RESEARCH SYNTHESIS: Measuring Attentiveness in Self-Administered Surveys

Adam J. Berinsky* Alejandro Frydman† Michele F. Margolis‡ Michael W. Sances§ D. Camilla Valerio¶

January 12, 2024

Abstract

The surge in online self-administered surveys has given rise to an extensive body of literature on respondent inattention, also known as careless or insufficient effort responding. This burgeoning literature has outlined the consequences of inattention and made important strides in developing effective methods to identify inattentive respondents. However, differences in terminology, as well as a multiplicity of different methods for measuring and correcting for inattention, have made this literature unwieldy. We present an overview of the current state of this literature, highlighting commonalities, emphasizing key debates, and outlining open questions deserving of future research. Additionally, we emphasize the key considerations that survey researchers should take into account when measuring attention.

For helpful comments, we thank Tom Robinson, John Kane, and panel participants at the 2023 Midwest Political Science Association Conference

*Mitsui Professor, Department of Political Science, Massachusetts Institute of Technology, berinsky@mit.edu.

†Ph.D. Candidate, Department of Political Science, Massachusetts Institute of Technology, afrydman@mit.edu.

‡Associate Professor, Department of Political Science, University of Pennsylvania, mmargo@sas.upenn.edu.

§Associate Professor, Department of Political Science, Temple University, msances@temple.edu.

¶Senior Research Support Associate, Department of Political Science, Massachusetts Institute of Technology, calarcon@mit.edu.

Introduction

Over the past 20 years, online surveys have exploded in popularity, quickly becoming the preferred survey mode for public opinion researchers, academics, and practitioners alike. These online surveys are typically conducted in an unsupervised manner, allowing respondents to more easily complete the survey. But perhaps taking a survey has become too easy. The surge in online self-administered surveys has given rise to an extensive body of literature on survey inattention, also known as careless or insufficient effort responding. Inattention is widely recognized as a significant problem for public opinion research. Inattentive respondents can introduce noise to the data, generating additional measurement errors that can affect scale reliability (Huang et al. 2015; Silber, Danner and Rammstedt 2019; Pyo and Maxfield 2021) and estimated treatment effects (Brühlmann et al. 2020; DeSimone and Harms 2018; Berinsky, Margolis and Sances 2014; Abbey and Meloy 2017; Hauser and Schwarz 2015; Maniaci and Rogge 2014).

This burgeoning literature outlines the consequences of inattention and has made important strides in developing effective methods to identify inattentive respondents. But it has also become unwieldy. As a result, it is often difficult for public opinion researchers to get a handle on current research, much less implement best practices. This paper presents an overview of the current state of the literature, emphasizing central debates and open questions deserving of future research. Additionally, it emphasizes the key considerations that survey researchers should take into account when measuring attention. We acknowledge the dynamic nature of survey research and aim to offer researchers a flexible framework for considering survey attention, independent of specific measurement strategies. While the bulk of the studies included in this review rely on online surveys, the lessons about survey attention are widely applicable to various self-administered survey formats.

We begin by providing some background on what constitutes survey inattention, triangulating differences in terminology used by scholars across fields who draw from the same theoretical frameworks. We then provide an overview of the two general approaches to measuring attention – direct and indirect measures[1] – and examine the challenges associated with these strategies. Next, we discuss the importance of using well-validated measurement strategies and how a multitude of factors, including both researcher decisions and respondent characteristics, can influence attention levels. Finally, we highlight lingering questions regarding how to handle inattentive respondents and discuss potential risks associated with measuring attention.

There are two general themes we hope to convey in this review. The first is the importance of transparency when using attention measurement strategies. Researchers first make decisions about how to measure attentiveness and then decide what to do with inattentive respondents. It is crucial for researchers to disclose any and all research decisions that affect the size and makeup of the sample. Second, researchers must adopt a holistic approach when measuring attention, factoring in the underlying causes of inattention within their survey and recognizing the potential ramifications of their chosen measurement approach on the final results. Importantly,

[1] Throughout the paper we discuss various direct and indirect measures. For detailed explanations, applications, and citations of specific attentiveness measures, please refer to Table 2 and Table 3.

addressing inattentive responding does not have a singular "solution", and this paper will not put one forward. Rather, our focus is on acknowledging inattentive responding as an inherent feature of self-administered surveys that researchers can minimize through considerate survey design, identify using different approaches, and manage once the data is collected. Our goal is therefore to provide researchers with a nuanced understanding of the inattentiveness literature and the ability to apply methodologies that align with their specific research goals.

Theory of Survey Inattention

Much of the research on survey inattention relies on existing theories of survey response behavior to conceptualize inattentive responding and understand the inferential threats posed by this type of behavior. While not all articles do so explicitly, the theoretical frameworks used by scholars to understand inattentive responding have roots in the theories of satisficing (Krosnick 1991) and the cognitive model of survey response (Tourangeau, Rips and Rasinski 2000). Inattention can be conceptualized as a subset of extreme satisficing behaviors where respondents may not even successfully complete the comprehension step of the response process when responding to a survey (Anduiza and Galais 2016). As we discuss in greater length in a later section, this conceptualization of inattention accounts for both individual ability and motivation as well as external features including survey design and survey-taking context. Additionally, it is important to clarify that inattention is not a dichotomous concept where one is either attentive or inattentive. In a later section, we highlight how the use of multiple indicators to measure respondent inattention can leverage to create a more granular index of response inattention.

In Table 1, we present a list of the various terms and definitions used across fields. Although scholars from different fields draw from the same theoretical frameworks when investigating survey inattention, there are terminological differences in how this type of behavior is discussed. Terms such as "inattentive responding," "insufficient effort responding," and "careless responding" are all used by researchers working in this space. For instance, Oppenheimer, Meyvis and Davidenko's (2009) seminal piece introducing instructional manipulation checks (IMCs) explicitly frames that strategy as a method to detect satisficing. Political scientists Berinsky, Margolis and Sances (2014) build on Oppenheimer, Meyvis and Davidenko's (2009) piece and categorize this same behavior as "inattentive responding," which they define as "respondents who offer careless or haphazard survey responses" (p.741). Meade and Craig (2012) use "careless responding" to refer to a subset of response bias termed content non-responsivity or "responding without regard to item content" (p.438). Huang et al. (2012), popularized "insufficient effort responding" – a term now widely used in psychological research – which describes a type of response behavior "in which the respondent answers a survey measure with low or little motivation to comply with survey instructions, correctly interpret item content, and provide accurate responses" (p.100). This type of response behavior includes both random and non-random careless responses. Other political scientists like Alvarez et al. (2019) also use an attention frame and raise concerns over inattentive respondents being a source of "non-sampling bias and response error" because these respondents are not "mindful in the survey process" (p.146). Importantly, across definitions, there is a consensus over a conceptual distinction between inattentive responding and dishonest responding (or response distortion), with the latter treated as a different type of response behavior (Arthur, Hagen and George 2021).

Table 1: Definitions of Attentiveness Used Throughout Disciplines

Term	Definition	Discipline	Citation
Inattentive / Low-Effort Responding	Low-quality responses from participants who devote less effort and offer quick answers	• Statistics • Psychology • Business • Political Science	• Olamijuwon (2021) • Buchanan & Scofield (2018) • Abbey & Li (2018) • Alvarez & Li (2021) • Breitshol & Steidelmuller (2018) • Bowling et al. (2016) • Kane, Velez & Barabas (2020) • Malone & Lusk (2018)
Insufficient Effort Responding (IER)	"Response set in which the respondent answers a survey measure with low or little motivation to comply with survey instructions, correctly interpret item content, and provide accurate responses" (Huang et al. 2016, p. 101). IER may manifest itself as either random responding or as non-random response patterns, and it encompasses both unintentional, occasional careless responding, and intentional 'speeding-through' of survey items" (Huang et al. 2015, p. 301).	• Psychology	• Huang et al. (2012/2015) • Bowling et al. (2021) • Toich et al. (2021) • Iaconelli & Wolters (2020)
Careless Responding	"Respondents intentionally or unintentionally answer survey items in a manner that does not accurately reflect their true sentiments" (Ward & Pond, p. 554)	• Psychology • Marketing	• Ward & Pond (2015) • Meade & Craig (2012) • Oppenheimer et al. (2009) • Weathers & Bardacki (2015) • Niessen, Meijer, & Tendeiro (2016)
Carless/Insufficient Effort (C/IE) Respondents	Response vectors resulting from lack of attention or effort, where the individual responds without sufficient attention to the content and semantic polarity of the items" (Arias et al. 2020, p. 2490). "C/IE responders are missing data that is not actually missing. They have provided a response when they might have well left that response blank" (Curran 2016, p. 4).	• Psychology • Marketing	• Arias et al. (2020) • Curran (2016) • Weathers & Bardacki (2015)
Straightlining	"One of the most prevalent forms of C/IE, where the person provides similar responses regardless of the content and direction of the item" (p. 2490).	• Psychology	• Arias et al. (2020)
Random Responding	"Respondent does not attend to the content of the item, but intentionally uses all response categories to appear to respond thoughtfully" (p. 2490)	• Psychology	• DeSimone & Harms (2018)
Noncompliant Responding	"Individuals who show too much or too little variation in their responses" (p. 86).	• Psychology	• Barends & de Vries (2019)

Stereotyped Response	"A response that does not accurately represent subjects' attitudes" (p. 63)	• Business	• Abbey & Meloy (2018)
Response Validity	"The extent to which scores reflect the thoughts and beliefs of respondents concerning the phenomena of interest...If reported scores are aligned with the respondent's cognitions, then response validity is high, whereas if the scores are misaligned with what the respondent actually thinks, perceives, or feels, response validity would be low" (p. 63).	• Business	• Edwards (2019)
Satisficing	"When responding to survey questions, rather than going through a deep mental process and coming up with an optimal answer, some respondents tend to take a short cut and stop the mental process prematurely" (Krosnick 1991, p. 33)	• Psychology	• Liu & Wronski (2018) • Krosnick (1991) • Conrad et al. (2017) • Anduiza & Galais (2017) • Moren & Pass (2020)
Content Nonresponsivity	"Careless respondents give answers of bad data quality regardless of the content of the question" (p. 2)	• Psychology	• Bruhlmann et al. (2020)

Whether researchers use "insufficient effort responding" or "inattention" to describe inattentive respondents, conceptually, they are referring to the same type of response behaviors. Though different fields have developed their own vocabularies to describe these behaviors, it is important to recognize the similarities in the underlying concepts being studied and not let semantic differences inhibit cross-field collaborations. In this review, consistent with Berinsky, Margolis and Sances (2014), we will use the term "inattentive responding."

Measuring Inattention

A key consideration facing scholars is *how* to measure attention. This decision is crucial because incorrectly flagging attentive respondents as inattentive (Type I error) and failing to identify inattentive respondents as such (Type II error) can pose threats to the validity of the measures used by researchers (Curran 2016; Abbey and Meloy 2017).[2]

Direct Measures

Direct measures (sometimes referred to as attention checks or screeners) involve explicitly asking survey respondents a question or series of questions to determine some level of attentiveness. Common strategies include instructional manipulation checks (IMCs), bogus items, and instructed response items (IRIs). Newer methods include Factual Manipulation Checks (Kane and Barabas 2019) and Mock Vignette Checks (Kane, Velez and Barabas 2023). See Table 2 for a list of the various direct methods along with relevant articles that serve as good examples for each strategy.

[2]Moreover, the consequences of inattention likely differ across studies. In a survey experiment, for example, inattentive respondents in the treatment group may not have actually been "treated."

Table 2: Direct Measures

Measure	Definition	Application	Citation
Instructional Manipulation Checks (IMC)	"A question embedded within the experimental materials that is similar to the other questions in length and response format (e.g. Likert scale, check boxes, etc.). However, unlike the other questions, the IMC asks participants to ignore the standard response format and instead provide a confirmation that they have read the instruction" (Oppenheimer, Meyvis, Davidenko 2009, p. 867)	When a big news story breaks people often go online to get up-to-the-minute details on what is going on. We want to know which websites people trust to get this information. We also want to know if people are paying attention to the question. To show that you've read this much, please ignore the question and select ABC News and The Drudge Report as your two answers (p. 740).	• Berinsky, Margolis & Sances (2014) • Hauser et al. (2016) • Anduiza & Galais (2017) • Paas, Dolnicar & Karlsson (2018) • Mancosu, Ladini & Vezzoni (2019) • Moreen & Paas (2020
Bogus Item/Trap Question / Infrequency Scale	"Items are constructed to appear face valid on a quick visual inspection, but obvious or absurd on deeper inspection" (Curran 2016, p. 13)	Ex: "I was born on February 30th."	• Meade and Craig (2012) • Huang et al. (2012) • Huang et al. (2015) • Liu & Wronski (2018) • Kim et al. (2018) • Curran and Hauser (2019)
Instructed Response Item	"Instructed response items [ask] respondents to provide a specific response and are often part of a grid of questions or a group of questions with the same scale... Instructed-response items evaluate respondents' compliance with simple and concise instructions" (Alvarez, Atkeson, Li 2019, p. 147)	For this item, select Response Option 5. If respondents do not endorse the instructed response option, it is assumed they are being inattentive (p.375).	• Bowling et al. (2016) • Beck, Albano & Smith (2018) • Alvarez et al. (2019) • Shamon & Berning (2020) • Berinsky et al. (2021) • Gummer, Rossman, & Silber (2021)
Factual Manipulation Checks	Ask respondents objective questions (i.e. questions with correct answers) about key elements of the experiment or about a study's contents (p. 234/5)	Study about student loan forgiveness asks, "According to the paragraph you just read, what is a potential consequence of the student loan forgiveness program?" Response options were in multiple-choice format (p. 240).	• Kane & Barabas (2019)
Mock Vignette Check	Feature content substantively similar to that of experimental vignettes in political science, and are followed by factual questions (mock vignette checks) that gauge respondents' attentiveness to the mock vignette. It is a design-based technique that is meant to overcome the risk of post-treatment bias when trying to measure attention.	Brief vignette about an unrelated topic meant to be similar to the treatment, followed by a series of factual questions related to the content of the vignette.	• Kane, Velez, & Barabas (2022)
Response Consistency	Asking respondents the same question twice in different forms to see if answers vary.	Survey asked respondent's age twice in two separate formats, the first question of the survey ("In what year were you born?") and the last question ("What is your age?") (p. 35).	• Liu & Wronski (2018)
Response Rounding	Questions asking respondents to type in a numeric answer in a textbox	On average, about how many hours do you watch television?	• Liu & Wronski (2018)
Knowledge Question	Questions based on common knowledge.	The logo for the Olympic Games is comprised of four interlocking rings (Response options: Yes, No) (p. 35).	• Liu & Wronski (2018)

While scholars have introduced a variety of direct measures, details about their implementation and best practices remain open questions. In an early paper on IMCs, Berinsky, Margolis and Sances (2014) note that because we think of attentiveness as a latent construct, these types of questions are subject to measurement error like any other survey question. As a result, they recommend using multiple IMCs and aggregating responses into an attentiveness scale rather than relying on a single measure. In a later paper, Berinsky et al. (2021) investigate both stand-alone items along with items embedded within a grid (instructed response items and bogus items) and recommend the use of a multi-item scale that includes questions with both high and low passage rates. Other scholars concur, calling for the use of multiple screeners (Thomas and Clifford 2017; Kane, Velez and Barabas 2023) with varying length and difficulty (Morren and Paas 2020). [3]

Of course, asking additional questions increases the survey length, thereby increasing the likelihood of inattentive responses and the cost of a survey. IMCs are stand-alone questions that take up more space in a questionnaire compared to bogus items (also known as infrequency items (Huang et al. 2015)) that can be inserted into a larger grid of questions. In fact, Liu and Wronski (2018) find that passing one attention check does not predict passing another one within the same survey and recommend only including a single question (though these authors stand apart from the field in advising against the use of multiple measures). Additionally, a series of articles have recommended using shorter and easier direct measures as opposed to longer ones that require a higher cognitive load (Liu and Wronski 2018; Mancosu,

[3] A challenge arises when using multiple items, as determining the appropriate threshold to identify a respondent as inattentive is not a straightforward process. See further discussion below.

Ladini and Vezzoni 2019; Ladini 2022; Anduiza and Galais 2016). Mancosu, Ladini and Vezzoni (2019) find that answer quality among respondents who pass an easy screener is similar to those who pass a medium or complex screener, recommending researchers resort to simpler and shorter questions.

Another implementation concern involves the placement of an attention check. Liu and Wronski (2018) find no difference in passage rates between questions at the beginning or the end of a survey, while Reyes (2022) recommends placing items later in the survey. If the survey includes an experiment, it may be important to ensure that the direct measure comes before the treatment is administered to avoid the risk of post-treatment bias (Montgomery, Nyhan and Torres 2018). Some argue that individual attention may wax and wane across a survey (Alvarez et al. 2019) so a direct measure should be asked as close as possible to the treatment, though Berinsky, Margolis and Sances (2014) explicitly test this question and find no evidence of significant variation in respondent attention.

Taken together, the research has not yet reached a consensus on how best to implement direct attention measures – including the optimal number, difficulty, and placement within the survey. An important final note is that these considerations are not independent of one another. For example, if a researcher has the space to use multiple measures then considerations about difficulty and placement will be different than if a researcher is only able to use a small number of measures.

Indirect Measures

Direct measures, however, are not the only way to measure attention on a survey. As opposed to direct measures, indirect measures do not involve explicitly asking respondents questions but instead are typically calculated post-hoc, using information such as response patterns and response times. There are a wide variety of indirect response measures, and a few studies have done comprehensive jobs summarizing and comparing them (Maniaci and Rogge 2014; Curran 2016; Niessen, Meijer and Tendeiro 2016; Abbey and Meloy 2017; DeSimone and Harms 2018; Leiner 2019; Goldammer et al. 2020; Brühlmann et al. 2020; Hong, Steedle and Cheng 2020; Ward and Meade 2023). Table 3 provides a list of indirect measures along with brief descriptions and examples.

Table 3: Indirect Measures

Measure	Definition	Applications
Response Time Approaches	Post-hoc method relying on survey completion times to detect inattentive responding.	• For a review, see (Matjašič, Vehovar and Manfreda 2018) • Response time (Meade and Craig 2012; Weathers and Bardakci 2015) • Floodlight detection (Dogan 2018) • Response Time Attentiveness Clustering (Read, Wolters and Berinsky 2022) • Response Time-based Latent Response Mixture Model (Ulitzsch et al. 2022)
Person-Fit Approaches	Assessing the fit between an individual's response and the underlying measurement model to detect responses that are not consistent with the measurement model.	• Person-Fit Statistics (ex. Guttman errors) (Niessen, Meijer and Tendeiro 2016) • Person-Fit Index (Beck, Albano and Smith 2019) • Interactive Cleansing Method (Patton et al. 2019)
Response Pattern Approaches	Identifying patterns of inconsistent or unreliable responding that suggests that a respondent is not paying attention or responding carefully.	• LongString Index (Johnson 2005; Huang et al. 2012) • Response Pattern Indices (Huang et al. 2012) • Individual Consistency/Inconsistency (Ward and Pond 2015) • Response Variance (Weathers and Bardakci 2015) • Random Response Share (Malone and Lusk 2018) • Semantic synonyms/antonyms (DeSimone, Harms and DeSimone 2015)
Outlier Approaches	Identifying and investigating individuals who are responding without sufficient effort that differs from their thoughtful counterparts in some way.	• Mahalanobis Distance (Ward and Pond 2015)

A key distinction between direct and indirect measures is that indirect measures do not require "correct" or "incorrect" answers to determine whether a respondent is flagged as inattentive. Instead, researchers rely on observed response behaviors (either response patterns or response times) believed to be indicative of inattention. Researchers then quantify the prevalence of these behaviors and assign a score to each respondent. They then determine whether someone is attentive or not by comparing respondents' scores against set thresholds. A main benefit of this strategy is that indirect measures do not take up any real estate on a survey and are unobtrusive, meaning that survey respondents do not know their behaviors are being monitored. The main drawback of this strategy is that researchers have to make two critically important decisions. First, researchers must decide what type of response behaviors they deem deviant and ensure that the observed behavior effectively captures inattention. Second, researchers must select a cut-off threshold used to flag respondents as inattentive. We describe the current state of research regarding both decisions below.

Response Patterns

One common indirect approach involves analyzing individual response patterns from long question batteries. These pattern-based indirect measures are popular in psychological research as respondents often complete long grids of questions. LongString analysis, in which researchers identify careless respondents by measuring the longest consecutive string of identical survey responses, is one response pattern strategy. This type of response behavior is also referred to as straightlining. The individual consistency approach is another strategy. Here, researchers identify respondents who provide contradictory or inconsistent responses within a long question battery. For a comprehensive review of different indirect measures and their application in psychology, see Ward and Meade (2023).

Studies that evaluate the effectiveness of multiple measures repeatedly find that while individual response pattern measures tend to be correlated, they do not all identify the same respondents as inattentive (Maniaci and Rogge 2014; DeSimone and Harms 2018). This should not be a surprise, as different methods are meant to identify different types of behavior. As mentioned earlier, LongString analysis can identify inattentive behavior that manifests through straightlining[4], while individual consistency measures identify a different, less obvious response pattern. Researchers employing a response-pattern strategy should therefore identify what sort of response-pattern behavior is indicative of inattention in their specific study.

Response Times

Another type of indirect measure of attentiveness employs individual response times.[5] In brief, researchers identify outliers – those whose time spent on a survey is far different from the majority of respondents – to identify inattentive respondents on a survey.

Importantly, inattentive respondents may either complete the survey too fast or too slow (Read, Wolters and Berinsky 2022). Respondents may speed through a

[4]Inattentive behavior like straightlining can also lead to *inflated* scale reliability due to artificial correlation between scale items.

[5]Matjsasic et al. (2018) provide a review of the state of the literature on response times.

survey without fully comprehending questions, or respondents may take an extended time to complete the survey, all while not focusing on the content of the survey due to distractions. Furthermore, there is no consensus about the best way to measure response time (entire survey, page time, question time, etc.) (Greszki, Meyer and Schoen 2015; Matjašič, Vehovar and Manfreda 2018; Leiner 2019).

Additionally, some researchers question the exclusive use of response time measures. Matjašič, Vehovas and Sendelbah (2021) find that many response time strategies are insufficient on their own when compared to direct measures and response pattern-based indirect measures. Researchers interested in using response times must therefore decide whether they are concerned about rushing respondents, distracted respondents, or both; care about the time spent on the survey as a whole or specific parts of the survey; and whether measuring response time alone is a sufficient strategy for identifying inattentive respondents.

Flagging Respondents and Setting Appropriate Cut-Offs

For both response time and response pattern-based indirect measures, decisions over appropriate cut-off points are critically important because such decisions can significantly influence which respondents are classified as attentive (Goldammer et al. 2020; Brühlmann et al. 2020; Chmielewski and Kucker 2020; Leiner 2019; DeSimone and Harms 2018; Curran 2016; Maniaci and Rogge 2014; Meade and Craig 2012; Wood et al. 2017). If the cut-off is too restrictive, researchers inflate the number of inattentive respondents by falsely flagging attentive respondents as inattentive. If the cut-off is too relaxed, then only the most inattentive respondents may be flagged, artificially deflating the level of inattention in the sample.

With this concern in mind, scholars have proposed new strategies to determine these thresholds for both response times (Ulitzsch et al. 2022; Dogan 2018; Read, Wolters and Berinsky 2022) and response patterns (Kim et al. 2018; Dunn et al. 2018; Patton et al. 2019; Yu and Cheng 2019; Schroeders, Schmidt and Gnambs 2022). One example is Read, Wolters and Berinsky (2022) who leverage per-question response times along with dimension reduction and an unsupervised clustering algorithm to classify attentive respondents without having to specify a specific speed threshold. Other examples include Ulitzsch et al. (2022) and Arias et al. (2020) who both propose model-based methods leveraging mixture models that do not depend on researcher-defined cut-offs. Another approach proposed by Harden, Sokhey and Runge (2019) uses response times as a measure of treatment compliance in surveys. This approach sidesteps the threshold discussion by conceptualizing inattention as a compliance problem. In doing so, researchers can estimate the causal average compiler effect (CACE) among attentive respondents.[6]

Decisions over cut-offs and flagging respondents are further complicated when scholars choose to use multiple indirect measures, as implementation is not straightforward (Meade and Craig 2012; DeSimone and Harms 2018). Due to the wide selection of indirect measures available, it is not always clear what the optimal combination of measures is. Moreover, different measures each have their own cut-off scores and may identify different sets of respondents. One recommendation for using multiple measures is the "multiple hurdle" approach where different methods are

[6]This approach does, however, add additional modeling assumptions associated with two-stage least squares estimation.

deployed sequentially to identify inattentive respondents (Curran 2016; Goldammer et al. 2020). Brühlmann et al. (2020) warn this approach may be too restrictive and lead to false positives. Additionally, there is concern over the aggregation of multiple indices that involve user-defined cutoffs as opposed to more objective indicators used in direct measures (Maniaci and Rogge 2014; Brühlmann et al. 2020). A different approach involves calculating multiple indirect measures, but then identifying careless respondents using the "best subset" of measures that optimizes a balance of sensitivity and specificity across different careless responding behaviors (Hong, Steedle and Cheng 2020).

All told, the research suggests that scholars should think carefully about what sorts of inattentive behaviors pose the greatest threats to their study and what kind of direct or indirect measure is most appropriate to identify respondents engaging in this behavior. Moreover, researchers should consider the empirical consequences of employing different cut-offs, different modeling approaches, and multiple measures of attentiveness.

Can We Increase Attentiveness?

In addition to trying to measure attention, researchers can attempt to increase or encourage attention through the use of warnings or instructions. These strategies forgo identifying a specific type of response behavior, and instead, rely on inducing more attentive responses through the survey design. As an example, researchers may include a warning to respondents that their answers will be reviewed for quality control. The evidence in support of using warnings through survey instructions to increase attention is mixed (Meade and Craig 2012; Ward and Pond 2015; Clifford and Jerit 2015; Breitsohl and Steidelmüller 2018; Paas, Dolnicar and Karlsson 2018; Ward and Meade 2018; Shamon and Berning 2020; Toich, Schutt and Fisher 2022; Bowling et al. 2021). A related approach involves introducing virtual proctors that encourage attentive responding, though evidence in support of this approach is limited (Francavilla, Meade and Young 2019). Researchers can also incorporate warnings with certain indirect measures like response time. For example, Conrad et al. (2017) implement an interactive prompting technique where respondents who answered faster than a certain response time threshold received a message encouraging them to answer carefully and to take their time. An extension of this strategy involves explicitly "training" respondents who fail attention checks through the use of additional instructions. While training respondents may increase attention check passage rates, Berinsky, Margolis and Sances (2016) find that training does not increase general attentiveness throughout the survey. Importantly, these strategies used to increase attention are not mutually exclusive from the other measurement approaches discussed earlier. Future work should explore how a combination of both approaches can be used to improve data quality, encouraging attentiveness upfront while retaining traditional measures for identifying inattentive respondents when necessary.

Validation Strategies

When measuring attention, researchers should ensure that the strategy chosen appropriately captures the construct of interest. As a result, it is imperative for researchers to employ high-quality measures. For most researchers, this simply means selecting measurement strategies that have already been carefully validated. Scholars interested in developing new measurement strategies, on the other hand, ought to devote significant effort to show that their proposed measures are valid and reliable. While the number of novel methods to measure survey attention has grown, there is a lack of common validation strategies that facilitate comparison across different methods.

When developing new methods, the choice of appropriate validation approach depends on whether the proposed measurement strategy is direct or indirect.[7] A recent study by Kane, Velez and Barabas (2023) introduces a novel direct measure called "Mock Vignette Checks," providing a compelling example of proper validation. This new measurement strategy relies on using a pre-treatment "mock" experimental vignette that approximates the true treatment as a measure of attention. The authors employ a variety of validation strategies to test the construct and convergent validity of the measure. First, the authors replicate multiple experiments with known treatment effect sizes to evaluate treatment effect heterogeneity across levels of attention. This strategy is common as it allows researchers to compare results to past studies that have already been replicated (Berinsky, Margolis and Sances 2014; Greszki, Meyer and Schoen 2015; Brühlmann et al. 2020). In addition to

[7] This difference is due to the distinctions in how those questions measure inattention, as discussed earlier.

replication, the authors also compare mock vignette checks to other validated direct measurement strategies such as instructional manipulation checks along with a latent measure of attention such as response time. Importantly, these tests are conducted across multiple different sample providers.

Another way researchers have validated their direct measures of attention is by examining responses to factual survey questions that can be compared to external administrative data (Alvarez and Li 2023; Clifford and Jerit 2015). Alvarez and Li (2023) used a survey with validated voter turnout to assess the accuracy of the direct attentiveness measures. Using the behavioral data, the authors find that inattentive respondents are less likely to provide accurate reports of their voting history.

Some measurement strategies, such as response-time-based approaches, are especially difficult to validate due to the possibility that inattentive respondents may complete a survey either too *quickly* or too *slowly*. The literature on response time-based approaches generally lacks consensus on appropriate validation strategies (Meade and Craig 2012; Ulitzsch et al. 2022; Conrad et al. 2017; Dogan 2018; Matjašič, Vehovar and Manfreda 2018). In a recent paper Read, Wolters and Berinsky (2022) proposed a new measure called response-time attentiveness clustering (RTAC), which leverages dimension reduction and an unsupervised clustering algorithm to identify inattentive respondents. To validate this new measure, the authors used three strategies which included comparing open-ended questions across different response times, replicating a well-known survey experiment, and using a flipped scale in a series of ideological questions. To evaluate differences in scale reliability across levels of attention, the authors used Cronbach's alpha, a common tool used to validate indirect measures that rely on long scales (Hauser and Schwarz 2015; Ladini 2022).

There are a few important takeaways here. For practitioners interested in measuring attention, it is essential to only employ properly validated measures. On the other hand, methodologists interested in proposing new measures must spend significant effort validating these measures. When doing so, researchers must ensure that the strategies employed align with the goal of the measurement strategies being tested. Additionally, the validation strategies should ensure both construct and convergent validity. Finally, an undervalued component of developing a new measure involves making it easy for practitioners to correctly implement the validated method. Providing practitioners with ready-made example questions can facilitate the appropriate implementation of the measurement strategy (Berinsky, Margolis and Sances 2014; Kane, Velez and Barabas 2023).

Who is Inattentive and Why?

Aggregate rates of survey inattention vary widely across studies and depend on a multitude of factors related to researcher decisions, survey context, and respondent behavior. First, researcher decisions can influence passage rates at multiple points in the survey process. Most notably, the choice of measurement strategy has a notable impact on flagging inattentive respondents, as different strategies may identify different respondents (Maniaci and Rogge 2014; DeSimone and Harms 2018). As discussed earlier, researchers can also affect passage rates when using both direct and indirect measures through choices related to cut-off thresholds and decisions over question construction such as the length, difficulty, and placement of an item. If a researcher chooses a difficult direct measure, a greater number of respondents will be flagged as inattentive.

Second, the choice of sample provider used can affect levels of attention (Abbey and Meloy 2017; Brühlmann et al. 2020). Researchers have found variation in attentiveness rates across sample providers such as MTurk (Chmielewski and Kucker 2020) and Lucid (Ternovski et al. 2022), as well as comparisons between samples (Hauser and Schwarz 2016; Arndt et al. 2022). Differences in sample screening methods across providers produce variations in sample quality, which can impact sample attentiveness. Sample quality may also fluctuate over time for a particular sample provider, resulting in temporal variation in passage rates (Ternovski et al. 2022; Peyton, Huber and Coppock 2022).

Third, scholars have explored how respondent-level characteristics correspond with individual-level attention. Evidence on the relationship between personality traits such as agreeableness or consciousness and attentiveness is mixed (Maniaci and Rogge 2014; Bowling et al. 2016; Palaniappan and Kum 2019). Bowling et al. (2016) argue that insufficient effort responding is a reflection of personality traits and find evidence in support of this claim – over time, individual levels of inattention were consistent, reflecting what they call "enduring individual differences" (Bowling et al. 2016). Other scholars disagree and find suggestive evidence that inattention may operate independently of personality traits and call for additional research on the subject that incorporates both individual traits along with survey context-specific factors such as the relevance of the survey to respondents, attention capacity, and other contextual factors (Maniaci and Rogge 2014; Palaniappan and Kum 2019). This research can build on past work linking survey attention to motivation (Bowling et al. 2021; Rios et al. 2017). Of interest is a potential interaction between personality and context where individuals with certain personality traits might be less likely to respond carefully if the survey is of little interest to them. Additionally, this question of internal motivation is also tied to the survey-taking *context*. For example, an individual's capacity to engage with a survey will likely vary depending on whether they take it uninterrupted at home or during their commute on public transportation. While no existing work directly examines this question, insights about the role of survey *design* in influencing attention can likely extend to the role of the survey-taking context.

Additionally, some studies have found demographic differences between attentive and inattentive respondents in terms of education, gender, and age (Berinsky, Margolis and Sances 2014; Anduiza and Galais 2016; Mancosu, Ladini and Vezzoni 2019; Alvarez et al. 2019; Paas, Dolnicar and Karlsson 2018). Other studies have found consistent differences along age and race lines. In general, studies have found that whiter, older, and more educated respondents are more likely to be identified as attentive. While these differences tend to be substantively small, they raise concerns about how adjustments for inattention may affect external validity (Alvarez et al. 2019; Thomas and Clifford 2017; Kane, Velez and Barabas 2023).

These findings further reinforce the importance of transparency throughout the research process. Since attentiveness measures are predicated, in part, on how researchers measure and determine attention, researchers must be clear about the choices they make in this arena. Moreover, potential differences between samples and demographic disparities between attentive and inattentive respondents underscore the importance of providing comprehensive details about the data collection process and the demographic characteristics of the attentive group.

Handling Inattentive Respondents

Following decisions about how to measure attentiveness on a survey, researchers must then consider the appropriate approach to handle the respondents identified as inattentive. If inattentive respondents are viewed as random noise, the easy way to handle them would be to drop them from the sample. Alvarez et al. (2019) present a useful framework for handling inattentive respondents and discuss four different strategies. These strategies include doing nothing and keeping all respondents, dropping inattentive respondents, dropping and re-weighting, and keeping all respondents and accounting for inattention through a model-based adjustment. Alvarez et al. (2019) review the limitations of each of these approaches and the situations in which they would be appropriate, highlighting the need for researchers to carefully consider the trade-offs associated with these decisions and consider how choices made align with the research goals. Alvarez and Li (2023) offer further guidance for researchers, suggesting they examine whether survey results are sensitive to attention levels and if attentiveness correlates with the outcome of interest. If either of these conditions holds, then dropping respondents is likely not appropriate.

Removing inattentive respondents can improve scale reliability, increase precision, reduce bias, and increase statistical power by reducing noise (Hauser and Schwarz 2015; Thomas and Clifford 2017; Abbey and Meloy 2017; Hong, Steedle and Cheng 2020; Gummer, Roßmann and Silber 2021; Pyo and Maxfield 2021). But under what conditions can these respondents be considered true noise? Importantly, no single rule dictates when dropping respondents is acceptable and caveats exist regarding the type of sample used and the different conditions that make removal acceptable.

Many scholars argue that researchers should not drop inattentive respondents because it could bias the estimates (Berinsky, Margolis and Sances 2014; Ward and Pond 2015; Anduiza and Galais 2016; Mancosu, Ladini and Vezzoni 2019; Tyler, Grimmer and Westwood 2022; Atsusaka and Stevenson 2023). If inattention is correlated with theoretically relevant demographic characteristics, dropping respondents can introduce bias in the sample. Additionally, removing respondents may significantly reduce the size of the sample, reducing statistical power. Silber, Danner and Rammstedt (2019) argue that removing inattentive respondents only marginally increases data quality so it may not be worth the potential risks. DeSimone and Harms (2018) make a similar argument. One common recommendation is to create an additive index of attention and report results at multiple levels of attention (Berinsky, Margolis and Sances 2014; Alvarez and Li 2023). Ultimately, researchers need to be explicit about their choices and, if they choose to drop inattentive respondents, they should report results among both the full sample and the attentive sample, along with their demographic compositions. These practices maximize transparency and allow readers to evaluate how robust the results are to inattention along with adjustments for inattention.

In a similar vein, to justify the exclusion of inattentive respondents, Buchanan and Scofield (2018) conducted a sensitivity analysis where they varied the proportion of the sample that was flagged as inattentive to better understand the consequences of removing these respondents. They find keeping respondents flagged as inattentive can increase noise in the data and reduce the study's statistical power. In a working paper, Tyler, Grimmer and Westwood (2022) propose a new statistical framework that allows for the identification of population attitudes and treatment effects from surveys that contain a mix of attentive and inattentive respondents without the risk of having to drop respondents. Leveraging cutting-edge machine learning and weighting techniques, the authors provide an example of statistical methods for handling inattention, a promising area for future research.

Finally, researchers should assess the possible implications of striving for a fully attentive sample. For example, consider a study on the persuasiveness of different frames of a campaign ad. In the real world, many individuals exposed to campaign messaging may not actually pay attention to the ads. To more accurately reflect the continuum of attention in the broader public, it may be necessary to retain inattentive respondents in the sample, as removing inattentive respondents may result in inflated treatment effect estimates. On the other hand, if the experimental treatment is simply a means to induce variation in something else, say emotion, inattentive respondents may contribute nothing to our ability to learn about the effects of emotions a researcher may be justified in adjusting for inattention in some way. This example highlights the importance of researchers considering the desired attention level within a sample.

Potential Risks of Measuring Attention

When determining what strategies to use to identify inattentive respondents, researchers should also consider the possibility that in addition to measuring attention, these strategies can potentially influence respondent behavior in unintended ways. Such risks are especially relevant when it comes to direct measures. Concerns about the unintended effects of attention checks connect to broader questions of survey design. As discussed earlier, researchers need to be careful when designing direct attention checks and consider trade-offs related to the difficulty, length, and placement of the question. Ideally, researchers should aim to implement attention checks that capture inattentive respondents without altering respondent behavior. Hauser and Schwarz (2015) argue that IMCs can impact systemic thinking, though Hauser and Schwarz (2016) later add an important caveat: this concern may only apply to certain complex reasoning tasks, not survey taking in general. Consequently, researchers must consider whether this threat is relevant to their specific research questions and survey instrument. In a review, Abbey and Meloy (2017) also warn of possible psychological consequences of obstructive attention checks that may generate unintended emotional responses. It should be noted that the consequences of attention checks are not necessarily purely negative; Shamon and Berning (2020) find that certain direct measures do not have demotivating effects on response behavior, but instead, have a motivational influence on respondents by increasing cognitive effort. While inducing attentive responses through the use of an attention check may improve data quality, researchers must also be mindful of the possibility of artificially inflating attentiveness within the sample.

A growing body of research has found that most direct measures seem to have little effect on respondent behavior–either positive or negative. There is little evidence that direct measures bias survey responses or threaten scale validity (Kung, Kwok and Brown 2018; Paas, Dolnicar and Karlsson 2018; Gummer, Roßmann and Silber 2021), lead to survey drop-off (Paas and Morren 2018), or deteriorate the survey-taking experience (Breitsohl and Steidelmüller 2018). These findings apply across a variety of direct methods (Gummer, Roßmann and Silber 2021; Paas, Dolnicar and Karlsson 2018; Huang et al. 2015). Encouragingly, Kane, Velez and Barabas's (2023) recent work finds that the effect of including a mock-vignette check on non-attention-related response behavior is close to zero.

While the growing evidence affirming the lack of negative consequences regarding attention checks further solidifies the value of these strategies, future work should explore the effects of variation in both the difficulty and number of attention checks on survey behavior. Gummer, Roßmann and Silber (2021) express concern over possible spillover effects due to too many attention checks. Additionally, researchers should continue to explore the relationship between attention checks and different types of survey tasks ranging from simple survey questions to more complex reasoning tasks (Hauser and Schwarz 2015; Kung, Kwok and Brown 2018). Researchers may also wish to avoid recycling common questions as certain sample providers may have subsets of experienced survey takers who may be familiar with attention checks, pass them, and then continue to provide inattentive, low-quality answers to the rest of the survey (Barends and de Vries 2019).

When thinking about the unintended consequences of measuring attention, researchers should keep a few other points in mind. Curran and Hauser (2019) urge researchers to account for the threat of false positives when using bogus or infrequency items. Liu and Wronski (2018) also find evidence of both false positives and false negatives when using instructed response items. When conducting cross-national research, Grau, Ebbeler and Banse (2019) argue for caution when measuring attention due to the possibility of cultural sources of variation in attention. Alvarez and Li (2023) also discuss the decision between different types of direct measures as a sort of bias/variance trade-off. They see IMCs as a more aggressive measure that may reduce bias in survey estimates, but introduce variance by flagging respondents who may still provide some attentive responses.

Conclusion

In this review, we examine a wide variety of strategies for measuring survey attention and debates over best practices when implementing these strategies. Scholars working within this literature have often focused on narrow research questions at the expense of developing a higher-level set of recommendations aimed toward helping public opinion researchers understand the central considerations when measuring attention and addressing inattentive respondents. This is evident in the use of varying terminology across disciplines to describe essentially identical classes of response behaviors. Rather than siloing small advances within different academic disciplines, future research on survey attention should strive for interdisciplinary collaboration, focusing on a common goal of improving survey data quality.

Throughout this piece, we stress the importance of survey researchers weighing multiple considerations when deciding how to measure inattention. Additionally, we emphasize the need for greater transparency when measuring survey attention. Researchers need to adopt a holistic approach to measuring attention, thinking deeply about how their research objectives relate to the significance of attentive respondents and how different measurement strategies might affect those goals. Factors such as the sample provider, survey content, and survey design should all influence researchers' decisions on how to address inattention.

Researchers must also exercise caution when flagging respondents as inattentive as decisions over the difficulty of an attention check or appropriate cut-off threshold will affect passage rates. It is crucial to remember that inattention is a property of both the individual and the survey context. Insufficient effort responding or inattention describes a specific type of response behavior, not a specific type of respondent, and can be induced by both individual-level characteristics as well as survey context.

When reporting the results of the survey, researchers need to disclose information about how they chose to measure attention. This includes information about what measures were used, and who was flagged as inattentive, along with researcher choices regarding cut-offs, if applicable. Additionally, if analyzing a subset of attentive respondents, researchers should report results among both the full sample and the attentive sample along with the demographic characteristics of each sample. To facilitate this process, we encourage researchers to include information about how survey attention will be measured and how inattentive respondents will be handled when pre-registering a survey or a survey experiment. Such practices will increase transparency when it comes to measuring attention and help researchers better understand the implications of different measurement strategies while minimizing the threat of p-hacking due to false positives among attentive and inattentive respondents.

Although the literature on survey inattention may be overwhelming, we have witnessed significant progress in recent years and there are some promising directions for future work. First, the field should focus on developing comprehensive recommendations for implementing both direct and indirect measures of attentiveness. This includes questions about the placement and difficulty of direct measures as well as suitable cut-off thresholds for indirect measures. A recent paper by Berinsky et al. (2021) is a good model for this type of work and provides concrete recommendations about the number and the level of difficulty of attention checks that should be used. At the same time, many ongoing debates remain unresolved, leaving room for innovation – especially regarding the question of cut-offs on indirect measures. Second, additional work should explore the relationship between individual-level traits, survey context, and survey inattention using longitudinal data (Paas, Dolnicar and Karlsson 2018; Goldammer et al. 2020; Gummer, Roßmann and Silber 2021). Variations in survey context may interact with individual-level predispositions to induce different response behaviors in distinct survey waves. Third, while evidence suggests that the behavioral consequences of direct measures are limited, the effects of these measures may vary depending on the type of survey question (Hauser and Schwarz 2015; Kung, Kwok and Brown 2018). Moreover, as we call for the use of multiple attention checks within a survey when possible, researchers should also investigate the potential behavioral consequences of both increasing the number and the level of difficulty of these measures. Fourth, scholars should continue leveraging state-of-the-art statistical modeling techniques to explore solutions for setting appropriate cut-off thresholds for indirect measures and removing inattentive respondents. Finally, concerns over survey inattention gained prominence as surveys shifted from primarily phone-based to online formats. In recent years, survey respondents have increasingly completed surveys online using their mobile phones instead of their computers. Phones and computers are fundamentally different devices, necessitating work on how survey inattention manifests on this newer medium.

References

Abbey, James D. and Margaret G. Meloy. 2017. "Attention by design: Using attention checks to detect inattentive respondents and improve data quality." *Journal of Operations Management* 53-56(1):63–70.
URL: *http://doi.wiley.com/10.1016/j.jom.2017.06.001*

Alvarez, R. Michael, Lonna Rae Atkeson, Ines Levin and Yimeng Li. 2019. "Paying Attention to Inattentive Survey Respondents." *Political Analysis* 27(2):145–162.
URL: *https://www.cambridge.org/core/product/identifier/S1047198718000578/type/journal$_a$rticle*

Alvarez, R Michael and Yimeng Li. 2023. "Survey Attention and Self-Reported Political Behavior." *Public Opinion Quarterly* 86(4):793–811.
URL: *https://academic.oup.com/poq/article/86/4/793/7024659*

Anduiza, Eva and Carol Galais. 2016. "Answering Without Reading: IMCs and Strong Satisficing in Online Surveys." *International Journal of Public Opinion Research* p. edw007.
URL: *https://academic.oup.com/ijpor/article-lookup/doi/10.1093/ijpor/edw007*

Arias, Víctor B., L. E. Garrido, C. Jenaro, A. Martínez-Molina and B. Arias. 2020. "A little garbage in, lots of garbage out: Assessing the impact of careless responding in personality survey data." *Behavior Research Methods* 52(6):2489–2505.
URL: *https://link.springer.com/10.3758/s13428-020-01401-8*

Arndt, Aaron D., John B. Ford, Barry J. Babin and Vinh Luong. 2022. "Collecting samples from online services: How to use screeners to improve data quality." *International Journal of Research in Marketing* 39(1):117–133.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S0167811621000203*

Arthur, Winfred, Ellen Hagen and Felix George. 2021. "The Lazy or Dishonest Respondent: Detection and Prevention." *Annual Review of Organizational Psychology and Organizational Behavior* 8(1):105–137.
URL: *https://www.annualreviews.org/doi/10.1146/annurev-orgpsych-012420-055324*

Atsusaka, Yuki and Randolph T. Stevenson. 2023. "A Bias-Corrected Estimator for the Crosswise Model with Inattentive Respondents." *Political Analysis* 31(1):134–148.
URL: *https://www.cambridge.org/core/product/identifier/S1047198721000437/type/journal$_a$rticle*

Barends, Ard J. and Reinout E. de Vries. 2019. "Noncompliant responding: Comparing exclusion criteria in MTurk personality research to improve data quality." *Personality and Individual Differences* 143:84–89.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S0191886919300972*

Beck, Mark F., Anthony D. Albano and Wendy M. Smith. 2019. "Person-Fit as an Index of Inattentive Responding: A Comparison of Methods Using Polytomous Survey Data." *Applied Psychological Measurement* 43(5):374–387.
URL: *http://journals.sagepub.com/doi/10.1177/0146621618798666*

Berinsky, Adam J., Michele F. Margolis and Michael W. Sances. 2014. "Separating the Shirkers from the Workers? Making Sure Respondents Pay Attention on Self-Administered Surveys: SEPARATING THE SHIRKERS FROM THE WORKERS?" *American Journal of Political Science* 58(3):739–753.
URL: *https://onlinelibrary.wiley.com/doi/10.1111/ajps.12081*

Berinsky, Adam J., Michele F. Margolis and Michael W. Sances. 2016. "Can we turn shirkers into workers?" *Journal of Experimental Social Psychology* 66:20–28.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S0022103115001201*

Berinsky, Adam J., Michele F. Margolis, Michael W. Sances and Christopher Warshaw. 2021. "Using screeners to measure respondent attention on self-administered surveys: Which items and how many?" *Political Science Research and Methods* 9(2):430–437.
URL: *https://www.cambridge.org/core/product/identifier/S2049847019000530/type/journal*$_a$*rticle*

Bowling, Nathan A., Jason L. Huang, Caleb B. Bragg, Steve Khazon, Mengqiao Liu and Caitlin E. Blackmore. 2016. "Who cares and who is careless? Insufficient effort responding as a reflection of respondent personality." *Journal of Personality and Social Psychology* 111(2):218–229.
URL: *http://doi.apa.org/getdoi.cfm?doi=10.1037/pspp0000085*

Bowling, Nathan A., Jason L. Huang, Cheyna K. Brower and Caleb B. Bragg. 2021. "The Quick and the Careless: The Construct Validity of Page Time as a Measure of Insufficient Effort Responding to Surveys." *Organizational Research Methods* p. 109442812110565.
URL: *http://journals.sagepub.com/doi/10.1177/10944281211056520*

Breitsohl, Heiko and Corinna Steidelmüller. 2018. "The Impact of Insufficient Effort Responding Detection Methods on Substantive Responses: Results from an Experiment Testing Parameter Invariance: IER DETECTION AND PARAMETER

INVARIANCE." *Applied Psychology* 67(2):284–308.
URL: *https://onlinelibrary.wiley.com/doi/10.1111/apps.12121*

Brühlmann, Florian, Serge Petralito, Lena F. Aeschbach and Klaus Opwis. 2020. "The quality of data collected online: An investigation of careless responding in a crowdsourced sample." *Methods in Psychology* 2:100022.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S2590260120300096*

Buchanan, Erin M. and John E. Scofield. 2018. "Methods to detect low quality data and its implication for psychological research." *Behavior Research Methods* 50(6):2586–2596.
URL: *http://link.springer.com/10.3758/s13428-018-1035-6*

Chmielewski, Michael and Sarah C. Kucker. 2020. "An MTurk Crisis? Shifts in Data Quality and the Impact on Study Results." *Social Psychological and Personality Science* 11(4):464–473.
URL: *http://journals.sagepub.com/doi/10.1177/1948550619875149*

Clifford, Scott and Jennifer Jerit. 2015. "Do Attempts to Improve Respondent Attention Increase Social Desirability Bias?" *Public Opinion Quarterly* 79(3):790–802.
URL: *https://academic.oup.com/poq/article-lookup/doi/10.1093/poq/nfv027*

Conrad, Frederick, Roger Tourangeau, Mick Couper and Chan Zhang. 2017. "Reducing speeding in web surveys by providing immediate feedback." *Survey Research Methods* Vol 11:45–61 Pages. Artwork Size: 45-61 Pages Publisher: European Survey Research Association.
URL: *https://ojs.ub.uni-konstanz.de/srm/article/view/6304*

Curran, Paul G. 2016. "Methods for the detection of carelessly invalid responses in survey data." *Journal of Experimental Social Psychology* 66:4–19.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S0022103115000931*

Curran, Paul G. and Kelsey A. Hauser. 2019. "I'm paid biweekly, just not by leprechauns: Evaluating valid-but-incorrect response rates to attention check items." *Journal of Research in Personality* 82:103849.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S009265661830309X*

DeSimone, Justin A. and P. D. Harms. 2018. "Dirty Data: The Effects of Screening Respondents Who Provide Low-Quality Data in Survey Research." *Journal of Business and Psychology* 33(5):559–577.
URL: *http://link.springer.com/10.1007/s10869-017-9514-9*

DeSimone, Justin A., P. D. Harms and Alice J. DeSimone. 2015. "Best practice recommendations for data screening: DATA SCREENING." *Journal of Organizational Behavior* 36(2):171–181.
URL: *https://onlinelibrary.wiley.com/doi/10.1002/job.1962*

Dogan, Volkan. 2018. Detecting Careless Respondents in Survey Data: Floodlight Detection of Careless Respondents. In *Back to the Future: Using Marketing Basics to Provide Customer Value*, ed. Nina Krey and Patricia Rossi. Cham: Springer International Publishing pp. 443–449. Series Title: Developments in Marketing Science: Proceedings of the Academy of Marketing Science.
URL: *http://link.springer.com/10.1007/978-3-319-66023-3*$_1$48

Dunn, Alexandra M., Eric D. Heggestad, Linda R. Shanock and Nels Theilgard. 2018. "Intra-individual Response Variability as an Indicator of Insufficient Effort Responding: Comparison to Other Indicators and Relationships with Individual Differences." *Journal of Business and Psychology* 33(1):105–121.
URL: *http://link.springer.com/10.1007/s10869-016-9479-0*

Francavilla, Nicole M., Adam W. Meade and Amanda L. Young. 2019. "Social Interaction and Internet-Based Surveys: Examining the Effects of Virtual and In-Person Proctors on Careless Response: Proctors and Careless Response." *Applied Psychology* 68(2):223–249.
URL: *https://onlinelibrary.wiley.com/doi/10.1111/apps.12159*

Goldammer, Philippe, Hubert Annen, Peter Lucas Stöckli and Klaus Jonas. 2020. "Careless responding in questionnaire measures: Detection, impact, and remedies." *The Leadership Quarterly* 31(4):101384.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S1048984320300114*

Grau, Ina, Christine Ebbeler and Rainer Banse. 2019. "Cultural Differences in Careless Responding." *Journal of Cross-Cultural Psychology* 50(3):336–357.
URL: *http://journals.sagepub.com/doi/10.1177/0022022119827379*

Greszki, Robert, Marco Meyer and Harald Schoen. 2015. "Exploring the Effects of Removing "Too Fast" Responses and Respondents from Web Surveys." *Public Opinion Quarterly* 79(2):471–503.
URL: *https://academic.oup.com/poq/article-lookup/doi/10.1093/poq/nfu058*

Gummer, Tobias, Joss Roßmann and Henning Silber. 2021. "Using Instructed Response Items as Attention Checks in Web Surveys: Properties and Implementa-

tion." *Sociological Methods & Research* 50(1):238–264.
URL: *http://journals.sagepub.com/doi/10.1177/0049124118769083*

Harden, Jeffrey J., Anand E. Sokhey and Katherine L. Runge. 2019. "Accounting for Noncompliance in Survey Experiments." *Journal of Experimental Political Science* 6(3):199–202.
URL: *https://www.cambridge.org/core/product/identifier/S2052263019000137/type/journal$_a$rticle*

Hauser, David J. and Norbert Schwarz. 2015. "It's a Trap! Instructional Manipulation Checks Prompt Systematic Thinking on "Tricky" Tasks." *SAGE Open* 5(2):215824401558461.
URL: *http://journals.sagepub.com/doi/10.1177/2158244015584617*

Hauser, David J. and Norbert Schwarz. 2016. "Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants." *Behavior Research Methods* 48(1):400–407.
URL: *http://link.springer.com/10.3758/s13428-015-0578-z*

Hong, Maxwell, Jeffrey T. Steedle and Ying Cheng. 2020. "Methods of Detecting Insufficient Effort Responding: Comparisons and Practical Recommendations." *Educational and Psychological Measurement* 80(2):312–345.
URL: *http://journals.sagepub.com/doi/10.1177/0013164419865316*

Huang, Jason L., Nathan A. Bowling, Mengqiao Liu and Yuhui Li. 2015. "Detecting Insufficient Effort Responding with an Infrequency Scale: Evaluating Validity and Participant Reactions." *Journal of Business and Psychology* 30(2):299–311.
URL: *http://link.springer.com/10.1007/s10869-014-9357-6*

Huang, Jason L., Paul G. Curran, Jessica Keeney, Elizabeth M. Poposki and Richard P. DeShon. 2012. "Detecting and Deterring Insufficient Effort Responding to Surveys." *Journal of Business and Psychology* 27(1):99–114.
URL: *http://link.springer.com/10.1007/s10869-011-9231-8*

Johnson, John A. 2005. "Ascertaining the validity of individual protocols from Web-based personality inventories." *Journal of Research in Personality* 39(1):103–129.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S0092656604000856*

Kane, John V. and Jason Barabas. 2019. "No Harm in Checking: Using Factual Manipulation Checks to Assess Attentiveness in Experiments." *American Journal of Political Science* 63(1):234–249.
URL: *https://onlinelibrary.wiley.com/doi/10.1111/ajps.12396*

Kane, John V., Yamil R. Velez and Jason Barabas. 2023. "Analyze the attentive and bypass bias: mock vignette checks in survey experiments." *Political Science Research and Methods* 11(2):293–310.
URL: *https://www.cambridge.org/core/product/identifier/S2049847023000031/type/journal_article*

Kim, Dale S., Connor J. McCabe, Brianna L. Yamasaki, Kristine A. Louie and Kevin M. King. 2018. "Detecting random responders with infrequency scales using an error-balancing threshold." *Behavior Research Methods* 50(5):1960–1970.
URL: *http://link.springer.com/10.3758/s13428-017-0964-9*

Krosnick, Jon A. 1991. "Response strategies for coping with the cognitive demands of attitude measures in surveys." *Applied Cognitive Psychology* 5(3):213–236.
URL: *https://onlinelibrary.wiley.com/doi/10.1002/acp.2350050305*

Kung, Franki Y.H., Navio Kwok and Douglas J. Brown. 2018. "Are Attention Check Questions a Threat to Scale Validity?: ATTENTION CHECKS AND SCALE VALIDITY." *Applied Psychology* 67(2):264–283.
URL: *https://onlinelibrary.wiley.com/doi/10.1111/apps.12108*

Ladini, Riccardo. 2022. "Assessing general attentiveness to online panel surveys: the use of instructional manipulation checks." *International Journal of Social Research Methodology* 25(2):233–246.
URL: *https://www.tandfonline.com/doi/full/10.1080/13645579.2021.1877948*

Leiner, Dominik Johannes. 2019. "Too Fast, too Straight, too Weird: Non-Reactive Indicators for Meaningless Data in Internet Surveys." *Survey Research Methods* pp. 229–248 Pages. Artwork Size: 229-248 Pages Publisher: Survey Research Methods.
URL: *https://ojs.ub.uni-konstanz.de/srm/article/view/7403*

Liu, Mingnan and Laura Wronski. 2018. "Trap questions in online surveys: Results from three web survey experiments." *International Journal of Market Research* 60(1):32–49.
URL: *http://journals.sagepub.com/doi/10.1177/1470785317744856*

Malone, Trey and Jayson L Lusk. 2018. "A simple diagnostic measure of inattention bias in discrete choice models." *European Review of Agricultural Economics* 45(3):455–462.
URL: *https://academic.oup.com/erae/article/45/3/455/4953441*

Mancosu, Moreno, Riccardo Ladini and Cristiano Vezzoni. 2019. "'Short is Better'. Evaluating the Attentiveness of Online Respondents Through Screener Questions in a Real Survey Environment." *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique* 141(1):30–45.
URL: *http://journals.sagepub.com/doi/10.1177/0759106318812788*

Maniaci, Michael R. and Ronald D. Rogge. 2014. "Caring about carelessness: Participant inattention and its effects on research." *Journal of Research in Personality* 48:61–83.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S009265661300127X*

Matjašič, Miha, Vasja Vehovar and Katja Lozar Manfreda. 2018. "Web Survey Paradata on Response Time Outliers: A Systematic Literature Review." *Metodološki zvezki* .

Matjašič, Miha, Vasja Vehovas and Anže Sendelbah. 2021. "Combining Response Times and Response Quality Indicators to Identify Speeders with Low Response Quality in Web Surveys.".

Meade, Adam W. and S. Bartholomew Craig. 2012. "Identifying careless responses in survey data." *Psychological Methods* 17(3):437–455.
URL: *http://doi.apa.org/getdoi.cfm?doi=10.1037/a0028085*

Montgomery, Jacob, Brendan Nyhan and Michelle Torres. 2018. "How Conditioning on Posttreatment Variables Can Ruin Your Experiment and What to Do about It." *American Journal of Political Science* 62(3):760–775.

Morren, Meike and Leonard J Paas. 2020. "Short and Long Instructional Manipulation Checks: What Do They Measure?" *International Journal of Public Opinion Research* 32(4):790–800.
URL: *https://academic.oup.com/ijpor/article/32/4/790/5680365*

Niessen, A. Susan M., Rob R. Meijer and Jorge N. Tendeiro. 2016. "Detecting careless respondents in web-based questionnaires: Which method to use?" *Journal of Research in Personality* 63:1–11.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S0092656616300344*

Oppenheimer, Daniel M., Tom Meyvis and Nicolas Davidenko. 2009. "Instructional manipulation checks: Detecting satisficing to increase statistical power." *Journal of Experimental Social Psychology* 45(4):867–872.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S0022103109000766*

Paas, Leonard J. and Meike Morren. 2018. "PLease do not answer if you are reading this: respondent attention in online panels." *Marketing Letters* 29(1):13–21.
URL: *http://link.springer.com/10.1007/s11002-018-9448-7*

Paas, Leonard J., Sara Dolnicar and Logi Karlsson. 2018. "Instructional Manipulation Checks: A longitudinal analysis with implications for MTurk." *International Journal of Research in Marketing* 35(2):258–269.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S016781161830003X*

Palaniappan, Kavitha and Isabel Yung Shen Kum. 2019. "Underlying Causes behind Research Study Participants' Careless and Biased Responses in the Field of Sciences." *Current Psychology* 38(6):1737–1747.
URL: *http://link.springer.com/10.1007/s12144-017-9733-2*

Patton, Jeffrey M., Ying Cheng, Maxwell Hong and Qi Diao. 2019. "Detection and Treatment of Careless Responses to Improve Item Parameter Estimation." *Journal of Educational and Behavioral Statistics* 44(3):309–341.
URL: *http://journals.sagepub.com/doi/10.3102/1076998618825116*

Peyton, Kyle, Gregory A Huber and Alexander Coppock. 2022. "The Generalizability of Online Experiments Conducted During the COVID-19 Pandemic." *Journal of Experimental Political Science* 9(3):379 – 394.

Pyo, Jimin and Michael G. Maxfield. 2021. "Cognitive Effects of Inattentive Responding in an MTurk Sample." *Social Science Quarterly* 102(4):2020–2039.
URL: *https://onlinelibrary.wiley.com/doi/10.1111/ssqu.12954*

Read, Blair, Lukas Wolters and Adam J. Berinsky. 2022. "Racing the Clock: Using Response Time as a Proxy for Attentiveness on Self-Administered Surveys." *Political Analysis* 30(4):550–569.
URL: *https://www.cambridge.org/core/product/identifier/S1047198721000322/type/journal$_a$rticle*

Reyes, Denise L. 2022. "Combatting carelessness: Can placement of quality check items help reduce careless responses?" *Current Psychology* 41(10):6858–6866.
URL: *https://link.springer.com/10.1007/s12144-020-01183-4*

Rios, Joseph A., Hongwen Guo, Liyang Mao and Ou Lydia Liu. 2017. "Evaluating the Impact of Careless Responding on Aggregated-Scores: To Filter Unmotivated Examinees or Not?" *International Journal of Testing* 17(1):74–104.
URL: *https://www.tandfonline.com/doi/full/10.1080/15305058.2016.1231193*

Schroeders, Ulrich, Christoph Schmidt and Timo Gnambs. 2022. "Detecting Careless Responding in Survey Data Using Stochastic Gradient Boosting." *Educational and Psychological Measurement* 82(1):29–56.
URL: *http://journals.sagepub.com/doi/10.1177/00131644211004708*

Shamon, Hawal and Carl Clemens Berning. 2020. "Attention Check Items and Instructions in Online Surveys: Boon or Bane for Data Quality?" *Survey Research Methods* pp. 55–77 Pages. Artwork Size: 55-77 Pages Publisher: Survey Research Methods.
URL: *https://ojs.ub.uni-konstanz.de/srm/article/view/7374*

Silber, Henning, Daniel Danner and Beatrice Rammstedt. 2019. "The impact of respondent attentiveness on reliability and validity." *International Journal of Social Research Methodology* 22(2):153–164.
URL: *https://www.tandfonline.com/doi/full/10.1080/13645579.2018.1507378*

Ternovski, John, Lilla Orr, Joshua Kalla and Peter Aronow. 2022. "A Note on Increases in Inattentive Online Survey-Takers Since 2020." *Journal of Quantitative Description: Digital Media* 2.
URL: *https://journalqd.org/article/view/2985*

Thomas, Kyle A. and Scott Clifford. 2017. "Validity and Mechanical Turk: An assessment of exclusion methods and interactive experiments." *Computers in Human Behavior* 77:184–197.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S074756321730506X*

Toich, Margaret J., Elizabeth Schutt and David M. Fisher. 2022. "Do you get what you pay for? Preventing insufficient effort responding in MTurk and student samples." *Applied Psychology* 71(2):640–661.
URL: *https://onlinelibrary.wiley.com/doi/10.1111/apps.12344*

Tourangeau, R., L. J. Rips and K. Rasinski. 2000. *The psychology of survey response.* Cambridge University Press.

Tyler, Matthew, Justin Grimmer and Sean Westwood. 2022. "A Statistical Framework to Engage the Problem of Disengaged Survey Respondents.".

Ulitzsch, Esther, Steffi Pohl, Lale Khorramdel, Ulf Kroehne and Matthias von Davier. 2022. "A Response-Time-Based Latent Response Mixture Model for Identifying and Modeling Careless and Insufficient Effort Responding in Survey Data." *Psychometrika* 87(2):593–619.
URL: *https://link.springer.com/10.1007/s11336-021-09817-7*

Ward, M.K. and Adam W. Meade. 2018. "Applying Social Psychology to Prevent Careless Responding during Online Surveys: PREVENT CARELESS RESPONDING." *Applied Psychology* 67(2):231–263.
URL: *https://onlinelibrary.wiley.com/doi/10.1111/apps.12118*

Ward, M.K. and Adam W. Meade. 2023. "Dealing with Careless Responding in Survey Data: Prevention, Identification, and Recommended Best Practices." *Annual Review of Psychology* 74(1):577–596.
URL: *https://www.annualreviews.org/doi/10.1146/annurev-psych-040422-045007*

Ward, M.K. and Samuel B. Pond. 2015. "Using virtual presence and survey instructions to minimize careless responding on Internet-based surveys." *Computers in Human Behavior* 48:554–568.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S0747563215000989*

Weathers, Danny and Ahmet Bardakci. 2015. "Can response variance effectively identify careless respondents to multi-item, unidimensional scales?" *Journal of Marketing Analytics* 3(2):96–107.
URL: *http://link.springer.com/10.1057/jma.2015.6*

Wood, Dustin, P. D. Harms, Graham H. Lowman and Justin A. DeSimone. 2017. "Response Speed and Response Consistency as Mutually Validating Indicators of Data Quality in Online Samples." *Social Psychological and Personality Science* 8(4):454–464.
URL: *http://journals.sagepub.com/doi/10.1177/1948550617703168*

Yu, Xiaofeng and Ying Cheng. 2019. "A change-point analysis procedure based on weighted residuals to detect back random responding." *Psychological Methods* 24(5):658–674.
URL: *http://doi.apa.org/getdoi.cfm?doi=10.1037/met0000212*