View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Open Research Exeter

Manipulated vs. Measured: Using an Experimental Benchmark to Investigate the Performance of Self-Reported Media Exposure

Jennifer Jerit
Jason Barabas
William Pollock
Susan Banducci
Daniel Stevens
Martijn Schooenvelde

June 9, 2015

Running Head: Manipulated versus Measured Media Exposure

Abstract

Media exposure is one of the most important concepts in the social sciences, and yet scholars have struggled with how to operationalize it for decades. Some researchers have focused on the effects of variously worded self-report measures. Others advocate the use of aggregate and/or behavioral data that does not rely on a person's ability to accurately recall exposure. The present study introduces the prototype of an experimental design that can be used to improve measures of exposure. In particular, we show how an experimental benchmark can be employed to (1) compare actual (i.e., manipulated) and self-reported values of news exposure; (2) assess how closely the self-reported measures approximates the performance of "true" exposure in an empirical application, and (3) leverage the experimental benchmark to investigate whether a variation in question wording improves the accuracy of self-reported exposure measures.

In order to study many contemporary political phenomena, it is crucial to know what political content a person has been exposed to in the mass media. Yet for decades, scholars have struggled to find the best way to operationalize media exposure. Concerns about measurement error and validity even have led some researchers to advocate abandoning the use of self-reported media use measures altogether (e.g., Price & Zaller, 1993). However, important advances have been made in the measurement of media exposure (e.g., Althaus & Tewksbury, 2007; Dilliplane, Goldman, & Mutz, 2013; Guess, 2015; Prior, 2009a). And because behavioral measures of media use are costly and hard to implement, social scientists are bound to rely on self-reports, at least for the near future. There also is unique value in survey-based measures of media exposure in terms being able to study people as they ordinarily encounter political information (Barabas & Jerit, 2009). Instead of rejecting the use of self-reported exposure measures, we need to understand why people do not provide more accurate responses.

This study seeks to deepen our understanding of the challenges of measuring media exposure through the use of an experimental benchmark. Scholars have used this approach profitably in a variety of other contexts (e.g., LaLonde, 1986; Green, Leong, Kern, Gerber, & Larimer, 2009; Arceneaux, Gerber, & Green, 2010), typically to compare the inferences from observational and experimental analyses. The random assignment of treatment and the resulting degree of internal validity is what makes the causal effect from an experiment the

point of reference (i.e., the "benchmark") against which other methods are evaluated.[1] In the present study, we employ an experimental benchmark to investigate the performance of self-reported exposure items. We find that there is some imprecision in the self-reported exposure item, but the use of a randomized media treatment allows us to identify which respondents misreport and why. Our study also leverages the experiment to assess the performance of two differently worded exposure measures. In this way, we hope our findings will inform the next generation of scholarship on communication exposure.

The Existing Literature

An experimental benchmark is a useful, but rarely employed, method for investigating the problems of self-reported exposure measures. To date, scholars have either tried to "fix" the problems with self-reported exposure items (e.g., Bartels, 1993) or make theoretically-inspired modifications to the wording of existing self-report measures, using convergent and/or predictive validity to gauge the degree of improvement (e.g., Althaus & Tewksbury, 2007; Dilliplane, Goldman, & Mutz, 2013; Prior, 2009a).[2] There is, however, a fundamental

---

[1] Conceptually, benchmarking studies are related to a wave of recent research making cross-study comparisons (e.g., Barabas & Jerit, 2010; Clifford & Jerit, 2014; Jerit, Barabas, & Clifford, 2013; Krupnikov & Levine, 2014).

[2] Convergent validity refers to the fit between independent measures of the same construct, while predictive validity refers to the ability of a construct to predict some criterion variable (usually political knowledge in the case of media exposure; see Dilliplane, Goldman, & Mutz, 2013).

ambiguity with both approaches because the researcher remains ignorant of a person's actual level of media exposure.

To shed light on the errors associated with recall measures, it is crucial to establish a baseline against which self-reports can be compared. Vavreck (2007) does this by randomly assigning exposure to a mobilization message in a web-based survey at one point in time, and later asking people in the treatment and control groups to self-report their exposure to the images from the experimental stimuli. Thus for each respondent, there is a true state of the world (i.e., their treatment assignment) as well as their *stated* level of exposure. In the end, Vavreck finds that people who over report media exposure also overstate their political participation—and that both forms of bias occur among those who have a high sense of civic duty and who believe they can influence government decisions (i.e., high self-efficacy).[3]

Investigations into the flaws of self-reported exposure measures are valuable, but they beg the question of why such biases exist. According to psychologists and survey researchers, the central problem with measuring media exposure is the *ability* of people to accurately recall the relevant behavior (e.g., Chang & Krosnick, 2003). One reason for the difficulty is the manner in which autobiographical memories of mundane behaviors are stored in the brain. Individual episodes of such behaviors are not stored in memory. "Instead," write Schwarz and

[3] Neilson ratings represent another alternative to self-reported exposure measures (Prior, 2009b), but this is an aggregate-level measure (i.e., the ratings represent the total audience for a particular program). Additionally, the ratings may contain error because they are based on "people meters," which require viewers to push a button to indicate the beginning and ending of their viewing.

Oyserman, “the various instances of closely related behaviors blend into one global knowledge-like representation that lacks specific time or location markers” (2001, p. 137).

The recommendation for researchers seeking to recover accurate recall of a mundane behavior is to design survey questions in a way that facilitates memory. Markus Prior comes to a similar conclusion in a study that identifies ability, rather than motivation or socially desirable responding, as the cause of over reporting: “Most respondents are incapable of recalling most or all episodes of news exposure so they *estimate* their exposure,” (2009a, p. 904, emphasis added). This leads to two sources of error in self-reported items: the measure can miss people who were actually exposed but do not remember the treatment *and* it can include people who were never exposed but mistakenly report exposure (Vavreck 2007, p. 332; also see Schwarz, 1999).

Notwithstanding these challenges, simple changes in questionnaire design may ameliorate the problem. Prior, for example, recommends that surveys “offer help with the estimation of news exposure” (2009a, p. 904). Likewise, Schwarz and Oyserman state that researchers can “improve the likelihood of accurate recall by restricting the recall task to a short and recent reference period and by providing recall cues…[such as] what happened, where it happened, and who was involved” (2001, p. 138). In the present study, we apply this advice in a laboratory experiment in which media exposure was both manipulated and measured (see van Elsas, Lubbe, van der Meer, & van der Brug, 2014 or Timotijevic, Barnett, Shepherd, & Senior, 2009 for other applications).

More generally, we will demonstrate how an experimental benchmark can be used fruitfully to (1) compare actual (i.e., manipulated) and self-reported values of news exposure; (2) assess how closely the self-reported measures approximates the performance of “true”

exposure in an empirical application (e.g., as an independent variable predicting knowledge); and (3) leverage the experimental benchmark to investigate whether a subtle variation in question wording improves the accuracy of self-reported exposure measures. It is our hope that these investigations will advance what we know about measuring exposure to mass communication.

Data and Study Design

We conducted a randomized laboratory experiment in which we controlled exposure to a news story. Because our goal was to compare actual (i.e., manipulated) and self-reported values of news exposure, we intentionally sought to implement our design in a controlled (i.e., laboratory) setting. We acknowledge that ordinary people consume news in settings that differ considerably from the one subjects experienced in our study. There have long been concerns about the realism of experimental research (e.g., Shadish, Cook, & Campbell, 2002). What our design may lack with respect to ecological validity (Morton & Williams, 2010) it compensates for in terms of the precision with which it can identify subjects who misreport their news exposure.

Respondents

Participants (n=296) were undergraduate students enrolled in political science classes at a large public university in the northeast region of the United States during the summer and fall 2014. An initial study (n=128) was administered in May 2014. The study was replicated

with different subjects later that same month (n=25) and again in November 2014 (n=143).[4] Subjects were recruited to participate in exchange for extra credit and instructed to sign up for the study through an online appointment system. The participants included 158 males and 135 females (with three people declining to indicate gender). The sample leaned slightly Democratic (on a 7-point partisan identification variable the mean was 5.1 with a standard deviation of 1.8). Table A1 provides additional details regarding sample characteristics.[5]

Procedures

Our study is a 2 (Exposure: news story vs. no story) X 2 (Wording of self-reported exposure question: general vs. specific) between-subjects design that took place across two waves, approximately two days apart. Thus, participants were randomly assigned to receive a story or not, and then they were subsequently randomly assigned to receive one of the two question wordings of the media exposure measure.

The study occurred over two time points. At Time 1, participants completed a self-administered questionnaire on a computer in a private room. At the conclusion of the survey, subjects were informed about the opportunity to participate in a short follow-up study that would take place online at a time and place of their choosing (with participants entered into a

---

[4] In the analyses below, we combine data from all three studies together and include indicators for the first and second experimental replications where appropriate. The studies were approved by the Institutional Review Board at <<Redacted>> (Application #: 580472).

[5] There were slight imbalances across conditions on partisanship, political knowledge, risk aversion, and presidential approval (see Table A1). We confirmed that all of our results obtain in models including controls for these variables.

drawing for a cash prize). Those who agreed to complete the follow-up questionnaire were sent a link to the study.[6]

Figure 1 about here.

Figure 1 shows some of the key design features of our study. At Time 1, participants were randomized to receive a story about job growth in the United States or no story (see Appendix for a screen shot of the stimulus).[7] The second experimental factor corresponds to a variation in question wording at Time 2, when participants answered a self-reported exposure item pertaining to the topic of job growth. There was general version of the exposure item, which read, "Thinking about the past few days, do you recall seeing any information about job growth in the United States?" and had answer choices of "Yes" and "No." The specific version was identical with the exception of a preamble that was intended to increase accuracy: "Thinking about the survey you completed a few days ago on campus, do you recall seeing any information about job growth in the United States?" The wording was intended to aid recall (Schwarz & Oyserman, 2001) by restricting the recall task to a short and recent

---

[6] The re-contact rate across the three studies was as follows. In the original study, 66% of the respondents from Time 1 completed the follow-up. The re-contact rates for the first and second replications were 80% and 55%, respectively. Auxiliary analyses indicate that attrition was random with respect to treatment assignment (see Table A1).

[7] The treatment was an edited version (338 words) of an actual *USA Today* story that appeared on the newspaper's website approximately one year before our study. We selected this article because it was geared toward young job seekers, and thus the topic should have had some appeal for our subjects.

reference period and by providing a cue regarding the location of the relevant behavior (e.g., a survey on campus).

Measures

Most of the key outcome measures appear at Time 2, however, the original study and the first replication included a manipulation check immediately following the treatment at Time 1 (Mutz, 2011). The manipulation check consisted of three items asking respondents about details from the news story, and we will employ these questions to establish the effectiveness of the news treatment (i.e., that it had the intended effect on subjects).

As for the other outcomes, the Time 1 survey included measures of several individual difference variables that have been shown to be related to how people seek, acquire, and process information. There were six-items assessing Need for Cognition (NC) (Petty & Cacioppo, 1986; $\alpha = .78$), four items on news elaboration (Eveland, 2002; $\alpha = .79$), and a seven-item social desirability battery (Crowne & Marlow, 1960; $\alpha = .50$). Risk aversion was measured with a single question asking respondents to place themselves on a seven-point scale that ranged from “extremely comfortable taking risks” to “extremely uncomfortable raking risks” (with all response options labeled). Additionally, there was an item asking participants how often they pay attention to what is going on in government and politics, with five response options ranging from “All the time” to “Never.” All of the individual difference items were measured prior to the treatment at Time 1.

The self-reported exposure question occurred at Time 2. As noted above, there was a general and specific version of the question, each with answer choices of “Yes” and “No.” After the self-reported recall item, all respondents received a follow-up asking how certain they were about seeing something about job growth (see Miller & Peterson, 2004 for more on

meta-attitudinal measures). There were three response options ranging from "Absolutely certain" to "Not certain." Those who recalled seeing something about job growth were asked an additional series of questions that probed the accuracy of their memories. The item (shown as a grid with "Yes" or "No" response options) asked whether the information pertained to "rates of job creation in different regions of the United States," "a decline in the number of jobs in the restaurant industry," or "a new federal program to train people receiving unemployment benefits." Only the job creation item should have been answered affirmatively based upon the content of the news article treatment.

To investigate the possibility of acquiescence bias, we included a placebo item that asked respondents whether they recalled seeing anything about human cloning in the United States.[8] Finally, we repeated one of the knowledge questions from Time 1. The item asked, "According to economic forecasts, which region of the country has reported the fastest job growth in recent years? The answer choices were "East," "West" (correct), "South," and "Midwest."

A novel feature of our study is that it both manipulates *and* measures exposure. As a result (and much like Vavreck [2007]), we can assess the degree to which a self-reported exposure measure identifies people who were exposed to the randomly-assigned treatment story. We also examine the performance of the manipulated and measured exposure items as independent variables in a model predicting treatment-relevant knowledge at Time 2. This is a

---

[8] Respondents received a general or specific version of this question, depending on treatment assignment, along with the certainty follow-up. We confirmed via media content analysis that there had been no media coverage of human cloning around the time of the study.

useful exercise insofar as researchers often use self-reported media exposure variables in this capacity (e.g., Romantan, Hornik, Price, Capella, & Viswanath, 2008). Finally, we compare two versions of the self-reported exposure (i.e., recall) question. Based on work done by Schwarz and colleagues, we expected that the specific form of the recall question would guide memory search and help treated subjects remember elements of the treatment story.

Empirical Results

We begin the discussion of our results by demonstrating that exposure to the news story had the intended effect. As expected, treated subjects were significantly more likely to answer the three manipulation check questions correctly (2.5 vs. 1.2; $|t| = 9.2$; $p < .01$).[9] Even more impressive, the effect of the treatment persists over time, with treated respondents at Time 2 significantly more likely than control respondents to answer the knowledge question correctly (.51 vs. .27; $|t| = 3.4$; $p < .01$). Having shown that the treatment was effective, we next consider whether the Time 2 recall item recovered exposure among people who were in the treatment group at Time 1.

Does a Self-Reported Measure Identify who was Exposed?

There was a tendency for treated respondents to be more likely to answer "yes" to the self-reported exposure question at Time 2. This is reassuring, for one would expect there to be a relationship between what actually happened in the study (e.g., getting the job growth story or not) and subjects' recollections. However, the correspondence is not particularly strong. In a probit model predicting self-reported exposure (i.e., answering "yes" to the recall question),

[9] All statistical tests are two-tailed unless otherwise noted.

the experimental treatment has positive and marginally significant effect (coeff = .30; p =.13).[10] We observe a similar pattern with certainty of recall, with exposure to the treatment significantly increasing one's certainty of recall on the self-reported exposure item (coeff = .45; p =.08).[11] Based on the results so far, we have some confidence that the self-reported exposure measure is valid. In a design where we controlled who viewed the job growth story, the treatment has the expected (albeit weak) relationship to self-reported exposure and certainty of exposure.

At the same time, there is some imprecision in the self-reports. In our case, more than half (57%) of the respondents in control group report seeing something about job growth (i.e., they answer "Yes" to the self-reported exposure question at Time 2 even though they were not exposed to the story in the experiment). Additionally, a third of treated respondents say they do not recall hearing anything about job growth, even though we know they were treated (and that the treatment was effective).[12] Consistent with research on the difficulty of estimating

[10] The marginal effect on the treatment indicator is .11. This model includes indicators for the first and second replications, along with the pretreatment measure of respondent attention. Excluding the latter (only including controls for the replications) results in a p-value of .14 for the treatment indicator (marginal effect = .11).

[11] The sign and significance of the coefficient on the treatment indicator remains unchanged in a model that excludes the pretreatment measure of attention.

[12] Notably, there is no difference in the amount of time treated subjects spent viewing the story according to self-reported recall status (i.e., answering "yes" or "no" to the recall question; p = .95).

mundane behaviors (e.g., Schwarz & Oyserman, 2001), many of the respondents in our study had difficulty figuring out whether they had been exposed to information about job growth.

Comparing Measured and Manipulated Exposure as an Independent Variable

Researchers often seek to measure what information people have been exposed to with the goal of relating self-reported exposure to other outcomes, such as learning and opinion change. Indeed, Dilliplane, Goldman, and Mutz write that "the gold standard for assessing the validity of media exposure is how well a measure predicts political knowledge *gain*" (2013, p. 238, emphasis original). Thus, we investigate the performance of the self-reported exposure measure in predicting knowledge about job growth at Time 2. We expected that exposure to job growth information would be positively related to giving correct answer to the knowledge question at Time 2. Tables 1 and 2 provide the relevant results.

Table 1 about here.

Table 1 shows the results of a probit analysis in which the dependent variable is knowledge at Time 2 and the independent variable is the treatment indicator denoting whether the respondent received the job growth story at Time 1. Terms for the first and second replications are also included in the second column of results, but the conclusion is the same across both models. Being randomly assigned to view the story on job growth at Time 1 is positively and significantly ($p < .01$) related to knowledge about this topic at Time 2. The marginal effect of the treatment indicator is substantively large (m.e. = .23), translating into a roughly 23 percentage point increase in knowledge for respondents exposed to the news story at Time 1.

Table 2 about here.

A different conclusion emerges from Table 2, where we conduct a similar analysis but use *measured* exposure as the independent variable predicting Time 2 knowledge.[13] Because the analysis is now observational, we also included controls for gender and Need for Cognition, two individual-level characteristics that have been shown to be related to observed levels of knowledge (Delli Carpini & Keeter, 1996; Bizer, Krosnick, Petty, Rucker, & Wheeler, 2000). Irrespective of model specification, self-reported exposure is *negatively* related to knowledge at Time 2, and the variable is statistically significant in two of the three model specifications (at $p < .10$ or better with the marginal effect ranging from -.12 to -.16). The difference in results across Tables 1 and 2 is, on its own, quite striking. But consider the broader implications of the analyses thus far: A randomly assigned treatment *manipulating* news exposure increases the likelihood of saying "Yes" to self-reported recall item and knowledge related to the treatment. Yet the self-reported recall item—whose sole purpose is to *measure* exposure to the news treatment—is negatively related to knowledge.

To illustrate the source of this discrepancy, Figure 2 displays levels of knowledge across experimental groups at Time 2.

Figure 2 about here.

The two leftmost columns ("Manipulated Exposure") show the percentage correct for treated and control subjects, and here the pattern mirrors previously-described findings. Treated subjects had significantly higher levels of knowledge than control subjects (51% versus 27%

[13] We ignore the variation in question wording and collapse across the general and specific wording conditions (the effect of question wording is examined in the next subsection).

resulting in a 24-point treatment effect). This difference represents the effect of manipulating exposure to job growth (i.e., randomly assigning some people to view the news story).

The remaining columns make analogous treatment vs. control comparisons, but now respondents are grouped according to whether they gave a positive response to the self-reported exposure question (Recall =Yes, middle set of columns) or a negative response (Recall =No, rightmost set of columns). Among those who said they had been exposed to information about job growth and who were in the treatment group, 46% give the correct answer to the knowledge question. In contrast, among those who said they had been exposed to information about job growth but were in the control group, the corresponding figure is only 20% (resulting in a 26-point treatment effect). The average percent correct among respondents who said they were exposed was just 34%, due in large part to the very low level of knowledge among people who over report their exposure.

A different pattern obtains for respondents who indicated they did *not* recall seeing anything about job growth (the rightmost set of columns in Figure 2). People who said they did not recall anything but who were in fact treated had high levels of knowledge (60% provide the right answer). Among those who said they were not exposed and who were in the control group, approximately a third (36%) gave the correct answer (resulting in a 24-point treatment effect).The overall percent correct for respondents who said they were not exposed is 46%. Notice that this figure is higher than corresponding figure for people who positively report exposure (34%, noted above). Thus, Figure 2 illustrates that two types of errors—

respondents who falsely recall exposure and those who mistakenly forget it—are the source of the discrepancy across Tables 1 and 2.[14]

Insofar as there are errors in self-reported measures of news exposure, it behooves researchers to "[understand] the bias better" (Prior 2012, p. 362). In our case, the randomized experiment permits an analysis of the individual-level correlates of misreporting. We are particularly interested in those who falsely recall exposure (i.e., over reporting) and those who mistakenly forget it (i.e., under reporting). In separate models predicting over reporting and under reporting, the only individual-level factor that is statistically significant is self-reported attention ($p < .05$ for both; complete results shown in the Table A2). Respondents who report paying "attention to what's going on in government in politics" are significantly more likely to over report exposure (i.e., say they recalled information even though they were in control group), while those who answer the attention question negatively under report (i.e., say they do not recall it even though they were in treatment group). Consistent with the work of Prior (2009a), social desirability was not related to over reporting, nor was risk aversion, media elaboration, Need for Cognition, or gender. In light of a difficult estimation task, respondents seem to use their habitual attention to news about government and politics as cue when answering the self-reported recall question (see Prior 2009a for a similar finding).[15]

---

[14] See Table A3 for a tabular presentation of the data shown in Figure 2.

[15] In explaining why more politically interested people over report media exposure, Prior states, they "rely too heavily on their generally high political involvement when estimating network news exposure without help. Even when they recall only a few episodes of news

Investigating the Performance of Differently Worded Exposure Questions.

Having shown that the self-reported exposure measure contains some error, we now leverage the experiment to examine whether question format (general versus specific wording) improves respondent accuracy. Because we know the "truth" (i.e., through a person's treatment assignment), we can assess the performance of the general and specific wordings on the initial recall question, the follow-up item on certainty of recall, and the three-item battery that probed respondents for their memories about the treatment story.

Overall, question wording did not affect the accuracy with which people self-reported their exposure to news stories about job growth. Among treated respondents, receiving the specific language did not increase a person's likelihood of giving an affirmative response to the self-reported exposure question (the mean value of *Recall* is .67 in both groups). Similarly, among the untreated, there was no difference in how respondents answered the self-reported exposure question across the specific and general conditions (.60 vs. .55; $|t| = .60$; $p = .55$). Question form also had no significant effects on certainty of self-reported recall ($p > .50$ in analogous t –tests). Despite the addition of language that was designed to prompt memory of the "survey on campus," there were no differences in accuracy among respondents receiving the general or specific wording either for the initial recall question or the certainty follow-up.

It is possible, however, that the specific language had more subtle effects when it comes to improving recall. Toward that end, we examined responses to the three-item battery

---

exposure, they may infer frequent exposure from their considerable interest in (and knowledge of) politics" (2009a, p. 901).

probing what respondents could remember about the treatment story.[16] In this analysis, the dependent variable is scored so that higher values denote level of accuracy (i.e., a score of three denotes respondents who said "yes" to the item on job growth and "no" to the restaurant and training program items). We predict performance on the scale with an ordered logit model including terms for *Treatment*, *Specific Question Wording*, and their interaction (*Treatment* X *Specific*).

Table 3 about here.

The specific form of the self-recall question improves the ability of people to identify elements of treatment story, but only among people who were treated. The coefficient on *Treatment* X *Specific* is positively signed and statistically significant (coeff = 1.6; p= .07). Translated into a more intuitive quantity, there is a nearly one-point increase (.82) in the dependent variable for those who were treated with the news story and received the specific form of the recall question. In an analysis of the placebo topic (not shown), the specific form also increased respondent accuracy. Few people said they had seen information about human cloning, but the specific question wording reduced the incidence even further (.06 vs. .01; $|t|$= 1.71; p = .08). The significant effect of question wording in these two analyses is noteworthy, both because the three-item battery was a difficult series of questions to answer correctly, and

[16] As noted earlier, those who said they recalled seeing something about job growth were asked whether the information pertained to "rates of job creation in different regions of the United States," "a decline in the number of jobs in the restaurant industry," or "a new federal program to train people receiving unemployment benefits." This question represents a more challenging recall task than the previous two items (overall recall and recall certainty).

because a floor effect on the placebo question makes it that much more difficult for question wording to have a significant effect.

Discussion

This study employed an experimental benchmark to better understand the challenges of using self-reported media exposure measures. Based upon the analyses reported here, there is evidence that the self-reported exposure variable contains error, but the experiment reveals that it is as much a product of under reporting as it is over reporting (e.g., Figure 2). There also is suggestive evidence that misreporting occurs because respondents reason from their habitual news attention to answer a more specific question about news exposure. Finally, an attempt to facilitate recall with modified question wording was only partially successful: accuracy on the basic recall question did not improve, however, the specific wording did help treated subjects better distinguish details from the news story they read at Time 1. It is important to emphasize that the insights from our analyses are derived from having an experimental benchmark (i.e., knowing which individuals were treated with the news story). Going forward, we believe the analytical leverage provided by an experimental design can be a powerful tool for uncovering the biases that may lurk in self-reported measures of exposure.

Conclusion

Media exposure has been described as "one of the most central concepts in all social science" (Prior, 2013, p. 632). Given its important place in research on public opinion and political behavior, there is room for methodological pluralism when it comes to understanding how best to measure this concept. Social science has benefited enormously from efforts to improve the wording of media use measures (e.g., Althaus & Tewksbury, 2007; Chang &

Krosnick, 2003; Dilliplane, Goldman, & Mutz, 2013), even if there is not complete consensus in that regard (e.g., Prior, 2013).[17] New technologies offer the promise of a behavioral measure by tracking media usage (e.g., LaCour & Vavreck, 2014), though some of these approaches seem obtrusive (e.g., if people must agree to use a new technology) and therefore might suffer from non-compliance.

Our contribution has been to develop a prototype of an experimental design that can be employed in a laboratory or online setting to explore a range of question wording variations and styles of media treatments. In general, we believe that much can be learned by comparing the performance of different exposure questions in a setting in which a news treatment has been delivered exogenously to some participants. The promise of being able to analyze the individual-level characteristics associated with misreporting, combined with the insights from the field of survey design (e.g., Schwarz & Oyserman, 2001; Tourangeau, Rips, & Rasinski, 2000) seems especially promising for improving how we measure media exposure.

[17] Another challenge facing researchers who use survey-based measures of exposure is the steep decline in response rates (Kohut, Keeter, Doherty, Dimock, & Christain, 2012), and the potential for unobserved factors to be correlated with survey participation and self-reported media use (irrespective of how the exposure question is worded).

References

Althaus, S. L., & Tewksbury, D. H. (2007). Toward a new generation of media use measures for the ANES. Report to the Board of Overseers, American National Election Studies.

Arceneaux, K., Gerber, A. S., & Green, D. P. (2010). A cautionary note on the use of matching to estimate causal effects: An empirical example comparing matching estimates to an experimental benchmark. *Sociological Methods and Research, 39*, 256-82.

Barabas, J., & Jerit, J. (2010). Are survey experiments externally valid? *American Political Science Review*, *104*, 226-242.

Barabas, J., & Jerit, J. (2009). Estimating the causal effects of media coverage on policy-specific knowledge. *American Journal of Political Science*, *53*, 73-89.

Bartels, L. M. (1993). Messages received: The political impact of media exposure. *American Political Science Review*, *87*, 267-285.

Bizer, G.Y., Krosnick, J.A., Petty, R.E., Rucker, D.D., Wheeler, C. (2000). Need for Cognition and Need to Evaluate in the 1998 National Election Survey Pilot Study. Report to the Board of Overseers, American National Election Studies

Chang, L., & Krosnick, J. A. (2003). Measuring the frequency of regular behaviors: Comparing the "typical week" to the "past week." *Sociological Methodology*, *33*, 55-80.

Clifford, S., & Jerit, J. (2014). Is there a cost to convenience? An experimental comparison of data quality in laboratory and online studies. *Journal of Experimental Science*, *1*, 120-131.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, *24*, 349-354.

Dilliplane, S., Goldman, S. K., & Mutz, D. C. (2013). Televised exposure to politics: New measures for a fragmented media environment. *American Journal of Political Science*, *57*, 236-248.

Delli Carpini, M.X., & Keeter, S. (1996). *What Americans know about politics and why it matters*. New Haven, CT: Yale University Press.

Eveland, W. P., Jr. (2002). News information processing as mediator of the relationship between motivations and political knowledge. *Journalism and Mass Communication Quarterly, 79*, 26-40.

Green, D. P., Leong, T. Y., Kern, H. L., Gerber, A. S., & Larimer, C. W. (2009). Testing the accuracy of regression discontinuity analysis using experimental benchmarks. *Political Analysis, 1*, 400-17.

Guess, A. M. (2015). Measure for measure: An experimental test of online political media exposure. *Political Analysis*, *23*, 59-75.

Jerit, J., Barabas, J., & Clifford, S. (2013). Comparing contemporaneous laboratory and field experiments on media effects. *Public Opinion Quarterly*, *77*, 256-282.

Kohut, A., Keeter, S. Doherty, C., Dimock, M., & Christian, L. (2012). "Assessing the representativeness of public opinion surveys." Pew Charitable Trusts Report, May 15.
.

Krupnikov, Y., & Levine, A. S. (2014). Cross-sample comparisons and external validity. *Journal of Experimental Political Science*, *1*, 59-80.

LaCour, M. J., & Vavreck, L. (2014). Improving media measurement: Evidence from the field. *Political Communication*, *31*, 408-420.

LaLonde, R. J. (1986). Evaluating the econometric evaluations of job training with experimental Data. *American Economic Review, 76*, 604-20

Miller, J., & Peterson, D. (2004). Theoretical and empirical implications of attitude strength. *Journal of Politics, 66*, 847-867.

Morton, R. & Williams, K. (2010). *Experimental political science and the study of causality: From nature to the lab*. New York: Cambridge.

Mutz, D. C. (2011). *Population-based survey experiments*. Princeton, NJ: Princeton University Press.

Petty, R. E., & Cacioppo, J. T. (1986). *Communication and persuasion: Central and peripheral routes to attitude change.* New York: Springer-Verlag.

Price, V., & Zaller, J. (1993). Who gets the news? Alternative measures of news reception and their implications for research. *Public Opinion Quarterly*, *57*, 133-164.

Prior, M. (2009a). Improving media effects research through better measurement of news exposure. *Journal of Politics*, *71*, 893-908.

Prior, M. (2009b). The immensely inflated news audience: Assessing bias in self-reported news exposure. *Public Opinion Quarterly*, *73*, 130-143.

Prior, M. (2012). Who watches presidential debates? Measurement problems in campaign effects research. *Public Opinion Quarterly*, *76*, 350-363.

Prior, M. (2013). The challenge of measuring media exposure: Reply to Dilliplane, Goldman, and Mutz. *Political Communication*, *30*, 620-634.

Romantan, A., Hornik, R., Price, V., Cappella, J., & Viswanath, K. (2008). A comparative analysis of the performance of alternative measures of exposure. *Communication Methods and Measures*, *2*, 80-99.

Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist, 54*, 93-105.

Schwarz, N., & Oyserman, D. (2001). Asking questions about behavior: Cognition, communication, and questionnaire construction. *American Journal of Evaluation*, *22*, 127-160.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-Experimental designs for generalized causal inference*. Boston: Houghton Mifflin.

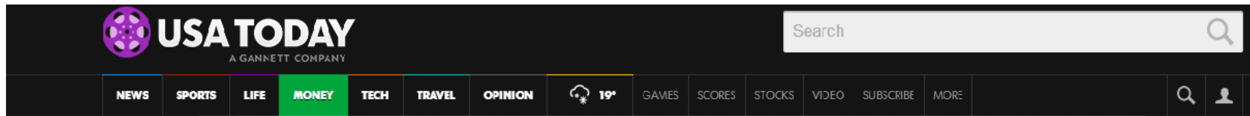
Timotijevic, L., Barnett, J., Shepherd, R., & Senior, V. (2009). Factors influencing self-report of mobile phone use: The role of response prompt, time reference and mobile phone use in recall. *Applied Cognitive Psychology*, *23*, 664-683.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. New York: Cambridge University Press.

van Elsas, E. J., Lubbe, R., van der Meer, T. W. G., & van der Brug, W. (2014). Vote recall: A panel study on the mechanisms that explain vote recall inconsistency. *International Journal of Public Opinion Research*, *26*, 18-40.

Vavreck L. (2007). The exaggerated effects of advertising on turnout: The dangers of self-reports. *Quarterly Journal of Political Science*, *2*, 325-343.

Appendix



West leads in U.S. job growth

Seven of the 10 states with the fastest job growth this year are expected to be in the West.

Paul Davidson, USA TODAY

Go west, young job-seeker.

Seven of the 10 states with the fastest job growth this year will be in the West, as the region benefits from a stronger housing recovery and continued gains in its bread-and-butter energy, technology and tourism industries, according to forecasts by IHS Global Insight.

The states, which generally led the nation with rapid payroll increases last year, as well, are North Dakota, Texas, Arizona, Colorado, Utah, Idaho and Oregon.

Now that states such as Arizona and Nevada have worked through most of their home foreclosures, residential construction is rebounding sharply, spawning thousands of new jobs, economists say. "It was down so far, and the housing market has finally stabilized," says Richard Wobbekind, head of business research at University of Colorado, Boulder.

Other factors are also at work. North Dakota and Texas are riding an oil boom after largely avoiding the recession's most punishing blows. Colorado and Utah, while enjoying a surge in oil and natural gas drilling, are also now high-tech centers helping satisfy Americans' appetite for mobile devices and applications.

Oregon is a semiconductor manufacturing hub. In Arizona, job growth is being fueled by a technology base that includes Apple's new 2,000-employee glass factory in Mesa, as well as surging tourism, now that rising household wealth is spurring more consumer spending.

As technology increasingly allows Americans to work remotely, the entire Western region is drawing more residents from other states who want to live amid scenic mountains and enjoy a better quality of life, says IHS economist Jim Diffley.

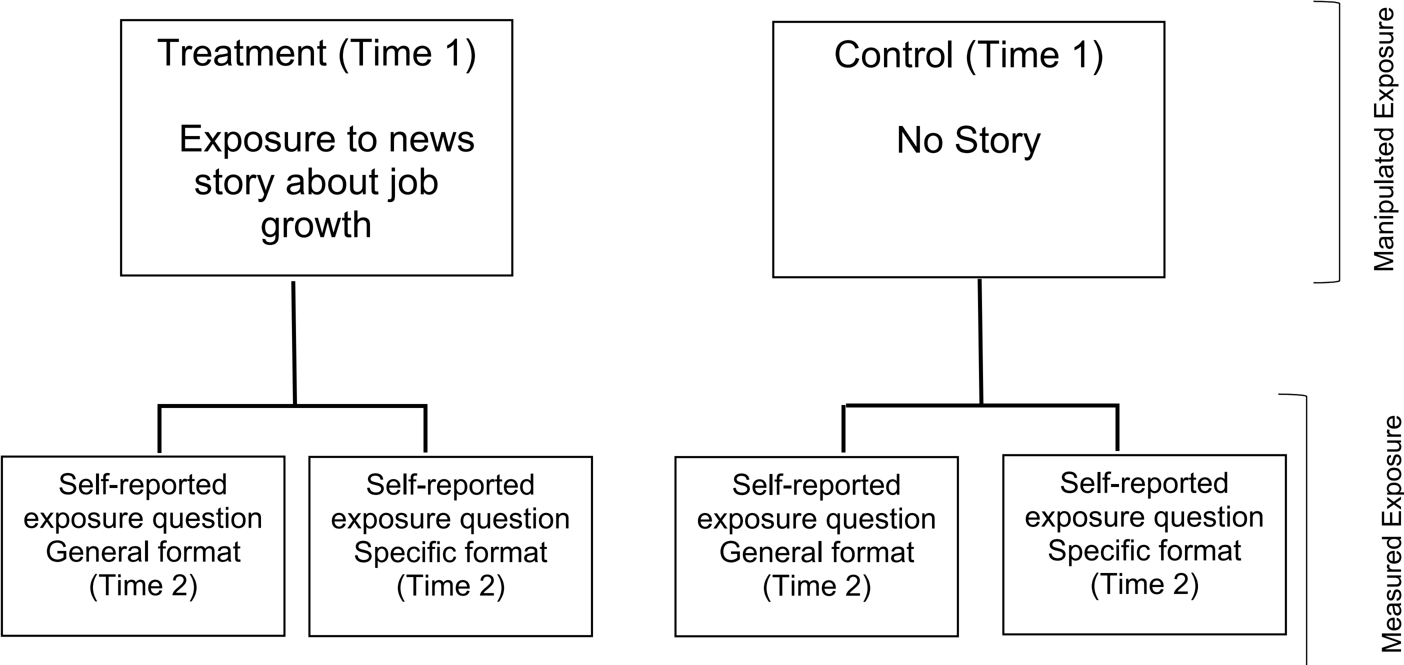
"They're just progressive, attractive places to live," Diffley says.

In the past, migration to the West was limited by the large number of Americans who couldn't move because they owed more on their mortgages than their homes were worth, says economist Chris Lafakis of Moody's Analytics. But the stock of so-called underwater homes has fallen dramatically.

"We expect 2014 to be the year when in-migration (to western states) picks up a lot," Lafakis says.

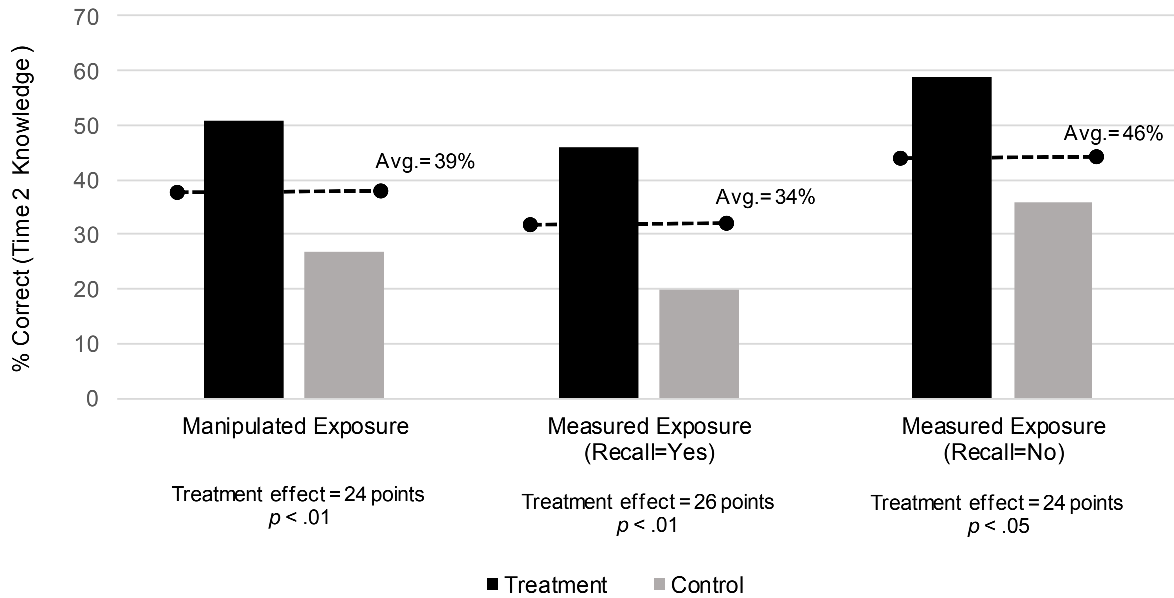
That will increase the need for local services and jobs.

Figure 1. Design Schematic



Note: There was randomization into condition at Time 1 and Time 2. This design was implemented identically in each of the three replications noted in the text. The overall N for the study is 296.

Figure 2. Percentage Correct on Wave 2 Knowledge Question, by Treatment Status and Self-Reported Exposure



Note: Columns represent the percentage giving the correct answer to the Wave 2 knowledge question. The average level of knowledge corresponds to the overall percent correct (i.e., across treatment and control conditions).

Table 1. Predicting Knowledge (Time 2) with Manipulated Exposure

	Coeff.		Coeff.	
Manipulated Exposure (Treatment =1)	.63	***	.62	***
	(.19)		(.19)	
First Replication			.01	
			(.33)	
Second Replication			.15	
			(.20)	
Constant	-.62	***	-.68	***
	(.14)		(.17)	
Log-Likelihood	-115.80		-115.53	
χ^2	10.89		11.46	
N	182		182	

Note: Cell entries denote probit coefficients with standard errors in parentheses.
* $p < .10$; ** $p < .05$; *** $p < .01$ (two-tailed)

Table 2. Predicting Knowledge (Time 2) with Measured Exposure

	Coeff.	Coeff.	Coeff.
Measured Exposure (Recall=Yes)	-.33 *	-.32	-.42 **
	(.19)	(.20)	(.20)
First Replication		.03	-.01
		(.32)	(.34)
Second Replication		.14	.05
		(.20)	(.21)
Male Respondent			.49 **
			(.20)
Need for Cognition			.26 *
			(.15)
Constant	-.09	-.17	-.34
	(.15)	(.18)	(.20)
Log-Likelihood	-119.80	-119.55	-114.15
χ^2	2.92	3.43	14.22
N	182	182	182

Note: Cell entries denote probit coefficients with standard errors in parentheses.
* $p < .10$; ** $p < .05$; *** $p < .01$ (two-tailed)

Table 3. The Effect of Question Form on Recall for Elements of Treatment Strory

	Coeff.
Treatment	-0.26
	(.48)
Specific	-0.25
	(.55)
Specific X Treatment	1.64 *
	(.89)
μ_1	-4.79
	1.06
μ_2	-2.51
	(.48)
μ_3	-0.45
	(.38)
Log-Likelihood	-99.57
χ^2	5.13
N	114

Note: Coefficients are ordered logit estimates (standard errors in parentheses). Model N is lower because the analysis is run on Wave 2 respondents who indicated they saw a story about job growth.

* $p < .10$; ** $p < .05$; *** $p < .01$ (two-tailed)

Table A1. Balance of Variables Across Conditions

	Control	Treatment	Diff.		N
Male	.53	.55	**-.01**		293
	(.04)	(.04)	**(.06)**		
Party ID	.63	.73	**-.10**	***	271
	(.03)	(.02)	**(.04)**		
Foreign-born	.28	.28	**.00**		293
	(.04)	(.04)	**(.05)**		
Socially Desirable Responding	.48	.51	**-.02**		293
	(.02)	(.02)	**(.02)**		
Need for Cognition	.57	.57	**.00**		294
	(.02)	(.02)	**(.02)**		
News Elaboration	.68	.67	**.00**		294
	(.02)	(.02)	**(.02)**		
Knowledge (Wave 1)	.51	.57	**-.06**	*	290
	(.02)	(.02)	**(.03)**		
Knowledge (Wave 2)	.72	.71	**.01**		181
	(.03)	(.03)	**(.04)**		
Internet consumption	.56	.56	**.00**		293
	(.02)	(.02)	**(.03)**		
Newspaper consumption	.12	.13	**-.01**		293
	(.02)	(.02)	**(.02)**		
Radio consumption	.13	.17	**-.03**		293
	(.02)	(.02)	**(.03)**		
TV consumption	.28	.32	**-.05**		293
	(.02)	(.02)	**(.03)**		
Obama approval	.58	.65	**-.07**	**	295
	(.02)	(.02)	**(.03)**		
Attention to politics	.44	.44	**.00**		295
	(.02)	(.02)	**(.03)**		
Risk aversion	.40	.35	**.05**	*	295
	(.02)	(.02)	**(.02)**		
Wave 2 Attrition	.34	.40	**-.06**		296
	(.04)	(.04)	**(.06)**		

*Note: Cell entries are means (rounded to hundredths) with standard errors in parentheses. All variables scaled to [0,1]. In a model predicting treatment assignment at time 2 (specific versus general), the model chi square is insignificant (p =.48).

** $p < .10$; ** $p < .05$; *** $p < .01$ (two-tailed)

Table A2. Factors that Predict Under and Over Reporting

	Control / Recall	Treatment / No Recall
Attention	.25 *	-.41 ***
	(.14)	(.15)
Socially Desirable Responding	-.07	.30
	(.21)	(.26)
Risk Aversion	.03	-.03
	(.08)	(.10)
News Elaboration	-.04	.10
	(.15)	(.19)
Need for Cognition	.03	.02
	(.18)	(.22)
Male	.04	-.24
	(.21)	(.22)
Knowledge (Wave 1)	-.13	.17
	(.10)	(.12)
Constant	-1.16 ***	.13
	(.55)	(.60)
Log Likelihood	-107.47	-71.85
χ^2	4.97	11.89
N	180	180

Note: Coefficients are probit estimates (standard errors in parentheses).
* $p < .10$; ** $p < .05$; *** $p < .01$ (two-tailed)

Table A3: Percent Correctly Answering Knowledge Question at Time 2, by Treatment Status and Self-Reported Exposure

	Manipulated	Measured	
	Actual Exposure	Self-Reported Exposure (Recall=Yes)	Self-Reported Exposure (Recall=No)
Treatment	51% (n=89)	46% (n=59)	60% (n=30)
Control	27% (n=93)	20% (n=54)	36% (n=39)
Overall	39% (n=182)	34% (n=113)	46% (n=69)

Note: Cell entries indicate percentage giving correct response to Wave 2 knowledge question, which asked: "According to economic forecasts, which region of the country has reported the fastest job growth in recent years?" Choices were "East," "West" (correct), "South," and