with Cohen’s d’s = \( .30 \), \( .36 \).

In a meta-analysis of assessment reactivity studies in the literature, McCambridge and Kypri (2011) determined that assessment reactivity effect sizes were slightly larger in college student samples compared to studies conducted in primary care and hospital emergency departments, therefore effect sizes were selected in accordance with the assessment reactivity studies with college student samples.

Based on the aforementioned findings, an effect size of \( .33 \) was assumed for all outcome variables. A priori power analysis using standard values for Type I error (.05) and Type II error (.95), 80% power, and effect size of \( .33 \) indicated that 176 participants would be needed for each study group. Based on the WBI completion rates reported in the literature (Hester et al., 2012; Kypri et al., 2004; Walters et al., 2007), a 10% drop out rate was allotted for participants who did not complete the WBI.
Considering participant attrition at *both* the WBI and follow up, a total of 853 randomized participants were required for the study to achieve adequate statistical power to detect small effect sizes.
CHAPTER 3: RESULTS

3.1 Recruitment

Figure 1 presents a flowchart of recruitment and retention in the study through the one month follow up. Of the 1,691 participants who consented to participate in the study, 1,463 (87%) completed the eligibility screening and 1,091 (75%) were determined to be eligible for the study. Of the 372 individuals who were not eligible, the reasons for ineligibility were the following: did not drink past NIAAA guidelines in the past month (n= 276), were not enrolled in an undergraduate program (n= 70), were enrolled in an undergraduate program, but not in the U.S. or Canada (n = 11), and were above 26 years of age (n = 15).
Figure 3. Recruitment of Study Participants and Participant Flow
Of the participants who were eligible, 856 (78%) completed the demographics questionnaire and were randomized to either the RA+WBI group (n=429) or the WBI only group (n=427). The randomization blocking on gender was successful; 31% of participants assigned to the RA+WBI condition were males and 30% of participants assigned to the WBI only condition were males.

Sixty-four percent of randomized participants completed the WBI. A chi square test of differences revealed that a significantly higher proportion of the control group completed the WBI (74%) compared to the experimental group (54%; \( \chi^2 = 9.10, p = .003 \)). The RA took a median of 11 minutes (mean = 17.1, SD = 15.6) to complete. A substantial minority of participants (19%) took over 1 standard deviation longer (33-61 minutes) to complete the assessments than the rest of the sample. There were only three questionnaires in the RA and it is likely that these participants took a break during the assessments and returned to complete them later. The WBI did not timestamp when participants started the program; only the time they completed the program was recorded. Thus, it was not possible to compute duration of the WBI for the WBI only group. However, it was possible to compute the average time duration from RA completion to WBI completion for those who were in the RA+WBI condition. For those participants, the WBI took a median of 16.1 minutes (mean = 19.1, SD = 11.1) to complete.

Of those who completed the WBI, 371 (68%) completed the follow up questionnaire and an additional 30 participants (5%) provided complete alcohol use data in the 30-day drinking calendar at follow up, but did not complete the RAPI and
PBSS that were positioned at the end of the questionnaire. There was no differential follow up completion rate by randomization group (67% completion for RA+WBI participants, 68% for WBI only participants).

### 3.2 Descriptive Statistics

Table 1 provides demographic characteristics and baseline alcohol use information for the final sample of 856 participants who were randomized in the study. The vast majority of participants (87%) reported hearing about the study through social media platform Facebook. Participants were from 30 U.S. states and four Canadian provinces and locations that garnered the most participants were Vermont (n=328), Ontario (n=109), New York (n=46), and Quebec (n=41).
Table 1: Participant Demographic Characteristics

<table>
<thead>
<tr>
<th>Demographic Characteristic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age; mean (standard deviation)</td>
<td>20.00 (1.23)</td>
</tr>
<tr>
<td>Year in School</td>
<td></td>
</tr>
<tr>
<td>Freshman = 112 (13%)</td>
<td></td>
</tr>
<tr>
<td>Sophomore = 244 (29%)</td>
<td></td>
</tr>
<tr>
<td>Junior = 295 (34%)</td>
<td></td>
</tr>
<tr>
<td>Senior = 202 (24%)</td>
<td></td>
</tr>
<tr>
<td>Information not provided = 3</td>
<td></td>
</tr>
<tr>
<td>Country in which participants were attending college</td>
<td></td>
</tr>
<tr>
<td>US = 686 (80%)</td>
<td></td>
</tr>
<tr>
<td>Canada = 170 (20%)</td>
<td></td>
</tr>
<tr>
<td>University of Vermont Students</td>
<td>327 (38 % of total sample)</td>
</tr>
<tr>
<td>Race, Ethnicity</td>
<td></td>
</tr>
<tr>
<td>Caucasian, non-Hispanic = 705 (82%)</td>
<td></td>
</tr>
<tr>
<td>Asian = 53 (6%)</td>
<td></td>
</tr>
<tr>
<td>Black/African American = 13 (2%)</td>
<td></td>
</tr>
<tr>
<td>Multiple races = 45 (6%)</td>
<td></td>
</tr>
<tr>
<td>Hispanic = 37 (4%)</td>
<td></td>
</tr>
<tr>
<td>Information not provided = 3</td>
<td></td>
</tr>
<tr>
<td>Source of Recruitment</td>
<td></td>
</tr>
<tr>
<td>Facebook = 746 (87%)</td>
<td></td>
</tr>
<tr>
<td>University News Email = 42 (5%)</td>
<td></td>
</tr>
<tr>
<td>Flyer = 31 (4%)</td>
<td></td>
</tr>
<tr>
<td>Friend = 21 (2.5%)</td>
<td></td>
</tr>
<tr>
<td>Twitter = 6 (0.6%)</td>
<td></td>
</tr>
<tr>
<td>Health Center = 4 (0.5%)</td>
<td></td>
</tr>
<tr>
<td>Counseling Center = 3 (0.4%)</td>
<td></td>
</tr>
<tr>
<td>Information not provided = 3</td>
<td></td>
</tr>
<tr>
<td>Highest number of drinks consumed in the past month in a single day</td>
<td>Mean = 10.79</td>
</tr>
<tr>
<td></td>
<td>Median = 8</td>
</tr>
<tr>
<td></td>
<td>SD = 6.79</td>
</tr>
</tbody>
</table>
3.3 Tests for Baseline Differences between Randomized Groups

There were no significant differences at baseline between the two randomized groups on the highest number of drinks consumed in one day in the past month \((t(406)=0.83, p = .41)\), age \((t(420)=1.15, p = 0.25)\), confidence in changing drinking behavior \((t(433)=1.58, p = 0.12)\), importance of changing drinking behavior \((t(405)=0.19, p = .85)\), residence status \((\chi^2_1 = 0.05, p = .82)\), Greek membership \((\chi^2_1 = 0.40, p = .53)\), class status \((\chi^2_1 = 0.62, p = .43)\), county in which participants were attending college \((\chi^2_1 = 2.24, p = .13)\), or enrolling in the study during the summer \((\chi^2_1 = 0.17, p = .68)\). There were equal proportions of athletes and nonathletes in each randomized group, and equal proportions of baseline hazardous and nonhazardous drinkers in each randomization group \((\text{AUDIT: } \chi^2_1 = 0.00, p = 1.00, \text{ athletic status: } \chi^2_1 = 0.00, p = 1.00)\).

3.4 Comparison of Participants who Completed the WBI Versus Those Who Did Not

Table 2 presents a comparison of participant baseline alcohol use for those who completed the WBI and those who did not for both randomized groups. In both groups, participants who did not complete the WBI had significantly higher baseline alcohol use compared to participants in each respective group who completed the WBI \((\text{RA+WBI: } t(420) = 2.38, p = 0.01; \text{ WBI only: } t(424) = 5.56, p = .002)\). However, participants from each randomized group who completed the WBI were not significantly different from each other on baseline alcohol consumption \((t(406) = 0.83, p = 0.41)\).
Table 2. Comparison of Participants Who Completed the WBI to those Who Did Not

<table>
<thead>
<tr>
<th>Completed the WBI</th>
<th>Did not complete the WBI</th>
</tr>
</thead>
<tbody>
<tr>
<td>RA+WBI condition</td>
<td>WBI only condition</td>
</tr>
<tr>
<td>Mean highest number of drinks consumed in a single day in the past month</td>
<td>9.65</td>
</tr>
</tbody>
</table>

Descriptive Statistics for the Outcome Measures

Participants drank a mean of 49.78 drinks (median = 34.00 drinks) in the past month as measured at follow up. Participants had a mean PBAC of 0.13 (median = 0.11) and drank past NIAAA guidelines a mean of 2.86 times (median = 1) in the past month at follow up. The mean RAPI score for the sample was 5.24 (median = 3) and the mean PBSS score was 45.15 (median = 45.00) at follow up.

3.5 Multiple Regression with Multiply Imputed Datasets

3.5.1 Results for Specific Aim 1

Results of the multiple regression analysis using multiply imputed datasets revealed that Hypothesis 1.1 was not supported: randomization group did not significantly predict the total number of drinks consumed in the past month (TOT), while controlling for the total number of drinks consumed in a single day in the past month at baseline, baseline AUDIT, athletic status, summer enrollment, and importance in changing drinking (Table 3). There were no significant differences in the mean expected log count of total drinks consumed in the past month for participants who
completed research assessments before the WBI compared to those who were not assigned to complete assessments before the WBI. The difference in expected log counts of the total number of drinks consumed in the past month was expected to be .04 units higher for participants in the experimental condition compared to the control condition, while holding the other variables constant in the model. Variables that did significantly predict TOT were baseline AUDIT, baseline highest number of drinks consumed in a single day in the past month, athletic status, and summer enrollment (Table 3).

Table 3

Main effect of randomization assignment on the total number of drinks consumed in the past month at follow up (TOT).

<table>
<thead>
<tr>
<th></th>
<th>Difference in log count</th>
<th>SE</th>
<th>p</th>
<th>95%CI (low, high)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>2.99</td>
<td>0.09</td>
<td>.001</td>
<td>2.81, 3.18</td>
</tr>
<tr>
<td>Randomization group</td>
<td>0.04</td>
<td>0.10</td>
<td>.69</td>
<td>-0.15, 0.22</td>
</tr>
<tr>
<td>Baseline AUDIT</td>
<td>0.74</td>
<td>0.11</td>
<td>.001</td>
<td>0.53, 0.95</td>
</tr>
<tr>
<td>Baseline highest drinks</td>
<td>0.05</td>
<td>0.01</td>
<td>.001</td>
<td>0.04, 0.06</td>
</tr>
<tr>
<td>Importance rating</td>
<td>-0.02</td>
<td>0.02</td>
<td>.40</td>
<td>-0.06, 0.02</td>
</tr>
<tr>
<td>Athlete</td>
<td>-0.26</td>
<td>0.13</td>
<td>.05</td>
<td>-0.53, -0.001</td>
</tr>
<tr>
<td>Summer Enrollment</td>
<td>-0.25</td>
<td>0.23</td>
<td>.02</td>
<td>-0.97, -0.08</td>
</tr>
</tbody>
</table>

*Note. DV: dependent variable; SE: standard error; p: p-value; CI: confidence interval.

Negative binomial regression coefficients are presented as the difference in log counts.
Results of the multiple regression analysis revealed that randomization assignment did not significantly predict peak blood alcohol concentration in the past month (PBAC), while controlling for the total number of drinks consumed in a single day in the past month at baseline, baseline AUDIT, athletic status, summer enrollment, and importance in changing drinking (Table 4). Hypothesis 1.2, that completing additional assessments in addition to the WBI would lead to lower PBAC at follow up compared to those who were not assigned to complete assessments before the WBI, was not supported. Participants in the experimental condition had a mean difference in the square root of the peak estimated blood alcohol content of .01 units compared to the control condition, while holding the other variables constant in the model. Participants in the experimental group had an estimated .0001 higher PBAC in the past month compared to control group participants. Variables in the model that did significantly predict PBAC were baseline AUDIT, baseline highest number of drinks consumed in the past month, and importance rating.
Table 4

Main effect of randomization assignment on peak estimated blood alcohol concentration (PBAC) in the past month.

<table>
<thead>
<tr>
<th>DV: Peak estimated blood alcohol concentration (PBAC) in the past month</th>
<th>$b$</th>
<th>SE</th>
<th>$p$</th>
<th>95%CI (low, high)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.24</td>
<td>0.02</td>
<td>.001</td>
<td>0.20, 0.27</td>
</tr>
<tr>
<td>Randomization group</td>
<td>0.01</td>
<td>0.01</td>
<td>.84</td>
<td>-0.03, 0.03</td>
</tr>
<tr>
<td>Baseline AUDIT</td>
<td>0.12</td>
<td>0.02</td>
<td>.001</td>
<td>0.08, 0.15</td>
</tr>
<tr>
<td>Baseline highest drinks</td>
<td>0.005</td>
<td>0.002</td>
<td>.001</td>
<td>0.002, 0.009</td>
</tr>
<tr>
<td>Importance rating</td>
<td>-0.008</td>
<td>0.004</td>
<td>.03</td>
<td>-0.015, -0.001</td>
</tr>
<tr>
<td>Athlete</td>
<td>-0.04</td>
<td>0.02</td>
<td>.08</td>
<td>-0.08, 0.01</td>
</tr>
<tr>
<td>Summer Enrollment</td>
<td>-0.03</td>
<td>0.04</td>
<td>.43</td>
<td>-0.11, 0.04</td>
</tr>
</tbody>
</table>

*Note.* DV: dependent variable; SE: standard error; $p$: $p$-value; CI: confidence interval. A square root transformation was applied to the dependent variable.

Multiple regression results indicated that the Hypothesis 1.3 was not supported; randomization assignment did not significantly predict the number of days in which participants drank past NIAAA guidelines in the past month (DP), while controlling for the total number of drinks consumed in a single day in the past month at baseline, baseline AUDIT, athletic status, summer enrollment, and importance in changing drinking (Table 5). Participants who completed additional assessments at baseline before the WBI were not significantly different in the expected log count of days they
drank past NIAAA guidelines compared to those who were not assigned to complete baseline assessments before the WBI. The difference in expected log counts of the number of times participants drank past NIAAA guidelines in the past month was expected to be .001 units higher for participants in the experimental condition compared to the control condition, while holding the other variables constant in the model. Confounding variables that did significantly predict DP were baseline AUDIT, baseline highest number of drinks consumed in the past month, and summer enrollment.

Table 5

Main effect of randomization assignment on the number of days participants drank past NIAAA guidelines (DP).

<table>
<thead>
<tr>
<th></th>
<th>Difference in log count</th>
<th>SE</th>
<th>p</th>
<th>95%CI (low, high)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.07</td>
<td>0.15</td>
<td>.61</td>
<td>-0.21, 0.35</td>
</tr>
<tr>
<td>Randomization group</td>
<td>0.001</td>
<td>0.13</td>
<td>.99</td>
<td>-0.26, 0.26</td>
</tr>
<tr>
<td>Baseline AUDIT</td>
<td>0.95</td>
<td>0.16</td>
<td>.001</td>
<td>0.64, 1.27</td>
</tr>
<tr>
<td>Baseline highest drinks</td>
<td>0.04</td>
<td>0.01</td>
<td>.001</td>
<td>0.03, 0.06</td>
</tr>
<tr>
<td>Importance rating</td>
<td>-0.003</td>
<td>0.03</td>
<td>.91</td>
<td>-0.06, 0.06</td>
</tr>
<tr>
<td>Athlete</td>
<td>-0.27</td>
<td>0.18</td>
<td>.14</td>
<td>-0.63, 0.09</td>
</tr>
<tr>
<td>Summer Enrollment</td>
<td>-0.80</td>
<td>0.34</td>
<td>.02</td>
<td>-1.47, -0.13</td>
</tr>
</tbody>
</table>

*Note.* DV: dependent variable; SE: standard error; p: p-value; CI: confidence interval.

Negative binomial regression coefficients are presented as the difference in log count.
3.5.2 Results for Specific Aim 2

The results of the multiple regression analysis using multiply imputed datasets indicated that Hypothesis 2 was not supported; randomization group did not significantly predict the number of times participants experienced alcohol-related problems in the past month (RAPI score), while controlling for the total number of drinks consumed in a single day in the past month at baseline, baseline AUDIT, and spring 2013 enrollment (Table 6). The difference in expected log counts of the number of alcohol related problems in the past month was expected to be .02 units lower for participants in the experimental condition compared to the control condition, while holding the other variables constant in the model. Baseline AUDIT and spring 2013 enrollment did significantly predict RAPI score at follow up.
### Table 6

Main effect of randomization assignment on the number of alcohol-related problems in the past month at follow up (RAPI score).

<table>
<thead>
<tr>
<th>Difference in log count</th>
<th>SE</th>
<th>p</th>
<th>95%CI (low, high)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1.42</td>
<td>0.16</td>
<td>0.01</td>
</tr>
<tr>
<td>Randomization group</td>
<td>-0.02</td>
<td>0.13</td>
<td>.88</td>
</tr>
<tr>
<td>Baseline AUDIT</td>
<td>0.41</td>
<td>0.14</td>
<td>.003</td>
</tr>
<tr>
<td>Baseline highest drinks</td>
<td>0.01</td>
<td>0.01</td>
<td>.32</td>
</tr>
<tr>
<td>Spring 2013 Enrollment</td>
<td>-0.33</td>
<td>0.13</td>
<td>.01</td>
</tr>
</tbody>
</table>

**DV:** Alcohol-related problems (RAPI score)

**Note.** DV: dependent variable; SE: standard error; p: p-value; CI: confidence interval.

Negative binomial regression coefficients are presented as the difference in log count.

### 3.5.3 Results for Specific Aim 3

Results from the multiple regression analysis using multiply imputed datasets revealed that randomization group did not significantly predict the use of protective behaviors in the past month (PBSS score) while controlling for the total number of drinks consumed in a single day in the past month as reported at baseline, baseline AUDIT, attending college in the U.S./Canada, importance in changing drinking, confidence in changing drinking, and spring 2014 enrollment (Table 7). Hypothesis 3 was not supported. Participants who completed additional assessments before the WBI were not significantly different in their use of protective behaviors in the past month, compared to those who were not assigned to complete baseline assessments before the
Participants in the experimental condition had a mean of .23 points higher scores on the PBSS compared to the mean PBSS score for participants in the control condition, while holding other variables constant in the model. Variables that did significantly predict PBSS score were baseline AUDIT and country in which participants were attending college (Table 7). Importance rating marginally predicted PBSS score.

Table 7
Main effect of randomization assignment on the use of protective behaviors related to alcohol consumption in the past month at follow up (PBSS score).

<table>
<thead>
<tr>
<th>DV: Use of protective behaviors (PBSS score)</th>
<th>b</th>
<th>SE</th>
<th>p</th>
<th>95%CI (low, high)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>45.37</td>
<td>2.78</td>
<td>.001</td>
<td>39.91, 50.82</td>
</tr>
<tr>
<td>Randomization group</td>
<td>0.23</td>
<td>1.25</td>
<td>.85</td>
<td>-2.21, 2.68</td>
</tr>
<tr>
<td>Baseline AUDIT</td>
<td>-4.75</td>
<td>1.28</td>
<td>.002</td>
<td>-7.25, -2.24</td>
</tr>
<tr>
<td>Baseline highest drinks</td>
<td>-0.15</td>
<td>0.10</td>
<td>.14</td>
<td>-0.36, 0.05</td>
</tr>
<tr>
<td>Importance rating</td>
<td>0.56</td>
<td>0.30</td>
<td>.07</td>
<td>-0.04, 1.15</td>
</tr>
<tr>
<td>Confidence rating</td>
<td>0.43</td>
<td>0.29</td>
<td>.14</td>
<td>-0.14, 1.01</td>
</tr>
<tr>
<td>Spring 2014 Enrollment</td>
<td>0.83</td>
<td>1.36</td>
<td>.54</td>
<td>-1.84, 3.49</td>
</tr>
<tr>
<td>U.S./Canadian student</td>
<td>-4.44</td>
<td>1.59</td>
<td>.01</td>
<td>-7.56, -1.32</td>
</tr>
</tbody>
</table>

Note. DV: dependent variable; b: unstandardized regression coefficient; SE: standard error; p: p-value; CI: confidence interval.
3.6 Moderation Analyses: Results for Specific Aim 4

A priori, four variables were identified as factors that might modify the relationship between randomization assignment and alcohol use/alcohol consequences/protective behaviors: baseline AUDIT, gender, importance in changing drinking, and confidence in changing drinking (Specific Aim 4). Although the main analyses found no significant differences in outcomes from the RA+WBI and WBI only group, it was plausible that assessment reactivity may have been present at some levels of the aforementioned variables, but not for others (for example, for females but not males). For this reason, the moderating effects of each variable were examined.

Baseline AUDIT

Baseline AUDIT did not significantly moderate the relationship between randomization group and TOT (difference in log count= -0.12, CI= -0.51, 0.28, p = .57), PBAC (b = -0.001, CI = -0.059, 0.057, p = .97), DP (difference in log count = -0.20, CI = -0.80, 0.39, p = .51), RAPI score (difference in log count = -0.44, CI = -0.95, 0.07, p = .10), or PBSS score (b= 2.45, CI = -2.15, 7.06, p = .30) after controlling for potential confounding variables. The moderating effects on TOT are displayed in Table 8 as an example of the analyses.
Table 8

Moderating effect of baseline AUDIT on the total number of drinks consumed in the past month at follow up (TOT).

<table>
<thead>
<tr>
<th>Difference in log count</th>
<th>SE</th>
<th>p</th>
<th>95%CI (low, high)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>2.96</td>
<td>0.10</td>
<td>0.001</td>
</tr>
<tr>
<td>Randomization group</td>
<td>0.10</td>
<td>0.16</td>
<td>0.54</td>
</tr>
<tr>
<td>Baseline AUDIT</td>
<td>0.79</td>
<td>0.13</td>
<td>0.001</td>
</tr>
<tr>
<td>Randomization*AUDIT</td>
<td>-0.12</td>
<td>0.21</td>
<td>0.57</td>
</tr>
<tr>
<td>Importance rating</td>
<td>-0.02</td>
<td>0.02</td>
<td>0.41</td>
</tr>
<tr>
<td>Athlete</td>
<td>-0.27</td>
<td>0.14</td>
<td>0.051</td>
</tr>
<tr>
<td>Summer Enrollment</td>
<td>-0.53</td>
<td>0.23</td>
<td>0.02</td>
</tr>
<tr>
<td>Baseline highest drinks</td>
<td>0.05</td>
<td>0.01</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Note. DV: dependent variable; SE: standard error; p: p-value; CI: confidence interval.

Negative binomial regression coefficients are presented as the difference in log count.

Gender

Gender did not significantly moderate the relationship between randomization group and TOT (difference in log count = -0.27, CI = -0.64, 0.11, p = .16), PBAC (b = -0.02, CI = -0.08, 0.04, p = .60), DP (difference in log count = -0.22, CI = -0.72, 0.28, p
= .40), RAPI score (difference in log count = -0.07, CI = -0.59, 0.46, p = .80), or PBSS score (b= -4.08, CI = -8.87, 0.72, p = .10) after controlling for potential confounding variables. The moderating effects on TOT are displayed in Table 9 as an example of the analyses.

Table 9

Moderating effect of gender on the total number of drinks consumed in the past month at follow up (TOT).

<table>
<thead>
<tr>
<th>Difference in log count</th>
<th>SE</th>
<th>p</th>
<th>95%CI (low, high)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>2.93</td>
<td>0.10</td>
<td>.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.73, 3.13</td>
</tr>
<tr>
<td>Randomization group</td>
<td>0.13</td>
<td>0.13</td>
<td>.30</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-0.12, 0.38</td>
</tr>
<tr>
<td>Gender</td>
<td>0.45</td>
<td>0.12</td>
<td>.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.23, 0.68</td>
</tr>
<tr>
<td>Randomization*Gender</td>
<td>-0.27</td>
<td>0.19</td>
<td>.16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-0.64, 0.11</td>
</tr>
<tr>
<td>Baseline AUDIT</td>
<td>0.76</td>
<td>0.11</td>
<td>.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.54, 0.97</td>
</tr>
<tr>
<td>Importance rating</td>
<td>-0.02</td>
<td>0.02</td>
<td>.28</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-0.07, 0.02</td>
</tr>
<tr>
<td>Athlete</td>
<td>-0.28</td>
<td>0.13</td>
<td>.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-0.54, -0.02</td>
</tr>
<tr>
<td>Summer Enrollment</td>
<td>-0.45</td>
<td>0.24</td>
<td>.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-0.91, 0.02</td>
</tr>
<tr>
<td>Baseline highest drinks</td>
<td>0.04</td>
<td>0.01</td>
<td>.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.03, 0.06</td>
</tr>
</tbody>
</table>

Note. DV: dependent variable; SE: standard error; p: p-value; CI: confidence interval.

Negative binomial regression coefficients are presented as the difference in log count.
Importance

Importance rating did not significantly moderate the relationship between randomization group and TOT (difference in log count = 0.04, CI= -0.10, 0.02, p=.37), PBAC (b = 0.003, CI = -0.010, 0.017, p = .63), DP (difference in log count = 0.02, CI = -0.10, 0.14, p = .73), RAPI score (difference in log count = -0.06, CI = -0.17, 0.06, p = .34), or PBSS score (b= 0.20, CI= -0.89, 1.29, p = .71) after controlling for potential confounding variables. The moderating effects on TOT are displayed in Table 10 as an example of the analyses.
Table 10

Moderating effect of importance rating on the total number of drinks consumed in the past month at follow up (TOT).

<table>
<thead>
<tr>
<th>Difference in log count</th>
<th>SE</th>
<th>p</th>
<th>95%CI (low, high)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>3.02</td>
<td>0.10</td>
<td>.001</td>
</tr>
<tr>
<td>Randomization group</td>
<td>-0.03</td>
<td>0.12</td>
<td>.80</td>
</tr>
<tr>
<td>Importance rating</td>
<td>-0.04</td>
<td>0.03</td>
<td>.23</td>
</tr>
<tr>
<td>Randomization*Importance</td>
<td>0.04</td>
<td>0.04</td>
<td>.37</td>
</tr>
<tr>
<td>Baseline AUDIT</td>
<td>0.74</td>
<td>0.11</td>
<td>.001</td>
</tr>
<tr>
<td>Athlete</td>
<td>-0.26</td>
<td>0.13</td>
<td>.05</td>
</tr>
<tr>
<td>Summer Enrollment</td>
<td>-0.52</td>
<td>0.23</td>
<td>.02</td>
</tr>
<tr>
<td>Baseline highest drinks</td>
<td>0.05</td>
<td>0.01</td>
<td>.001</td>
</tr>
</tbody>
</table>

Note. DV: dependent variable; SE: standard error; p: p-value; CI: confidence interval.

Confidence

Confidence rating did not significantly moderate the relationship between randomization group and TOT (difference in log count= -0.28, p=.16), PBAC (b = -0.002, CI= -0.013, 0.017, p = .85), DP (difference in log count = -0.01, CI = -0.13, 0.11, p = .86), RAPI score (difference in log count = -0.03, CI = -0.16, 0.11, p = .68), and PBSS score (b= -0.10, CI= -1.30, 1.09, p = .86) after controlling for potential...
confounding variables. The moderating effects on TOT are displayed in Table 11 as an example of the analyses.

Table 11

Moderating effect of confidence rating on the total number of drinks consumed in the past month at follow up (TOT).

<table>
<thead>
<tr>
<th></th>
<th>Difference in log count</th>
<th>SE</th>
<th>p</th>
<th>95%CI (low, high)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>2.79</td>
<td>0.27</td>
<td>.001</td>
<td>2.25, 3.33</td>
</tr>
<tr>
<td>Randomization group</td>
<td>0.29</td>
<td>0.37</td>
<td>.43</td>
<td>-0.43, 1.01</td>
</tr>
<tr>
<td>Confidence rating</td>
<td>0.03</td>
<td>0.03</td>
<td>.41</td>
<td>-0.04, 0.09</td>
</tr>
<tr>
<td>Randomization*Confidence</td>
<td>-0.03</td>
<td>0.04</td>
<td>.37</td>
<td>-0.12, 0.06</td>
</tr>
<tr>
<td>Baseline AUDIT</td>
<td>0.75</td>
<td>0.11</td>
<td>.001</td>
<td>0.53, 0.96</td>
</tr>
<tr>
<td>Importance rating</td>
<td>-0.01</td>
<td>0.02</td>
<td>.45</td>
<td>-0.06, 0.27</td>
</tr>
<tr>
<td>Athlete</td>
<td>-0.26</td>
<td>0.14</td>
<td>.06</td>
<td>-0.53, 0.01</td>
</tr>
<tr>
<td>Summer Enrollment</td>
<td>-0.52</td>
<td>0.23</td>
<td>.02</td>
<td>-0.97, -0.07</td>
</tr>
<tr>
<td>Baseline highest drinks</td>
<td>0.05</td>
<td>0.01</td>
<td>.59</td>
<td>0.04, 0.06</td>
</tr>
</tbody>
</table>

*Note. DV: dependent variable; SE: standard error; p: p-value; CI: confidence interval.

Negative binomial regression coefficients are presented as the difference in log count.
3.7 Validation of WBI Efficacy: Change in Peak Estimated Blood Alcohol Content from Baseline to Follow Up

To confirm that the magnitude of the WBI effect on alcohol use at follow up was consistent with the literature, the change in PBAC from baseline to follow up was evaluated for all participants who completed the WBI. Results from the repeated measures ANCOVA analysis with baseline AUDIT, randomization group, and importance rating included as confounding variables demonstrated that there was a significant decrease in PBAC from baseline to follow up, with a moderate Cohen’s d effect size of .41 (F = 28.34, partial η² = .10, p=.001). The effect size was consistent with small to moderate effect sizes that have been found previously in the literature. Randomization group was not significantly associated with PBAC (F = 0.17, partial η² = .001, p=.68). Thus, there were significant WBI effects but not assessment reactivity effects.
CHAPTER 4: DISCUSSION

4.1 Main Findings: Specific Aims 1-3

The purpose of the current study was to test whether baseline research assessments administered online would have an additive effect on college student alcohol use and related problems when combined with a web-based brief alcohol intervention. Mechanisms of assessment reactivity were examined to determine under what conditions reactivity was most strongly observed. In contrast to the hypotheses, participants who completed the RA plus WBI were not significantly different than participants randomized to the WBI only group at follow up when compared on the total number of drinks consumed (Hypothesis 1.1), peak estimated blood alcohol content (Hypothesis 1.2), the number of days in which participants drank past NIAAA guidelines (Hypothesis 1.3), alcohol-related problems (Hypothesis 2), or protective behaviors related to alcohol consumption used (Hypothesis 3) in the past month. A repeated measures ANCOVA confirmed that there were significant effects of the WBI on peak estimated blood alcohol content as measured from baseline to follow up, however there were no significant effects of randomization assignment on peak blood alcohol content in the model.

The findings are in contrast to the majority of experimental studies of assessment reactivity in college student samples that found evidence of reactivity in the form of lower alcohol consumption, alcohol-related problems, and use of protective behaviors at follow up compared to control conditions (Hester et al., 2012; Kypri et al., 2007; McCambridge & Day, 2008; McCambridge & Kypri, 2011; Walters, Vader, Harris, & Jouriles, 2009). Multiple factors likely contributed to the lack of reactivity
effects. First, the current study was designed to test the effects of RA when administered in combination with a WBI. However, many of the assessment reactivity studies with college students tested the effects of RA alone (not in combination with an active treatment) and were compared to a control condition that received neither assessment nor treatment. Small to moderate assessment reactivity effects have been reliably found using this study paradigm (e.g., McCambridge & Day, 2008; Walters, Vader, Harris, & Jouriles, 2009). It may be that assessment reactivity is more prominent in the absence of treatment; that is, when RA essentially serve as their own intervention. This interpretation is also consistent with the reported observation of assessment reactivity effects during follow up periods in intensive alcohol treatment studies, as assessments conducted at these points were typically the only therapeutic contact participants encountered at that time.

The findings from the current study can serve as evidence validating WBI efficacy when alcohol use outcomes are measured within subjects using an RA+WBI design. The findings suggest that brief pretreatment assessment for research purposes does not bias within subjects estimates of WBI efficacy. However, it should be noted that the results validate the within groups measure to a single follow up assessment only, and should not be considered evidence for the use of multiple follow up assessments. Multiple assessment points can produce assessment reactivity effects (Clifford & Davis, 2012; Clifford et al., 2007).

Many studies in the literature determined WBI efficacy in two ways: 1) by evaluating within subjects change in alcohol consumption pre to post-treatment and 2) measuring alcohol consumption between an RA+WBI group and an RA only control
group. The findings do not validate between groups estimates of efficacy when a RA+WBI group is compared to an RA only control group. Some researchers have questioned whether assessment reactivity is concern if both treatment and control groups receive baseline RA (Magill et al., 2012). In this line of reasoning, the assumption is that if assessment reactivity is present, it is present to the same degree in both groups. Solomon (1949) and Campbell (1957) described this assumption as problematic, because assessments may interact with intervention effects and lead to a different degree of assessment reactivity than may be observable in an RA only control group. In the current study, assessments administered in combination with a WBI did not produce significant assessment reactivity effects; however previous research has suggested that assessments conducted in the absence of an intervention have produced assessment reactivity. A differential impact on treatment and control conditions could dampen the detection of accurate treatment effects when a comparison is made across groups. Researchers may actually find lower estimates of effect size than are truly present when comparing alcohol use between groups, however this conjecture should be experimentally tested.

Another factor that may have contributed to the lack of assessment reactivity effects may have been the web-based administration of the assessments. One study in the literature used the same experimental design as the current study except that the assessments and BI were conducted live, in person. This study randomized participants to complete a live TLFB + live BI or a BI only (Carey et al., 2006). At follow up participants in the TLFB+BI condition had significantly lower drinking and related problems compared to the BI only condition (Carey et al., 2006). It may be that an in-
person assessment has a greater impact on participants than web-administered assessments. For example, an in-person discussion of alcohol use and related problems could have more of an influence on motivation to change drinking and may lead to assessment reactivity effects at follow up. Future research on in-person and web-based administration of assessments could further delineate under what conditions assessment reactivity may arise.

In this regard, some researchers have found evidence of reactivity to live, phone-administered assessments versus paper and pencil assessments. Heather, Whitton, and Robertson (1986) conducted a study of a mailed self-help manual for heavy drinkers with follow up assessments administered either live via telephone or through mailed paper or pencil assessments. The researchers reported a greater reduction in alcohol consumption and alcohol-related problems in a group that had live, phone-based follow up assessments compared to a group that received the mailed assessment. The researchers questioned whether there may have been reactivity effects to the live interviews. In considering the current findings and the previous literature, there may be differential reactivity effects that are generated from in-person versus web-based research assessments. It may be that interpersonal interaction provided by a live assessment may have a greater impact on motivation to change alcohol use and lead to reactivity effects, whereas web-based assessment may have less of an impact. This conjecture should be further investigated with experimental tests and the results could indicate whether web administration is a valuable way of minimizing reactivity effects that may be more prominent from in-person assessment administration.
Another factor that may influence the degree of assessment reactivity observed is the number of follow up assessments conducted during a study. In the current study, there was one baseline assessment point and one follow up assessment point. Researchers have found evidence that the greater the number of assessment points at follow up, the greater reactivity is observed (Clifford et al., 2007). It may be that the single RA was not enough to influence assessment reactivity effects. In this regard, Kypri et al. (2007) found evidence of assessment reactivity on alcohol consumption at a 12 month follow up, but not at the study’s initial 6 month follow up assessment. The authors remarked that it may have taken multiple assessment points to observe reactivity effects in their study. Although evidence suggests that a greater number of assessments produces greater reactivity, experimental tests have demonstrated assessment reactivity to single, brief questionnaires. There may not have been assessment reactivity effects in the current study because effects of RA in combination with a WBI were not substantial in comparison to only a WBI and because there was one follow up assessment point.

Finally, the length of the RA may have contributed to the lack of reactivity effects. The RA used in the current study were designed in accordance with previous WBI efficacy trials in the literature. The RA took participants about 10 minutes to complete, which was about 50% of the time it took to complete the intervention. It is possible that the brevity of the RA allowed for the avoidance of reactivity effects. However, it is possible that longer assessments may garner reactivity. Clifford et al (2007) evaluated the influence of the length of follow up assessments on alcohol consumption in participants who completed intensive alcohol treatment. The
researchers found that participants randomized to the lengthy assessment group had greater reactivity to the assessments, as demonstrated by significantly greater proportion of days abstinent, drinks per drinking day, and proportion of heavy drinking days than the other groups, compared to the brief assessment group. It is unclear to what degree these findings would generalize to a sample of college students completing baseline RA online. No studies in the literature have directly measured reactivity to brief versus lengthy baseline RA. An experimental test using a college student sample could indicate the optimal length of RA that researchers could use to collect baseline data online while avoiding reactivity effects.

4.2 Moderating Analyses: Specific Aim 4

Although there were no significant main effects, the moderating effects of gender, baseline AUDIT indication of hazardous alcohol use, importance in changing drinking, and confidence in changing drinking were each evaluated because it was plausible that assessment reactivity could be present at some levels of these variables, but not at others (Specific Aim 4). Hypothesis 4 was not supported; none of the four variables significantly moderated the relationship between randomization assignment and any of the five outcomes examined. The results are in contrast to several studies in the literature. Magill et al. (2012) reported a statistically significant difference in alcohol consumption for participants with high precollege AUDIT scores compared to their low scoring peers; however the effect size observed was very small. Two studies found evidence of greater assessment reactivity for females; females decreased their consumption significantly more than males (Magill et al., 2012; Murray et al., 1988).
Murray et al (1988) observed substantial differences in reported smoking behavior between participants assessed at 5 time points compared to their unassessed peers, however the differences between males and females in the Magill et al (2012) study were very small. The strong gender effects in the Murray et al (1988) study were found using a sample of adolescent smokers. It is possible that younger participants, particularly young girls, may react more strongly to assessments by thinking about their behavior than adult participants, which could lead to greater reactivity. The Magill et al (2012) sample was comprised of college students who had experienced a recent adverse alcohol event, in contrast to the current study in which students with and without recent negative alcohol consequences were included. It may be that the motivational effects of both a recent negative consequence and the assessments were enough to produce reactivity in the Magill et al (2012) study, particularly in students with high baseline AUDIT scores and females. Also, Magill et al (2012) conducted RA in person, which may have garnered more reactivity than the web-based RA used in the current study.

Researchers have hypothesized that assessment reactivity occurs because assessment increases participant motivation to change drinking behavior (Clifford & Davis, 2012; Schrimsher & Filtz, 2011). Results of the current study indicated that ratings of importance and confidence in changing drinking did not moderate any of the outcomes measured. It may be that importance and confidence in changing drinking do not have substantial effects on the relationship between RA and alcohol outcomes examined in the study however it may be the case that other motivation factors do contribute to reactivity effects, such as readiness to change. It may also be the case that single item questions of importance and confidence are not adequate proxy measures
for motivation. As mentioned previously, the WBI came preprogrammed to ask for ratings of importance and confidence at the end of the WBI and thus these ratings were used opportunistically in the analyses. Given the hypothesis that exists in the field about a mechanism of motivation in assessment reactivity, further investigation into motivational factors related to assessment reactivity is warranted.

In addition to assessment reactivity effects, pre-post measures of peak blood alcohol content in the past month were derived for all participants to confirm that the WBI effects in the current study were similar to those observed in the literature. The repeated measures ANCOVA did indicate that there was a moderate effect size for decreased peak blood alcohol consumption from baseline to follow up, consistent with the small to moderate effect sizes observed in the literature (Khadjesari et al., 2011; Riper et al., 2009; White et al., 2010), but there were no significant effects when randomization assignment was included as a covariate in the model. These findings support the conclusion that there were significant WBI effects on alcohol consumption as measured by PBAC, but there were not significant assessment reactivity effects.

4.3 Future Studies

Future research could expand upon the current study design to further elucidate under what conditions assessment reactivity may be observed. Solomon (1949) proposed that researchers use a four condition experimental design to test assessment reactivity effects alone and in combination with a treatment. In what has been termed the Solomon four group design, study participants are randomized to an assessment + treatment group, treatment only group, assessment only group, or a no assessment no
intervention group. While there were not sufficient resources to conduct a Solomon four group designed experiment for the current study, this type of study using a college student sample would allow for an evaluation of assessment effects in combination and in isolation of a WBI within the same sample. Such a design could also facilitate an investigation into the mechanisms of assessment reactivity.

Further research would also be useful to determine whether there are differential reactivity effects between live, in-person administration of RA and web-based RA in one sample. If the results of such a study suggested there were no additive effects of RA to an intervention when conducted online, but there were additive effects when conducted in-person, researchers may prefer to use the web-administration to avoid reactivity effects and still be able to gather pretreatment alcohol use data on participants. Alternatively, researchers may choose to capitalize on the reactivity effects of in-person assessment for clinical purposes.

Additional investigation into the effects of the length of assessments on reactivity is needed. While the current study showed that brief assessments did not bias WBI efficacy estimates, it would be useful for researchers to know the length of assessments they could use to optimize pretreatment data collection while avoiding reactivity effects. It would also be useful to know whether very brief assessments could be conducted with an RA only control group that would not lead to substantial reactivity effects. Results from these studies could further inform researchers about how research methodology may or may not produce assessment reactivity.

Further research into the factors and mechanisms that influence assessment reactivity is also needed. Given the hypothesis in the field that assessment may increase
motivation to change behavior, a more detailed inquiry into motivational factors that relate to assessment reactivity could provide insight into mechanisms of action. The current study used a single item rating of importance and confidence; validated questionnaires that measure readiness to change or other aspects of motivation may allow for a more thorough examination of the relationship between assessment, motivation, and alcohol use.

4.4 Limitations

There were several limitations of the current study that warrant discussion. First, the study used a convenience sample and thus may not be representative of the college student population in the U.S. and Canada. The sample was skewed demographically towards females (70%) who were not racial or ethnic minorities (82%). While some WBI efficacy studies have observed similar proportions of females and Caucasian students (e.g., Butler & Correia, 2009), other studies have obtained more balanced samples (e.g., Doumas et al., 2010; Hustad et al., 2010; Walters et al., 2007). The skew toward Caucasian females may be related to the recruitment techniques used in the study. Recruitment was conducted primarily through social media and the majority of the sample was recruited through Facebook. Even at UVM where both social media and traditional recruitment techniques were utilized, most of the participants were recruited through Facebook (72%). Recent research has determined that the predominant users of Facebook are Caucasian females (Wells & Link, 2014). It may be that the characteristics of those recruited for the current study are representative of college students who are more active Facebook users. However, almost all college students (93-99%) report participation in social media site Facebook (Roblyer, McDaniel,
Webb, Herman, & Witty, 2010; Sheldon, 2008; Sponcil & Gitimu, 2013). Also, there were no significant differences in demographic characteristics and baseline alcohol use in the UVM students recruited through social media and traditional methods. Thus, more research is required to determine whether the skew in the sample was a result of the social media recruitment or was random and sample-specific.

A second limitation to the current study is that a substantial proportion of the participants randomized in the study did not complete the WBI (46%) and were essentially lost after randomization. The participants who were lost were heavier drinkers at baseline compared to those who did complete the WBI. Loss of heavy drinkers has been a significant problem in alcohol research to date, and the current study was no exception. In addition, participants were more likely to complete the WBI if they were randomized to the WBI only condition as opposed to the RA+WBI condition. It is likely that the differential drop out was due to the length of time required for participation and the values of the incentives provided. Participation for the WBI only group took about 15-20 minutes, whereas participation for the RA+WBI group took about 25-30 minutes. Incentives for completing the WBI were entry into up to three iPad mini drawings. It appears that the majority of participants were willing to spend 15-20 minutes of their time for entry into iPad mini drawings, but only about half of study participants were willing to spend 30 minutes of their time for the same incentive. Studies that increase the incentive value for the drawings, or change the incentive structure to providing every person who completes the WBI a monetary incentive may observe higher WBI completion rates than were obtained for the current study.
In this regard, almost 70% of participants were willing to spend about 10 minutes on the follow up questionnaire for entry into one of four $500 Amazon.com gift card drawings. The completion rate was somewhat lower than expected and researchers who use higher drawing incentive values of $750 or $1,000 may have higher follow up completion rates. The follow up rate was also lower than has been obtained in other WBI efficacy studies, which generally have observed at least 75% follow up completion rates (e.g., 96%: Neighbors et al., 2009; 75%: Walters et al., 2007), however there have been several exceptions (62%: Doumas, Kane, et al., 2011; 38%: Ekman et al., 2011). Researchers in the studies with higher completion rates provided reasonably large monetary incentives to each participant for completing follow up assessments and this likely contributed to higher retention rates.

4.5 Strengths

The study had several strengths. First, features of the study were designed in accordance with WBI efficacy studies in the literature to aid in generalizability. The study used an assessment battery that has been used extensively in the WBI efficacy literature (e.g., Doumas & Andersen, 2009; Doumas et al., 2010; Ekman et al., 2011; Neighbors et al., 2009; Walters et al., 2007). Also in accordance with the literature, the assessment battery was designed to take about 50% as long as the actual intervention. The WBI that was utilized has been investigated in the literature frequently and is commercially available. The study was adequately powered to detect small effects and experimental manipulations were maintained.
The study used a large number of undergraduate students from the U.S. and Canada and participants randomized to the experimental and control conditions were not significantly different at baseline on any demographic or alcohol use variables. There was a substantial drop out rate after randomization; however the characteristics of participants who completed the WBI did not differ between groups. Thus, it is likely that the findings were not biased by baseline differences between these groups.

Although 27% of alcohol use data and 32% of RAPI and PBSS data were missing at follow up, data imputation was utilized to cope with the missing data. Also, twelve variables with theoretical and statistical significance were evaluated for potential inclusion in the analytic models. Variables that influenced the beta or p values of the main outcome and/or that significantly predicted the outcome were included in the analyses in order to account for other factors that might influence the main outcomes of interest. With the combination of the inclusion of confounding variables and use of data imputation, it is likely that there was not bias from these sources and that there were no significant assessment reactivity effects.

4.6 Implications

The current study contributes to the literature by identifying an experimental condition under which assessment reactivity may not be present and does not appear to cloud the detection of WBI efficacy. The results are positive for WBI researchers, who may be justified in conducting brief assessments online to collect information about participant alcohol use pretreatment without biasing within subjects estimates of WBI efficacy. The results also have positive implications for WBI dissemination. WBIs have
been widely distributed to universities nationwide and abroad to be used as general services for university students. Universities using these programs may likely observe similar effect sizes to those reported in the literature. However, WBI effectiveness studies are needed.

The study did not test reactivity effects when a WBI treatment group is compared with an assessment only control group and the results of the current study do not serve as evidence validating between groups estimates of efficacy. Further research is needed to test whether there are differential assessment reactivity effects between treatment and assessment only controls. The greater the understanding we as a research field gain of the conditions under which assessment reactivity is and is not produced, the more we will be equipped to conduct methodologically sounds studies in a manner that limits or avoids the influence of research design methodology on study findings.

The study was funded by the Vermont Conferences on Primary Prevention (Tera Fazzino, PI).
References


doi:10.1136/bmj.325.7369.870

group drinking in clinical trials of brief alcohol intervention: Implications for bias
toward the null. *Drug and Alcohol Review, 29*(5), 498–507. doi:10.1111/j.1465-
3362.2010.00174.x

value of readiness, importance, and confidence in ability to change drinking and


Bingham, C. R., Barretto, A. I., Walton, M. A., Bryant, C. M., Shope, J. T., &
Raghunathan, T. E. (2010). Efficacy of a Web-Based, Tailored, Alcohol
Prevention/Intervention Program for College Students: Initial Findings. *Journal of

College Student Drinkers. *Journal of Consulting and Clinical Psychology, 68*(4),
728–733. doi:10.1037/0022-006X.68.4.728


Three Indexes of Risky Drinking for Predicting Alcohol Problems in College


doi:10.1080/07448489709595589


doi:10.1080/10413201003666454


efficacy and benefit-cost analysis. *Alcoholism, Clinical and Experimental Research, 26*(1), 36–43.


123


doi:10.1093/biomet/63.3.581


