Disentangling Administration Errors From Scale Development Errors in Survey Research

Dan Friesner
North Dakota State University
University of Akron

Carl S. Bozman
Gonzaga University

Matthew McPherson
Gonzaga University

In a recent manuscript, Friesner, et al. (2023) used the concept of information entropy to assess the quantity of information in survey responses. They demonstrate how assessments of the quantity of information can be used to identify possible errors in a survey's administration. A major limitation of their methodology is that it assumes the survey items used to elicit consumer preferences were created appropriately and contained a meaningful quantity of information. The current study addresses this limitation by incorporating a methodology developed by Friesner et al. (2021) into the Friesner et al. (2023) methodology. The combined methodology is applied to the same data studied in both Friesner et al. (2021) and Friesner et al. (2023), which allows for a direct comparison of the quantity of information gained/lost from survey administration versus scale development. The results indicate that the survey used in the empirical application exhibits flaws in both scale design and survey administration.

Keywords: survey design, information entropy, scale development, Tobit regression

INTRODUCTION

Surveys are among the most popular means available to social scientists to collect information on self-reported human behaviors. This is especially true in the field of marketing, where surveys are an essential means by which firms assess various aspects of customer satisfaction and brand loyalty (Batra et al., 2001; Beatty et al., 1985; Bradley et al., 2015). For specific service-related industries – especially those firms which offer entertainment events - several months of planning are required to execute the event successfully. And as the number of participants increases (if the event is initially successful), the more varied are the consumer perceptions of the experience (Demming, 1944; Johnson et al., 2006; Suchman, 1962; Wikman and Warneryd, 1990; Wiseman, 1972). The increased variation in perceptions makes it difficult to measure the quality of the participant's experience, and to adequately use that information to improve future experiences. Adequately capturing consumer preferences becomes more challenging and critical to firms' long run success in service-related industries.

A critical component of assessing consumers' preferences is eliciting (maximally) informative survey responses. Greater quantities of information that can be gleaned from such surveys not only allows for more information to guide process improvement but also make optimal use of time and effort allocated to survey development and implementation (Draugalis et al., 2008; Wing et al., 2018). For many large-scale survey projects, researchers often conduct a pilot study using the survey, and use the data to assess whether further revisions are necessary to ensure that the survey and its administration procedures produce maximally informative data (Abd Gani et al., 2020; González-Cabrera et al., 2020; van Teijlingen & Hundley, 2001; van Teijlingen et al., 2001). Identifying those consumers who give more informative responses also allows survey researchers to proactively identify and address sources of bias that commonly occur in survey data (Demming, 1944; Fenner et al., 2020; Hendra and Hill, 2019; Marquis et al., 1986). Suppose different types of consumers systematically give different quantities of information in their responses. In that case, the researcher may choose to adopt a stratified sampling design (a priori), adopt methods of analysis (such as generalized methods of moments) to weigh the more informative responses more heavily than other responses when assessing survey data (ex post), or a combination of the two (Imbens and Lancaster, 1996; Tripathi, 2011). The current manuscript assumes the existence of such a pilot study (or a survey that is administered repeatedly over time), and that data are available from this pilot study to further assess and refine the survey.

A recent strand of the survey research literature attempts to assess the "quantity" of information contained in survey responses and use it to improve the design and administration of surveys. In the context of the literature, the "quantity" of information compares the actual distribution of survey responses (whether for a single survey item, aggregated across respondents, or multiple (related) survey items aggregated across a single respondent) to the researcher's a priori expectation for this distribution in the absence of (or before) a survey's administration (Dahl and Osteras, 2010). The more consistent the two distributions, the lower the quantity of information contained in the survey because there is little to be gained from administering the survey in the first place. The more the two distributions diverge, the greater the quantity of information available to be gleaned from administering a survey. As a result, it is worthwhile to apply descriptive and inferential statistical methods to the data (which is referred to as assessing the "quality" of information or the magnitudes of the inter-relationships that exist in the data).

Friesner et al. (2023) used the concept of information entropy (Shannon, 1948; Janes, 1957, 1982; Golan et al., 1996a,b; Golan, 2006; Dahl and Osteras, 2010) to assess the quantity of information in survey responses. They demonstrate how assessment of the quantity of information can be used to identify possible errors in a survey's administration, especially when specific groups of respondents give fundamentally different quantities of information when completing the same survey items. A major limitation of their analysis is that it assumes the items and response scales used to elicit consumer preferences were created appropriately and contain a meaningful quantity of information.

Friesner et al. (2021) use the same literature to demonstrate how the quantity of information can inform the aggregation of related survey items into scales. Items that do not contain a meaningful (and statistically significant) quantity of information may (if additional criteria are met) be eliminated from the survey to reduce its length (which helps to improve response rates) while maintaining the same quantity of information. A major limitation of this study is that it assumes that the survey was administered appropriately and that all groups of respondents give the same quantity of information in their responses.

Considered cumulatively, each of the studies mentioned above is limited in that it addresses only one issue, holding constant, or assuming away, the possibility of the other issue. That is, Friesner et al. (2021) assume that the survey was administered appropriately, and that all groups of respondents give the same quantity of information in their responses. Friesner et al. (2023) assess whether different groups of respondents give the same quantities of information in their responses to the same items, assuming that all survey items and scales being assessed contain a reasonable quantity of information. These assumptions have important implications for survey design. For example, suppose a researcher adapts a survey by altering the content in a scale. In that case, it may change who (or which groups) respond to the survey and provide different quantities of information. Concomitantly, suppose a researcher alters the survey's administration based on groups that provide more or less informative responses. In that case, it may change the information in the survey items and the scales constructed based on those items.

This study proposes an econometric methodology that addresses both of these issues simultaneously.[1] It is applied to the same data studied in both Friesner et al. (2021) and Friesner et al. (2023). This allows for directly comparing the quantity of information gained/lost from survey administration versus scale development.

The reminder of this paper proceeds in five steps. The next step summarizes the Friesner et al. (2021) and Friesner et al. (2023) methodologies. A new methodology is then developed, which combines both of the former methodologies. The third section describes the data used to implement the combined methodology. More specifically, the data are drawn from a customer satisfaction survey administered via non-stratified sampling at the 2018 Hoopfest Basketball Tournament, which was used in both Friesner et al. (2021) and Friesner et al. (2023). The fourth section contains the empirical results. The conclusion summarizes the study findings, discusses implications for both survey design and public policy and identifies possible directions for future research in this area of inquiry.

METHODOLOGY

Assumptions and Hypotheses

The current study, as well as those of Friesner et al. (2021) and Friesner et al. (2023), operate under several assumptions. First, we assume the existence of a pilot study or a survey that has been administered repeatedly over time, whose data can be used to assess the current state of the survey and identify additional refinements to that survey.

Friesner et al. (2021) and Friesner et al. (2023) studies operate under a general assumption of ignorance about the data-generating process being assessed. More specifically, these studies assume that a survey is designed and administered such the distribution of responses for a given survey item or scale (collected over a set of individuals), or a given individual (collected over a set of survey items or scales), is uniform. Uniform responses minimize the likelihood of leniency, common method variance and/or framing biases (among other design issues), which reduce the sensitivity of the survey item(s) or scale(s) being analyzed (Ballard, 2019; Entman, 2007; Jordan and Troth, 2020). Under this assumption, information entropy can be characterized as:

$$H(p) = -\sum_{k=1}^{K} p_k log_2(p_k) \tag{1}$$

where k indicates one of k = 1,…,K possible responses to a given survey item and p_k is the proportion of responses that fall into category k for a given sample of data. It is further assumed that all proportions are proper; that is: $0 \leq p_k \leq 1$ and $\sum_{k=1}^{K} p_k = 1$. It is straightforward to show that information entropy is maximized when $p_k = 1/k$ for every k (i.e., when the proportions are uniform).

For any value of k items, entropy H(p) is calculated for each category k. Because the absolute magnitude of the information entropy metric lacks an obvious baseline against which it can be interpreted, it is typically normalized by dividing H(p) by the maximum possible value for entropy, achieved when each p_k are uniformly distributed. The result is a normalized entropy that can be expressed as a proportion, with higher values denoting greater captured entropy and a lower quantity of information contained in the metric. Similarly, lower normalized values indicate less captured entropy and higher quantities of information.

When applied specifically to survey design and administration, we are interested in evaluating whether, over a series of similar survey items, individual respondents give more or less information content than other types of respondents. This requires an adaptation of (1). To adapt the base entropy measure, define i=1,…,n as observations/respondents, and l = 1,..,L as the number of survey items in a given scale. We define k = 1,…,K as the number of options a respondent can give when responding to a specific survey item (i.e., yes/no questions yield k = 2). Additionally, we assume that all proportions continue to be proper. This allows for a more nuanced definition of each proportion:

$$p_{ik} = \frac{\sum_{l=1}^{L} D_{kli}}{L_i} \tag{2}$$

Equation (1) can also be re-specified as:

$$H_i(p) = -\sum_{k=1}^{K} p_{ki} log_2(p_{ki}) \tag{1b}$$

And, as noted previously, the new entropy measure (1b) can be expressed as a percentage of information captured in the scale by each respondent:

$$Normalized\ H_i(p) = \frac{H_i(p)}{Max(H_i(p))} \tag{3}$$

Note that, when entropy is maximized, the quantity of information is minimized. Thus, when equation (3) is closer to unity, the survey item(s) in question contains less information. When it is closer to zero, a greater proportion of available information is captured by the survey item(s) in question. In other words, equation (3) represents the proportion of the quantity of information that is *not* captured by the survey item(s) in question. We use this notation because it was used by Friesner et al. (2016) and Friesner et al. (2023). However, future researchers may wish to subtract this proportion from one, allowing for a much more intuitive measure of the quantity of information (i.e., the quantity captured by the survey item(s) in question).

The study's primary hypotheses follow directly from the assumption of ignorance and the use of information entropy to characterize the quantity of information. More specifically, this study operates under the general assumption that the design and administration of the survey used to generate data are reasonably well-designed. Within the context of this study, as well as Friesner et al. (2021) and Friesner et al. (2023), the null and alternative hypotheses are:

$\boldsymbol{H_0}$***:*** *No mean differences in information entropy exist across different types of consumers with different incentives and/or the construction of a scale.*

$\boldsymbol{H_A}$***:*** *Not* H_0.

Combining the Friesner et al. (2021) and Friesner et al. (2023) Methodologies

Friesner et al. (2021) suggest using backward selection to characterize the quantity of information in a scale. They specifically postulate a four-step process. First, given the existence of a survey scale (and a well-defined, appropriate survey administration process), equation (3) can be constructed using all possible survey items that may contribute to the scale. This can be considered as a variable, which Friesner et al. (2021) define as $Normalized_{All}\ H_i(p)$. The next step in their process is to calculate equation (3) using the same scale. However, eliminating those z ≥ 1 survey items from the scale that the research suspects may not contain a positive quantity of information. Friesner et al. (2021) define this variable as $Normalized_{Reduced}\ H_i(p)$. Third, the researcher should assess the difference between the normalized entropy measure calculated in Step 1 versus Step 2.[2] That is:

$$DifNormalized\ H_i(p) = Normalized_{All}\ H_i(p) - Normalized_{Reduced}\ H_i(p) \tag{4}$$

The variable DifNormalized $H_i(p)$ characterizes the amount of information contained in the z ≥ 1 survey items omitted from the scale. Lastly, parametric or non-parametric hypothesis tests can be applied to DifNormalized $H_i(p)$to assess whether the omitted items contain a quantity of information that is statistically different from a hypothesized amount. In most practical applications, the researcher assesses whether DifNormalized $H_i(p)$ is significantly different from zero.

Friesner et al. (2023) assume that the scale and the items that comprise it are constructed appropriately, and instead seek to determine whether specific groups of respondents (or the characteristics of specific types of respondents) provide greater or lesser quantities of information than other types of respondents. To do so, they calculate Normalized $H_i(p)$ and use it as the dependent variable in a reduced form, linear in parameters regression analysis, where the independent variables in the regression characterize those respondent-specific characteristics. Because Normalized $H_i(p)$ is bound to the unit interval, the authors utilized a Tobit model with two sided censoring:

$$Normalized\ H_i(p) = \alpha + \sum_{q=1}^{Q} \beta^q X_i^q + \varepsilon_i \tag{5}$$

where X^q is one of Q variables characterizing a specific respondent characteristic, α represents an estimated intercept, β^q is one of q = 1,..,Q slope parameters to be estimated, and ε_i is a random error term (over i =1,…,n observations). Under the null hypothesis that the survey is appropriately administered, β^q (whether assessed individually or jointly across or a subset of the Q parameter estimates) should not be significantly different from zero.

Disentangling survey design errors from survey administration errors proceeds logically using a stepwise fashion. First, simple t-tests can be used to assess whether (for a given set of z ≥ 1 survey items omitted from the scale) DifNormalized $H_i(p)$ is significantly different from zero. We note that an equivalent means to assess this hypothesis would be to conduct matched sample t-tests assessing the difference between $Normalized_{All}\ H_i(p)$ and $Normalized_{Reduced}\ H_i(p)$.[3] Next, one can adapt the Friesner et al. (2023) analysis by specifying the following Tobit regression:

$$DifNormalizedH_i(p) = \alpha + \sum_{q=1}^{Q} \beta^q X_i^q + \varepsilon_i \tag{6}$$

As before, under the null hypothesis that the survey is appropriately administered, β^q (whether assessed individually or jointly across or a subset of the Q parameter estimates) should not be significantly different from zero.

To combined the two methodologies, we employ them sequentially. First, we employ Friesner et al.'s (2021) methodology to determine whether a positive quantity of information exists in a particular survey item, as characterized by DifNormalized $H_i(p)$, and without holding any specific respondent-specific characteristics constant. This provides a general baseline against which to assess the information quantity in the methodology's second step.

Next, we estimate equation (6) using the DifNormalized $H_i(p)$ as the dependent variable, both overall and for each of the individual survey questions being analyzed. Several hypothesis tests are implemented to assess the study's null hypothesis. First, the statistical significance of the model's intercept can be used to determine whether, after controlling for respondent-specific characteristics and the potential censoring of the dependent variable, a positive quantity of information exists in a given survey item, as characterized by DifNormalized $H_i(p)$, that is unique to that survey item. In the absence of data at the points of censoring (in this model, either values of zero or one), and when all other regressors are statistically insignificant, the estimated intercept represents the quantity of information contained in a survey item. In all other cases, the statistical significance of the intercept (qualitatively) indicates whether a unique quantity of information is contained in a variable that is not systematically attributable to either the censoring of the dependent variable or the effects of the other model regressors. The former is simply an artifact of using maximum entropy as a baseline for evaluation (which, in turn, is an artifact of the underlying assumption of ignorance), while the latter (if present) is indicative of possible survey administration issues. Because the researcher has no prior information about whether the Tobit disturbance term or the other model covariates will be statistically significant, we do not attempt to interpret the magnitude of the estimated intercept. Instead, we take the economical approach of only assessing the statistical significance of the model intercept. If the intercept is not statistically different from zero, the evaluated survey item does not contain a unique, positive quantity of information, holding constant the possible censoring of the dependent variable and the effects of the other regressors.

Next, the statistical significance of the regressors, whether considered jointly using chi-square tests or individually using t-tests, can be used to assess whether respondent-specific charactistics influence the quantity of information (DifNormalized $H_i(p)$). Holding constant the overall quantity of information in a specific survey item, if the parameter estimates for the regressors are statistically significant, then survey administration errors may exist in the data, even after controlling for the unique quantity of information in the survey item overall. In this way, it is possible to disentangle the assessment of survey design issues from survey administration issues.

DATA

Data are drawn from customer satisfaction surveys at the 2018 Hoopfest Basketball Tournament in Spokane, WA. Hoopfest is the largest 3-on-3 amateur basketball tournament in the world. Attendance at the event is typically estimated at approximately 250,000 individuals (Schnell, 2014). Hoopfest routinely uses surveys administered at the event to assess customer satisfaction among all attendees (players, spectators, etc.). The same base survey has been in use since approximately 2006, and only minor changes are typically made in the survey each year. As such, the survey has a design and administration process vetted in previous empirical studies and whose sample size should be adequately powered to conduct complex statistical analyses (Dillman, 2000, pp. 207). More information on the design and administration of this survey can be found in Bozman et al. (2010), Kurpis et al. (2010), and Friesner et al. (2016).

Related to the current analysis, the data are also interesting because the survey provided a unique context that allows the methodology to potentially disentangle survey administration issues from scale development issues and triangulate against the literature using similar data. The survey contains measures of incentives to attend, including economic and personal incentives, which allows for a robust set of variables that can be used to characterize respondent-specific characteristics. Additionally, respondents were asked to complete a Carre, et al. (2013) and Jolliffe and Farrington (2006) two-item Basic Empathy Scale (BES). Six additional survey items were included in the Hoopfest survey as proposed additions to the empathy scale. These six items attempted to characterize the various aspects of overall empathy. Two items were proposed to assess cognitive empathy ("I recall my personal experience when I observe someone else in a similar situation" and "I almost always understand the motives behind the actions of another person"), two items attempted to assess affective empathy ("I feel happy when I see smiles on other people's faces" and "I am sad when I observe someone in distress") and two items were proposed to assess behavioral empathy ("I get agitated when I see someone in distress" and "I help others when I see they need help"). All six proposed items use the same five-point response scale as the original two BES items. The original BES scale, combined with these six potential additions, yields an interesting natural experiment to assess both survey design and administration issues. Friesner et al. (2023) and Friesner et al. (2021) utilize data drawn from the 2018 Hoopfest survey. The current analysis also uses this same data source to ensure comparability with these studies. Additionally, because the former study utilizes data across a larger number of variables (i.e., those used as regressors), it has fewer observations than the latter study. Because this study uses the same regressors as in Friesner et al. (2023), it also uses the same number of observations as in that study.

Hoopfest personnel managed and approved the survey's construction, collected the data, and subsequently provided a de-identified dataset to the authors. Because the data were not directly collected by the authors, and the authors cannot identify respondents, the institutional review boards (IRBs) of the author's institutions (at the time the research was conducted) do not classify research with this data as human subjects research and did not require IRB approval.

Friesner et al. (2023) noted that the survey is designed to be adequately statistically powered. Given the properties of this particular event, (Dillman, 2000, pp. 207) suggests a target sample size of 385 observations. Hoopfest administrators randomly identified 500 potential respondents, of which 437 agreed to complete the survey (87% response rate). After eliminating missing or mis-measured data in each of the selected variables to be used in the analysis, Friesner et al. (2023) obtained a working data set with 336 observations (67% completion rate).

RESULTS

Table 1 contains the variable names, variable descriptions, and descriptive statistics for each variable used in the analysis. Panel A contains the names, definitions, and descriptive statistics for each of the dependent variables used in the analysis, which consist of the $DifNormalizedH_i(p)$ variables for each of the six proposed survey items. Additionally, Panel A contains $DifNormalizedH_i(p)$ variables for the traditional, two-item BES scale (BESBaseEN) and the BES scale with all eight survey items (BESAllEN). Both of these variables provide a baseline against which the individual survey items are assessed. Among the six proposed items, two align with a specific aspect of empathy (cognitive, affective, and behavioral). Three additional variables were created that characterize the quantity of information (normalized to the unit interval) across both proposed items in each aspect of empathy. These are referred to as BESQCogEN (for cognitive empathy), BESQAffEN (for affective empahy), and BESQBehEN (for behavioral empathy, respectively). Panel B contains the names, definitions, and descriptive statistics for each respondent specific characteristics, which are used as regressors in the Tobit regression analyses. Because the information in Panel B is identical to that presented in Friesner et al. (2023), we refer the interested reader to that study for a detailed discussion of these variables. Because the information contained in Panel A is based on a smaller set of observations compared to Friesner et al. (2021), the information in Panel A will deviate slightly from what is reported in that manuscript. However, we note that all variable names and definitions are identical to those from Friesner et al. (2021). Additionally, and not surprisingly, the descriptive statistics reported in Panel A are very close (but not identical) in magnitude to those described in Friesner et al. (2021).

TABLE 1
DESCRIPTIVE STATISTICS

Variable	**Description**	**Mean**	**Median**	**Std. Dev.**
Dependent Variables (Each Represents $DifNormalizedH_i(p)$)				
BESAllEN	Normalized Entropy Calculation - All Emotional Connection Scales	0.496	0.441	0.266
BESBaseEN	Normalized Entropy Calculation - BES Emotional Connection Items (“I get caught up in other people’s feelings easily” & “I can often understand how people are feeling even before they tell me”)	0.378	0.000	0.486
BESQCEN	Normalized Entropy Calculation - All Emotional Connection Items Except Item C (“I recall my personal experience when I observe someone else in a similar situation”)	0.502	0.406	0.266
BESQDEN	Normalized Entropy Calculation - All Emotional Connection Items Except Item D (“I get agitated when I see someone in distress”)	0.513	0.406	0.259
BESQEEN	Normalized Entropy Calculation - All Emotional Connection Items Except Item E (“I feel happy when I see smiles on other people’s faces”)	0.515	0.406	0.262
BESQFEN	Normalized Entropy Calculation - All Emotional Connection Items Except Item F (“I almost	0.521	0.406	0.266

	always understand the motives behind the actions of another person”)			
BESQGEN	Normalized Entropy Calculation - All Emotional Connection Items Except Item G (“I help others when I see they need help”)	0.503	0.406	0.261
BESQHEN	Normalized Entropy Calculation - All Emotional Connection Items Except Item H (“I am sad when I observe someone in distress”)	0.498	0.406	0.266
BESQCogEN	Normalized Entropy Calculation - All Emotional Connection Items Except New Cognitive Items (Items C&F)	0.530	0.461	0.266
BESQBehEN	Normalized Entropy Calculation - All Emotional Connection Items Except New Behavioral Items (Items D&G)	0.519	0.461	0.255
BESQAffEN	Normalized Entropy Calculation - All Emotional Connection Items Except New Affective Items (Items E&H)	0.524	0.461	0.261
Covariates				
AGE	Respondent Age in Years	34.497	31.000	13.677
FEMALE	Binary Variable Identifying Female Respondents	0.500		
PLAY	Binary Variable Identifying Respondents Attending Hoopfest as Players	0.387		
WATCH	Binary Variable Identifying Respondents Attending Hoopfest to Watch Games	0.574		
VOLUN	Binary Variable Identifying Respondents Attending Hoopfest as a Volunteer	0.033		
OTHROLE	Binary Variable Identifying Respondents Attending Hoopfest for Another Reason	0.006		
HRS05	Binary Variable Identifying Respondents Attending Hoopfest for 0-5 Hours	0.250		
HRS611	Binary Variable Identifying Respondents Attending Hoopfest for 6-11 Hours	0.676		
HRS12U	Binary Variable Identifying Respondents Attending Hoopfest for 12 or More Hours	0.074		
NPEOPLE	Number of Individuals Who Attended Hoopfest with the Respondent	4.330	3.000	4.132
OVERNGT	Binary Variable Identifying Respondents Who Stay Overnight away from Home to Attend Hoopfest	0.378		
HOME	Binary Variable Identifying Respondents Who Stay Home and Attend Hoopfest	0.622		
HOTEL	Binary Variable Identifying Respondents Who Stay Overnight at a Hotel to Attend Hoopfest	0.167		
FAMILY	Binary Variable Identifying Respondents Who Stay Overnight with Family to Attend Hoopfest	0.199		
CAMPGR	Binary Variable Identifying Respondents Who Stay Overnight at a Campground to Attend Hoopfest	0.003		

OTHACC	Binary Variable Identifying Respondents Who Stay Overnight in Other Accommodations to Attend Hoopfest	0.009		
LODGING	Total Lodging Expenditures	82.369	0.000	215.261
LODGPAY	Binary Variable Identifying Respondents Who Pay for Lodging	0.188		
MEALS	Number of Meals Purchased	4.179	3.000	5.475
NFOOD	Binary Variable Identifying Respondents Who Purchased Food at Hoopfest	0.890		
FOODCOST	Total Food Expenditures	96.929	40.000	246.363
PURCH	Total Expenditures on Other Items	57.932	15.000	115.598
PURCHDV	Binary Variable Identifying Respondents Who Purchase Other Items	0.539		
VERYSAT	Binary Variable Identifying Respondents Who Are Very Satisfied with Hoopfest	0.717		
MODSAT	Binary Variable Identifying Respondents Who Are Moderately Satisfied with Hoopfest	0.271		
OTHSAT	Binary Variable Identifying Respondents Who Express Less than Moderate Satisfaction with Hoopfest	0.012		
DEFATT	Binary Variable Identifying Respondents Who Definitely Intend to Attend Hoopfest Next Year	0.625		
PROBATT	Binary Variable Identifying Respondents Who Probably Intend to Attend Hoopfest Next Year	0.307		
OTHATT	Binary Variable Identifying Respondents Who Express a Less than Probable Intention to Attend Hoopfest Next Year	0.068		
MOBILEAPP	Binary Variable Identifying Respondents Who Downloaded the Free Hoopfest Mobile Application	0.622		
Number of Observations:		336		

Table 2 contains the results of the simple hypothesis tests to assess the overall quantity of information in each of the dependent variables. The results closely (but not exactly) follow those of Friesner et al. (2021). Matched sample t-tests indicate that, with the exception of the sixth and final proposed survey item ("I am sad when I observe someone in distress"), all of the differences are statistically significant. The mean value for BESAllEN is 0.496, which implies that, at the mean, all eight survey items collective capture 1 – 0.496 = 0.504, or 50.4% of the available quantity of information. Means for the additional six proposed survey items exhibit mean values exceeding 0.496. This implies that each of these six survey items captures a smaller quantity of available information than all eight items assessed jointly. For example, survey item C (BESQCEN) only captures 1 – 0.502 = 0.498, or 48% of available information. The only case where a collection of survey items provides more information is when one examines only those two survey items contained in the original BES scale. Those two items capture 1 – 0.378 = 0.622 or 62.2% of available information. This implies (but does not prove) that at least one of the six proposed survey items do not contain a positive quantity of information, that one or more of these six items capture the same quantity of information or both.

TABLE 2
MATCHED SAMPLE HYPOTHESIS TESTS

Variable 1	**Mean**	**Variable 2**	**Mean**	**Mean Difference**	**t-Statistic**	**Prob.**	
BESAllEN	0.496	BESBaseEN	0.378	0.118	5.28	<0.001	**
BESAllEN	0.496	BESQCEN	0.502	-0.006	-2.08	0.038	**
BESAllEN	0.496	BESQDEN	0.513	-0.016	-4.79	<0.001	**
BESAllEN	0.496	BESQEEN	0.515	-0.019	-5.21	<0.001	**
BESAllEN	0.496	BESQFEN	0.521	-0.025	-6.24	<0.001	**
BESAllEN	0.496	BESQGEN	0.503	-0.007	-2.33	0.020	**
BESAllEN	0.496	BESQHEN	0.498	-0.002	-0.74	0.461	
BESAllEN	0.496	BESQCogEN	0.530	-0.034	-6.69	<0.001	**
BESAllEN	0.496	BESQBehEN	0.519	-0.022	-4.93	<0.001	**
BESAllEN	0.496	BESQAffEN	0.524	-0.028	-5.37	<0.001	**

Note: ** Indicates statistical significance at the 5 percent level or better
* Indicates statistical significance at the 10 percent level or better

Table 3 contains the first set of Tobit regressions, which examine the quantity of information in the BESAllEN variable and the quantity of information contained in the two original BES survey items considered jointly (i.e., BESAllEN – BESBaseEN). The first series of columns in Table 3 explains the normalized proportion for the quantity of information across all eight survey items (the two original items and the six new proposed items) taken collectively. The model's intercept (coefficient value: 0.589) and the Tobit disturbance term (coefficient value: 0.281) are statistically significant from zero. The chi-square test assessing the joint significance of the respondent-specific characteristics is also statistically significant. These results suggest three inferences. First, and holding constant the other regressors in the model, the eight survey items capture a positive and statistically significant proportion of the available information. The significance of the Tobit disturbance term coefficient suggests that the censoring of the quantity of information measure impacts the magnitudes of the estimates. Therefore the use of the Tobit model is likely to be an appropriate choice. Lastly, the joint significance of the other regressors suggests that respondents with specific characteristics do provide different quantities of information in their responses. That is, holding constant the design of the survey and the censoring of the dependent variable, failure to address the differences may lead to survey administration errors. Examination of those coefficient estimates that are statistically significant at a 5% level or better indicates that those individuals who report spending 5 or fewer hours at Hoopfest (HRS05) provide a higher quantity of information (recall that a higher value for the dependent variable indicates a lower quantity of information, so a negative coefficient estimate leads to an increase in the quantity of information) than those who stay for between 6-12 hours. Concomitantly, the coefficient estimate for HRS12U is negative and significant, which implies that those individuals who attend the event for 12 or more hours provide significantly lower quantities of information in their responses. Those who pay for lodging (LODGPAY) provide significantly lower information quantities, keeping the other specified regressors constant. However, the coefficient for ln(LODGING) is significant and negative, which indicates that (holding the other regressors in the model constant) as respondents spend more money on lodging expenses, the quantity of information in their responses increases. At the 10% level, a similar effect is present among respondents who purchase other items (PURCHDV) and at the 5% significance level, among those who report the magnitudes of those purchases (ln(PURCH)). Those who report being moderately satisfied (MODSAT; significant at the 5% level) and less than moderately satisfied (OTHSAT; significant at the 10% level) exhibit positive coefficient estimates, which implies that respondents with these satisfaction ratings provide significantly lower quantities of information compared to those individuals who report high levels of satisfaction (the omitted reference category), holding the other explanatory variables in the model constant. Lastly, the PROBATT coefficient estimate is positive and statistically significant at the 5% level. Thus, holding the other specified regressors constant, those who report an intention to attend Hoopfest next year provide lower quantities of information in their responses.

TABLE 3
TOBIT ANALYSIS OF ALL SIX SURVEY ITEMS CONSIDERED COLLECTIVELY

Dependent Variable:	**BESAllEN [Mean = 0.496]**					**BESAllEN-BESBaseEN [Mean = 0.118]**				
Regressor	**Coeff.**	**Std. Error**	**t-Statistic**	**P-value**		**Coeff.**	**Std. Error**	**t-Statistic**	**P-value**	
Intercept	0.589	0.082	7.16	<0.001	**	-0.041	0.115	-0.36	0.722	
AGE	0.000	0.001	-0.27	0.786		0.003	0.002	1.32	0.189	
FEMALE	-0.070	0.036	-1.93	0.054	*	-0.012	0.051	-0.24	0.810	
PLAY	0.040	0.042	0.95	0.340		0.035	0.058	0.60	0.546	
HRS05	-0.141	0.039	-3.61	<0.001	**	-0.007	0.055	-0.12	0.902	
HRS12U	0.134	0.065	2.05	0.041	**	-0.056	0.089	-0.63	0.528	
NPEOPLE	-0.005	0.004	-1.26	0.209		0.013	0.006	2.29	0.022	**
FAMILY	-0.077	0.044	-1.75	0.079	*	-0.072	0.062	-1.16	0.245	
LODGPAY	0.507	0.241	2.11	0.035	**	0.102	0.340	0.30	0.764	
Ln (LODGING)	-0.093	0.042	-2.24	0.025	**	-0.011	0.059	-0.19	0.851	
MEALS	-0.004	0.004	-1.00	0.318		0.002	0.005	0.40	0.690	
NFOOD	-0.101	0.092	-1.10	0.271		-0.042	0.129	-0.32	0.745	
Ln(FOOD COST)	0.016	0.021	0.78	0.435		0.023	0.030	0.77	0.441	
ln(PURCH)	-0.040	0.023	-1.75	0.080	*	-0.026	0.032	-0.82	0.414	
PURCHDV	0.193	0.097	1.98	0.048	**	0.036	0.136	0.26	0.792	
MODSAT	0.077	0.039	1.96	0.050	**	-0.032	0.055	-0.57	0.566	
OTHSAT	0.271	0.156	1.73	0.083	*	0.089	0.218	0.41	0.684	
PROBATT	0.104	0.039	2.69	0.007	**	0.002	0.054	0.04	0.969	
OTHATT	0.018	0.072	0.26	0.798		-0.040	0.102	-0.39	0.695	
MOBILE-APP	0.004	0.036	0.10	0.920		0.024	0.050	0.48	0.629	
Tobit Disturb. Term	0.281	0.012	22.95	<0.001	**	0.400	0.015	25.92	<0.001	**

Log-Likelihood Function			-98.777					-168.898		
Restricted Log-Likelihood Function			-128.003					-177.261		
Chi-Square Test Statistic Value			58.454	<0.001	**			16.726	0.608	
Degrees of Freedom			19					19		
Number of Observ.			336					336		

Note: ** Indicates statistical significance at the 5 percent level or better
* Indicates statistical significance at the 10 percent level or better

The second set of columns in Table 3 examines the quantity of information jointly contained in the original two BES survey items (i.e., BESAllEN – BESBaseEN). The results here are starkly different from the prior regression. The Tobit disturbance term is statistically significant at the 5% level, which implies that the decision to control for the censoring of the dependent variable is likely to be appropriate. However, the estimated intercept for the model is statistically insignificant. The chi-square test assessing the joint significance of the other regressors is also statistically significant from zero. These results suggest (but do not prove) that, when assessed collectively, the original BES survey items (and after accounting for censoring and the other covariates) do not contain a unique quantity of information that differs significantly from zero. There also do not appear to be any respondent-specific characteristics that provide greater or lesser quantities of information. In more practical terms, the results of this regression suggest that the two original BES survey items, when considered collectively, do not contain a significant quantity of information and may be poorly designed (holding the model's specification constant). However, no evidence exists to support the potential existence of a survey administration error (again, holding the model's specification constant).

Table 4 contains the second set of Tobit regressions, which assess the quantity of information captured in proposed survey items C (i.e., BESAllEN – BESQCEN), D (i.e., BESAllEN – BESQDEN), and E (i.e., BESAllEN – BESQEEN), when considered individually. The first set of columns examines the quantity of information contained in survey item C ("I recall my personal experience when I observe someone else in a similar situation"). The Tobit disturbance term is statistically significant at the 5% level, indicating that the decision to adjust for the censoring of the dependent variable was appropriate. However, neither the model's intercept, nor the chi-square test for joint regressor significance, is statistically insignificant at the 5% level. Holding the specification of the model constant, survey item C does not contain a statistically significant quantity of information, and the quantity of information does not vary significantly across groups of respondents. The second and third sets of columns contain results for Tobit regressions explaining the quantity of information in survey items D ("I get agitated when I see someone in distress") and E ("I feel happy when I see smiles on other people's faces"), respectively. The results for both regressions are similar to those for survey item C. More specifically, the parameter estimate for the Tobit regression is statistically significant. Still, the estimate for the model's intercept, and the chi-square test assessing the joint significance of the other model regressors, are both statistically insignificant at the 5% level. Thus, holding the specification of the model constant, survey items D, and E do not contain a statistically significant quantity of information, and the quantity of information does not vary significantly across respondents for either survey item.

TABLE 4
TOBIT ANALYSIS OF SURVEY ITEMS C, D, AND E

Dependent Variable:	**BESAllEN-BESQCEN [Mean = -0.006]**					**BESAllEN-BESQDEN [Mean = -0.016]**					**BESAllEN-BESQEEN [Mean = -0.019]**				
Regressor	**Coeff.**	**Std. Err.**	**t-Stat.**	**P-value**		**Coeff.**	**Std. Err.**	**t-Stat.**	**P-value**		**Coeff.**	**Std. Err.**	**t-Stat.**	**P-value**	
Intercept	-0.007	0.015	-0.47	0.639		-0.015	0.017	-0.84	0.399		0.016	0.018	0.87	0.383	
AGE	0.0003	0.000	1.32	0.186		0.0001	0.000	0.28	0.778		-0.0004	0.000	-1.32	0.189	
FEMALE	0.003	0.006	0.46	0.642		-0.019	0.008	-2.51	0.012	**	-0.013	0.008	-1.62	0.105	
PLAY	-0.005	0.007	-0.70	0.485		-0.002	0.009	-0.21	0.831		-0.006	0.009	-0.69	0.490	
HRS05	0.003	0.007	0.40	0.690		0.006	0.008	0.74	0.459		-0.023	0.009	-2.62	0.009	**
HRS12U	0.008	0.011	0.75	0.454		0.006	0.013	0.43	0.667		-0.001	0.014	-0.07	0.940	
NPEOPLE	0.001	0.001	1.20	0.229		0.0002	0.001	0.19	0.853		-0.0004	0.001	-0.41	0.685	
FAMILY	-0.008	0.008	-1.01	0.314		0.019	0.009	2.00	0.046	**	-0.014	0.010	-1.45	0.148	
LODGPAY	-0.064	0.043	-1.48	0.139		0.084	0.051	1.63	0.104		-0.012	0.054	-0.22	0.828	
Ln(LODGING)	0.009	0.007	1.20	0.230		-0.013	0.009	-1.47	0.143		0.0003	0.009	0.04	0.971	
MEALS	-0.0004	0.001	-0.60	0.549		-0.0003	0.001	-0.41	0.680		-0.0004	0.001	-0.53	0.594	
NFOOD	-0.021	0.016	-1.27	0.204		0.002	0.020	0.11	0.913		-0.015	0.020	-0.73	0.468	
Ln(FOOD COST)	0.002	0.004	0.66	0.508		0.0000	0.005	-0.01	0.993		0.002	0.005	0.44	0.658	
ln(PURCH)	0.0000	0.004	0.00	0.998		-0.002	0.005	-0.47	0.636		-0.004	0.005	-0.82	0.410	
PURCHDV	0.003	0.017	0.20	0.845		-0.004	0.021	-0.18	0.858		0.006	0.021	0.28	0.777	
MODSAT	-0.004	0.007	-0.52	0.604		0.008	0.008	0.99	0.325		-0.011	0.009	-1.29	0.197	
OTHSAT	0.017	0.027	0.60	0.547		-0.017	0.033	-0.52	0.603		0.017	0.034	0.50	0.619	
PROBATT	0.005	0.007	0.71	0.478		0.002	0.008	0.25	0.800		0.013	0.009	1.56	0.119	
OTHATT	-0.018	0.013	-1.38	0.168		0.002	0.015	0.15	0.883		0.031	0.016	1.94	0.053	*

MOBILE APP	-0.002	0.006	-0.34	0.735		0.003	0.008	0.39	0.698		0.017	0.008	2.16	0.030	**
Tobit Disturb. Term	0.050	0.002	25.92	<0.001	**	0.060	0.002	25.92	<0.001	**	0.063	0.002	25.92	<0.001	**
Log-Likelihood Function			527.113					466.023					451.179		
Restricted Log-Likelihood Function			517.354					456.236					438.375		
Chi-Square Test Statistic Value			19.519	0.424				19.574	0.421				25.608	0.141	
Degrees of Freedom			19					19					19		
Number of Observ.			336					336					336		

Note: ** Indicates statistical significance at the 5 percent level or better
* Indicates statistical significance at the 10 percent level or better

Table 5 contains the second set of Tobit regressions, which assess the quantity of information captured in proposed survey items F (i.e., BESAllEN – BESQFEN), G (i.e., BESAllEN – BESQGEN), and H (i.e., BESAllEN – BESQHEN), when considered individually. The first set of columns in Table 5 examines the quantity of information contained in survey item F ("I almost always understand the motives behind the actions of another person"). The Tobit disturbance term is statistically significant at the 5% level, indicating that the decision to adjust for the censoring of the dependent variable was appropriate. The estimated intercept's magnitude is -0.050, and it is statistically significant at the 5% level. The significant and negative coefficient estimate for the intercept indicates that item F contains a positive quantity of information, holding the other regressors constant. The chi-square test probability value for this regression is 0.140. Thus, when considered jointly, all included respondent-specific characteristics are not statistically significant. Thus, potential survey administration issues do not appear to exist for survey item F. The second and third sets of columns in Table 5 contain results for Tobit regressions explaining the quantity of information in survey items G ("I help others when I see they need help") and H ("I am sad when I observe someone in distress"), respectively. The results for both of these regressions are similar to those for survey items D and E. The parameter estimate for the Tobit disturbance term in each regression is statistically significant. However, in each regression, the estimate for the model's intercept, and the chi-square test assessing the joint significance of the other model regressors, are both statistically insignificant at the 5% level. Once again, holding the specification of the model constant, survey items G and H do not contain a statistically significant quantity of information, and the quantity of information does not vary significant across respondents for either survey item.

As noted earlier, the six proposed survey items align with three domains of empathy: cognitive empathy, behavioral empathy, and affective empathy. Two of the six proposed items align with each of these domains. In an ideal world, this eight item scale would cover overall empathy (as characterized by the two previously validated BES items, and each of the three empathy domains. It is therefore also interesting to empirically assess whether the three domain scales (each of which is comprised of two proposed survey items) provides a significant quantity of information and is free of possible survey administration errors. Table 6 presents three sets of Tobit regressions to address this issue. The first set of columns in Table 6 examine the quantity of information contained in the cognitive empathy construct (BESAllEN-BESCogEN), which consists of proposed survey items C ("I recall my personal experience when I observe someone else in a similar situation") and F ("I almost always understand the motives behind the actions of another person") considered jointly. The Tobit disturbance term is statistically significant at the 5% level, indicating that the decision to adjust for the censoring of the dependent variable was appropriate. The estimated intercept's magnitude is -0.056, and it is statistically significant at the 5% level. The significant and negative coefficient estimate indicates that the cognitive empathy items jointly contain a positive quantity of information, holding the other regressors constant. The chi-square test probability value for this regression is 0.522. Thus, when considered jointly, all included respondent-specific characteristics are not statistically significant. Thus, potential survey administration issues do not appear for these two survey items.

The second set of columns in Table 6 assesses the quantity of information contained in the behavioral empathy scale (BESAllEN-BESBehEN), which consists of survey items D ("I get agitated when I see someone in distress") and G ("I help others when I see they need help"), considered jointly. In this regression, the Tobit model parameter estimate is statistically significant. However, neither the model's intercept nor the chi-square test for joint regressor significant are statistically insignificant at the 5% level. Thus, holding the specification of the model constant, survey item E does not contain a statistically significant quantity of information, and the quantity of information does not vary significantly across groups of respondents.

The final columns in Table 6 assess the quantity of information contained in the affective empathy scale (BESAllEN-BESAffEN), which consists of survey items E ("I feel happy when I see smiles on other people's faces") and H ("I am sad when I observe someone in distress"), considered jointly. The Tobit disturbance term is statistically significant at the 5% level. However, the model's intercept is not statistically significant from zero. Thus, at the mean, the affective empathy scale does not appear to add significant information, holding the other specified regressors constant. Additionally, the chi-square test produces a probability value that rejects the null hypotheses that the regressors, considered jointly, do not predict the dependent variable. This indicates that, at a 5% significance level and holding the other regressors constant, different groups of respondents provide significantly different quantities of information. Examining the signs and significance of the individual parameter estimates, we find that female respondents and those who spend five or fewer hours at Hoopfest exhibit negative and significant parameter estimates. This indicates that these respondents provide more information when responding to the affective empathy items, compared to male respondents and those who attend for more than 5 hours, respectively. Additionally, the coefficient estimate for OTHATT is positive and significant, implying that individuals less likely to attend Hoopfest next year provide significantly lower quantities of information in their responses (compared to the reference group), holding the other specified regressors constant. Considered cumulatively, responses to the affective empathy items (considered jointly) both contain a statistically insignificant quantity of information and vary based on the type of respondent.

DISCUSSION AND CONCLUSIONS

This manuscript extends the methodologies of Friesner et al. (2023) and Friesner et al. (2021) to simultaneously determine whether a survey's responses (whether a pilot survey or an established survey that is administered repeatedly over time) contain a positive quantity of information, as well as whether the quantity of information in the survey's responses vary systematically by the type of survey respondent. The latter is a necessary (but not sufficient) condition for a well-designed survey. Concomitantly, the lack of significant information represents a flaw in a survey's design. The presence of the latter is a potential flaw in the survey administration process. If one or both issues exist, future revisions to the survey and its administration process may be necessary to correct these issues. The proposed methodology is operationalized using a customer satisfaction survey administered regularly over time at a large, amateur sporting event. The survey contained (among other items) a two-item, previously validated empathy scale, as well as six survey items (using the same response options) that are proposed to assess various aspects of empathy (cognitive, behavioral, and affective).

The results of the analysis are three-fold. First, univariate hypothesis tests indicate that seven of the eight survey items, when considered individually and without controlling for respondent-specific factors) contain a statistically significant, positive quantity of information. These eight items collectively explain approximately (1-0.496 = 0.504) 50.4%% of the available quantity of information. Individual survey items capture between 47.6% and 50.2% of the available quantity of information in each of those items, respectively.

Second, Tobit regression analyses indicate that, after controlling for the censoring of the information entropy measure, as well as various respondent characteristics, these eight items, taken collectively, continue to capture a significant amount of the available quantity of information in these survey items. Respondents with specific characteristics were significantly more likely to give relatively higher or lower quantities of information in their responses, which raises the possibility that changes may need to be made to the survey's administration processes should these items be retained in the survey.

TABLE 5
TOBIT ANALYSIS OF SURVEY ITEMS F, G, AND H

Dependent Variable:	**BESAllEN-BESQFEN [Mean = -0.025]**					**BESAllEN-BESQGEN [Mean = -0.007]**					**BESAllEN-BESQHEN [Mean = -0.002]**				
Regressor	**Coeff.**	**Std. Err.**	**t-Stat.**	**P-value**		**Coeff.**	**Std. Err.**	**t-Stat.**	**P-value**		**Coeff.**	**Std. Err.**	**t-Stat.**	**P-value**	
Intercept	-0.050	0.020	-2.48	0.013	**	-0.020	0.015	-1.33	0.184		0.016	0.013	1.23	0.220	
AGE	0.0001	0.0003	0.18	0.854		-0.0001	0.0003	-0.57	0.567		0.000	0.000	-0.23	0.820	
FEMALE	-0.010	0.009	-1.08	0.279		0.002	0.007	0.25	0.805		0.000	0.006	0.00	0.996	
PLAY	0.013	0.010	1.27	0.205		0.001	0.007	0.11	0.914		0.003	0.007	0.42	0.676	
HRS05	-0.010	0.010	-1.06	0.290		0.014	0.007	1.96	0.050	**	0.000	0.006	-0.06	0.955	
HRS12U	0.005	0.015	0.34	0.730		-0.008	0.012	-0.72	0.470		-0.005	0.010	-0.54	0.593	
NPEOPLE	-0.0001	0.001	-0.14	0.887		0.0002	0.001	0.21	0.833		0.000	0.001	-0.08	0.939	
FAMILY	-0.005	0.011	-0.47	0.635		-0.006	0.008	-0.79	0.429		-0.007	0.007	-1.04	0.299	
LODGPAY	-0.040	0.059	-0.68	0.495		0.052	0.044	1.18	0.238		0.063	0.038	1.66	0.097	*
ln(LODGING)	0.009	0.010	0.88	0.378		-0.009	0.008	-1.13	0.258		-0.011	0.007	-1.59	0.112	
MEALS	0.000	0.001	0.29	0.772		0.0001	0.001	0.21	0.832		0.000	0.001	-0.38	0.707	
NFOOD	0.041	0.022	1.83	0.068	*	-0.024	0.017	-1.45	0.146		-0.013	0.014	-0.92	0.358	
ln(FOODCOST)	-0.002	0.005	-0.36	0.718		0.008	0.004	2.00	0.046	**	0.001	0.003	0.16	0.877	
ln(PURCH)	-0.002	0.006	-0.30	0.761		-0.004	0.004	-0.92	0.356		-0.003	0.004	-0.91	0.365	
PURCHDV	0.007	0.024	0.30	0.767		0.019	0.018	1.09	0.274		0.014	0.015	0.92	0.359	
MODSAT	0.003	0.010	0.34	0.735		0.006	0.007	0.89	0.373		-0.001	0.006	-0.09	0.930	
OTHSAT	0.020	0.038	0.54	0.589		0.011	0.028	0.38	0.705		0.005	0.024	0.19	0.846	
PROBATT	-0.003	0.009	-0.36	0.720		-0.002	0.007	-0.34	0.735		-0.006	0.006	-0.99	0.323	
OTHATT	-0.030	0.018	-1.71	0.087	*	0.010	0.013	0.76	0.446		0.017	0.011	1.52	0.129	

MOBILE APP	-0.007	0.009	-0.84	0.401		0.007	0.006	1.07	0.286		-0.006	0.006	-1.16	0.246	
Tobit Disturb. Term	0.069	0.003	25.92	<0.001	**	0.052	0.002	25.92	<0.001	**	0.045	0.002	25.92	<0.001	**
Log-Likelihood Function			420.411					518.771					566.455		
Restricted Log-Likelihood Function			407.586					510.720					558.933		
Chi-Square Test Statistic Value			25.650	0.140				16.103	0.650				15.044	0.720	
Degrees of Freedom			19					19					19		
Number of Observ.			336					336					336		

Note: ** Indicates statistical significance at the 5 percent level or better
* Indicates statistical significance at the 10 percent level or better

TABLE 6
TOBIT ANALYSIS OF SURVEY ITEMS GROUPED BY COGNITIVE, BEHAVIORAL AND AFFECTIVE DOMAINS

Dependent Variable:	BESAllEN-BESCogEN [Mean = -0.034]					BESAllEN-BESBehEN [Mean = -0.022]					BESAllEN-BESAffEN [Mean = -0.028]				
Regressor	Coeff.	Std. Err.	t-Stat.	P-value		Coeff.	Std. Err.	t-Stat.	P-value		Coeff.	Std. Err.	t-Stat.	P-value	
Intercept	-0.056	0.026	-2.16	0.031	**	-0.040	0.023	-1.69	0.092	*	0.039	0.026	1.52	0.127	
AGE	0.000	0.000	0.94	0.346		0.000	0.000	0.09	0.930		0.000	0.000	-1.08	0.281	
FEMALE	-0.006	0.011	-0.50	0.615		-0.016	0.010	-1.53	0.126		-0.025	0.011	-2.24	0.025	**
PLAY	0.009	0.013	0.72	0.470		0.000	0.012	0.02	0.987		-0.009	0.013	-0.72	0.471	
HRS05	-0.009	0.012	-0.72	0.475		0.022	0.011	1.97	0.049	**	-0.028	0.012	-2.28	0.022	**
HRS12U	0.020	0.020	1.00	0.319		-0.003	0.018	-0.16	0.875		-0.018	0.020	-0.90	0.370	
NPEOPLE	0.001	0.001	0.53	0.597		0.000	0.001	0.28	0.780		0.000	0.001	-0.28	0.777	
FAMILY	-0.015	0.014	-1.08	0.281		0.016	0.013	1.28	0.199		-0.024	0.014	-1.71	0.088	*
LODGPAY	-0.096	0.077	-1.26	0.209		0.145	0.069	2.09	0.036	**	0.081	0.076	1.06	0.290	
Ln (LODGING)	0.017	0.013	1.28	0.202		-0.023	0.012	-1.90	0.057	*	-0.015	0.013	-1.16	0.247	
MEALS	-0.0002	0.001	-0.16	0.872		-0.0002	0.001	-0.19	0.852		-0.001	0.001	-0.86	0.392	
NFOOD	0.008	0.029	0.28	0.778		-0.027	0.026	-1.04	0.299		-0.039	0.029	-1.36	0.175	
Ln(FOODCOST)	0.003	0.007	0.42	0.677		0.008	0.006	1.35	0.176		0.004	0.007	0.57	0.570	
ln(PURCH)	-0.002	0.007	-0.27	0.784		-0.007	0.007	-1.13	0.257		-0.011	0.007	-1.59	0.111	
PURCHDV	0.009	0.031	0.31	0.757		0.021	0.028	0.77	0.442		0.036	0.031	1.17	0.241	
MODSAT	0.003	0.012	0.22	0.823		0.015	0.011	1.29	0.196		-0.012	0.012	-0.95	0.341	
OTHSAT	0.050	0.049	1.01	0.313		-0.003	0.044	-0.07	0.944		0.032	0.049	0.66	0.510	
PROBATT	-0.002	0.012	-0.20	0.845		0.002	0.011	0.16	0.875		0.009	0.012	0.71	0.477	
OTHATT	-0.050	0.023	-2.18	0.029		0.008	0.021	0.38	0.706		0.062	0.023	2.70	0.007	**

MOBILE APP	-0.011	0.011	-0.96	0.337		0.012	0.010	1.15	0.252		0.015	0.011	1.36	0.173	
Tobit Disturb. Term	0.090	0.003	25.92	<0.001	**	0.081	0.003	25.92	<0.001	**	0.090	0.003	25.92	<0.001	**
Log-Likelihood Function			331.648					367.023					333.393		
Restricted Log-Likelihood Function			322.641					357.591					316.843		
Chi-Square Test Statistic Value			18.013	0.522				18.865	0.466				33.101	0.023	**
Degrees of Freedom			19					19					19		
Number of Observ.			336					336					336		

Note: ** Indicates statistical significance at the 5 percent level or better
* Indicates statistical significance at the 10 percent level or better

Third, Tobit regressions analyzing the individual survey items (as well as scales based on combinations of two survey items each) yielded very different results when compared to the univariate hypothesis tests. When considered individually, and after controlling for both the censoring of the information entropy variable and respondent characteristics, only the cognitive empathy construct (comprised of survey items C and F), and question F yielded a significant, positive quantity of information. And in each of these regressions, no significant evidence is found to suggest that respondents with specific characteristics provide greater or lesser quantities of information in their responses. These results suggest that survey item F and the cognitive empathy scale (which appears to be largely determined by survey item F) may be a well-designed and administered survey item. However, within the Tobit regression analysis, no other survey items yielded a significantly positive quantity of information. Additionally, the regressions for the affective empathy scale (comprised of survey items E and H) indicated that specific types of respondents (including females, those attending Hoopfest for 5 hours or less, and those who are less likely to attend the event in subsequent years) provided significantly different quantities of information in their responses compared to other types of respondents. This suggests that, if these items remain in the survey an adjustment in the survey's administration processes should be adjusted to account for these differences.

The results of this study have important implications for the survey design and applied marketing literature. The most important implication of this analysis is that survey design and administration errors may be present in a particular survey and its administration processes. These possible errors may be independent or intertwined. If intertwined, they may confound or compound each other. Therefore, it is important to assess survey data for both types of errors. The current manuscript provides a straightforward means to assess data for both issues simultaneously.

The second implication of this study follows directly from the first implication. Friesner et al. (2021) provide a simple means – using univariate hypothesis tests – to characterize and assess the quantity of information contained in one or more survey items. Their methodology fails to address the fact that several related survey items may characterize the same underlying quantity of information. Moreover, if that quantity of information is significant, it will make all survey items appear to contain a statistically significant quantity of information. Regression analysis is a more appropriate technique, because it allows the research to characterize the quantity of information. In this study's empirical application, we found that the cognitive empathy survey items (survey items C and F) jointly contained a unique quantity of information. However, this was primarily due to the quantity of information provided by survey item F, rather than survey item C.

The results of this study also provide an interesting opportunity for future research. In the absence of censoring and when the parameter estimates for all covariates are jointly statistically insignificant from zero, the model's intercept represents the quantity of information contained in that variable. When censoring occurs, when another parameter estimate is statistically distinct from zero, or both, then adjustments must be made when calculating the quantity of information in a given survey item (i.e., calculating an analog of a marginal effect). While not pursued in this manuscript, it is a straightforward extension of our methodology to undertake these calculations. However, this issue identifies an opportunity for future research. More specifically, the Tobit disturbance term is statistically significant in each of the regressions presented in this paper. Thus, adjusting for the possible censoring of the dependent variable significantly impacts the quantity of information in each survey item. The possibility of censoring arises from equation (3). To interpret overall characterizations of the quantity of information, each entropy value is expressed as a proportion of its theoretical maximum (i.e., the state of ignorance, when response distributions are uniform). The magnitude of the maximum possible entropy value (i.e., the denominator in equation (3)) is determined by the total number of response options available to respondents for a given survey item (evaluated using uniform probabilities). In the case of our empathy survey items, respondents evaluated each survey item using a 1 to 5 scale. A different number of response options tautologically changes the maximum entropy for that item, and in turn changes whether a given response is censored. This implies two related things. First, the statistical significance of the Tobit disturbance term is likely controlling for the number of possible response options available to choose from across those survey items being evaluated. Second, if this is the case, a change in the number of response options (including corresponding changes to the interpretation/definitions of those response options) across the same set of survey items, may impact the normalized quantity of information captured by those survey items and response scales. This creates an opportunity to evaluate how the researcher's choice of response scales impacts the quantity of information collected in that survey. Ideally, response scales should strike a balance between parsimony (i.e., keeping possible response options more intuitive, which means fewer response options) and information (i.e., allowing more response options potentially allows for collecting a larger quantity of information). Future research that extends this paper's methodology to evaluate that tradeoff would provide a meaningful contribution to the survey design and applied marketing literatures.

The current manuscript also exhibits several limitations, which should be noted by the reader and addressed in future research. Most notably, this manuscript focuses exclusively on the quantity of information. However, the quantity of information is only one aspect of survey design and administration. Researchers are also interested in what Dahl and Osteras (2010) discuss as the "quality of information"; namely the inter-relationships between variables, holding constant the total amount of information in those responses. The latter is typically assessed using a wide array of alternative empirical techniques (for example, see Hair et al. (2006) for various techniques). No attempt in this manuscript is made to assess potential tradeoffs between the quantity of information available for analysis and the quality of information extracted from that analysis. It may be the case that changes in survey design and administration may increase both or the quantity of information at the expense of diminished quality of information.

Second, the data used in this analysis are drawn from a single survey designed to be administered at a unique amateur sporting event. As such, the results and inferences in this study may be specific to this survey and this specific event. The specific demographic and other control variables collected through this survey also limit the results. Replications of this analysis using different surveys, administered in various contexts, is necessary to establish the generalizability of the results in this manuscript.

Third, all empirical results presented in this manuscript are crucially dependent on the specification of the regression models. Suppose the assumptions underlying the Tobit model are inappropriate. In that case, if the model omits important control variables or if the linear-in-parameters specification used in the Tobit model is incorrect, the results contained in this manuscript will be biased. Future replications of this study are necessary to determine whether these limitations are of concern.

Lastly, this manuscript used a specific process suggested in the literature to normalize entropy, which allows the quantity of information to be interpreted simply and intuitively. However, alternative normalization processes may exist that are equally intuitive and do not create censoring. Future research is encouraged to develop such alternative normalizations. In doing so, the censoring process is avoided, and regression analyses can be estimated using ordinary least squares (or another, equally simple means). This would not allow the reader to more easily obtain information entropy measure(s) and interpret regression results.

ACKNOWLEDGEMENT

This research was conducted while Dr. Friesner was on the faculty at North Dakota State University.

ENDNOTES

1. Administration errors that can be evaluated using the methodology described in this manuscript can potentially go beyond sampling design (i.e. using stratified instead of simple random sampling for efficiency). For example, researchers could evaluate survey responses from different interviewers to discern interviewer bias (and potentially enhance training to mitigate any issues).
2. The null hypothesis is one of no difference. Thus, the formulation of (3) is appropriate to capture these trends. However, if the researcher is concerned with the magnitude of any differences, and if those differences are statistically significant, we expect the magnitude of DifNormalized $H_i(p)$ to be negative, since a larger number of questions should capture a greater proportional amount of available information. That is, since Normalized $H_i(p) = \frac{H_i(p)}{Max(H_i(p))}$ represents the quantity of information that is not captured in the survey

item(s) of interest, we also expect $Normalized_{All}\ H_i(p) < Normalized_{Reduced}\ H_i(p)$, since adding additional survey items should potentially add an additional quantity of information.

3. If the researcher was interested in applying non-parametric hypothesis tests, Wilcoxon-signed rank tests may also be used in place of matched sample hypothesis tests.

REFERENCES

Abd Gani, N., Rathakrishnan, M., & Krishnasamy, H. (2020). A pilot test for establishing validity and reliability of qualitative interview in the blended learning English proficiency course. *Journal of Critical Reviews*, *7*(5), 140–143.

Ballard, A. (2019). Framing bias in the interpretation of quality improvement data: Evidence from an experiment. *International Journal of Health Policy and Management*, *8*(5), 307–314.

Batra, R., Homer, P., & Kahle, L. (2001). Values, susceptibility to normative influence, and attribute importance weights: A nomological analysis. *Journal of Consumer Psychology*, *11*(2), 115–128.

Beatty, S., Kahle, L., Homer, P., & Misra, S. (1985). Alternative management approaches to consumer values: The list of values and the Rokeach value survey. *Psychology and Marketing*, *2*(3), 181–200.

Bozman, C.S., Kurpis, L.V., & Frye, C. (2010). Hoopfest: Using longitudinal economic impact data to assess the success of a strategic reorientation. *Sport Management Review*, *13*, 65–81.

Bradley, K., Peabody, M., & Sampson, S. (2015). Quality control in survey design: Evaluating a survey of educators' attitudes concerning differentiated compensation. *International Journal of Assessment Tools in Education*, *2*(1), 3–21.

Carré, A., Stefaniak, N., D'Ambrosio, F., Bensalah, L., & Besche-Richard, C. (2013). The Basic Empathy Scale in Adults (BES-A): Factor structure of a revised form. *Psychological Assessment*, *25*(3), 679–691.

Dahl, F., & Osteras, N. (2010). Quantifying information content in survey data by entropy. *Entropy*, *12*(2), 161–163.

Deming, W.E. (1944). On errors in surveys. *American Sociological Review*, *9*(4), 359–369.

Dillman, D.A. (2000). *Mail and internet surveys: The tailored design method.* New York, NY: John Wiley and Sons.

Draugalis, J., Coons, S., & Plaza C. (2008). Best practices for survey research reports: A synopsis for authors and reviewers. *American Journal of Pharmaceutical Education*, *72*(11), Article 11. Retrieved from https://www.ajpe.org/content/72/1/11

Entman, R. (2007). Framing bias: Media in the distribution of power. *Journal of Communication*, *57*(1), 163–173.

Fenner, K., Hyde, M., Crean, A., & McGreevy, P. (2020). Identifying sources of potential bias when using online survey data to explore horse training, management, and behaviour: A systematic literature review. *Veterinary Sciences*, *7*(3), Article 140. https://doi.org/10.3390/vetsci7030140

Friesner, D., Bozman, C., McPherson, M., Valente, F., & Zhang, A. (2021). Information entropy and scale development. *Journal of Survey Statistics and Methodology*, *9*(5), 1183–1203.

Friesner, D., Bozman, C., McPherson, M., Valente, F., & Zhang, A. (2023). Information entropy as a quality control tool in survey research. *Journal of Marketing Development and Competitiveness*, *17*(1), 46–61.

Friesner, D., Valente, F., & Bozman, C.S. (2016). Using entropy-based information theory to evaluate survey research. *Journal of Marketing Development and Competitiveness*, *10*(3), 32–48.

Golan, A. (2006). Information and entropy econometrics – A review and Synthesis. *Foundations and Trends in Econometrics*, *2*(1–2), 1–145.

Golan, A., Judge, G., & Miller, D. (1996). *Maximum entropy econometrics: Robust estimation with limited data.* New York, NY: John Wiley and Sons.

Golan, A., Judge, G., & Perloff, J. (1996). A maximum entropy approach to recovering information from multinomial response data. *Journal of the American Statistical Association*, *91*(434), 841–853.

González-Cabrera, M., Ortega-Martínez, A., Martínez-Galiano, J., Hernández-Martínez, A., Parra-Anguita, L., & Frías-Osuna, A. (2020). Design and validation of a questionnaire on communicating bad news in nursing: A pilot study. *International Journal of Environmental Research and Public Health*, *17*(2), Article 457. https://doi.org/10.3390/ijerph17020457

Hair, J., Black, W., Babin, B., Anderson, R., & Tatham, R. (2006). *Multivariate data analysis* (6th Ed.). Upper Saddle River, NJ: Pearson/Prentice Hall.

Hendra, R., & Hill, A. (2019). Rethinking response rates: New evidence of little relationship between survey response rates and nonresponse bias. *Evaluation Review*, *43*(5), 307–330.

Imbens, G., & Lancaster, T. (1996). Efficient estimation and stratified sampling. *Journal of Econometrics*, *74*(2), 289–318.

Jaynes, E. (1957). Information theory and statistical mechanics. *Physics Review*, *106*(4), 620–630.

Jaynes, E. (1982). On the rationale of maximum-entropy methods. *Proceedings of the IEEE*, *70*(9), 939–952.

Johnson, T., Cho, Y., Holbrook, A., O'Rourke, D., Warnecke, R., & Chavez, N. (2006). Cultural variability in the effects of question design features on respondent comprehension of health surveys. *Annals of Epidemiology*, *16*(9), 661–668.

Jolliffe, D., & Farrington, D.P. (2006). Development and validation of the basic empathy scale. *Journal of Adolescence*, *29*, 589–611.

Jordan, P., & Troth, A. (2020). Common method bias in applied settings: The dilemma of researching in organizations. *Australian Journal of Management*, *45*(1), 3–14.

Kahle, L., Beatty, S., & Homer, P. (1986). Alternative measurement approaches to consumer values: The list of values (LOV) and values and life style (VALS). *Journal of Consumer Research*, *13*(3), 405–409.

Kahle, L.R. (1983). *Social Values and Social Change: Adaptation to Life in America*. New York, NY: Praeger.

Kurpis, L.V., Bozman, C.S., & Kahle, L.R. (2010). Distinguishing between amateur sport participants and spectators: The list of values approach. *International Journal of Sport Management and Marketing*, *7*(3/4), 190–201.

Marquis, K., Marquis, M., & Polich, J. (1986). Response bias and reliability in sensitive topic surveys. *Journal of the American Statistical Association*, *81*(394), 381–389.

Schnell, L. (2014, June 27). *Hoopfest, the world's largest 3-3 tourney, turns 25 this weekend*. Retrieved from https://www.si.com/college-basketball/2014/06/27/spokane-hoopfest-3-3-basketball-tournament-25-years

Shannon, C. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*, 37–423.

Suchman, E.A. (1962). An analysis of "bias" in survey research. *The Public Opinion Quarterly*, *26*(1), 102–111.

Tripathi, G. (2011). Generalized method of moments (GMM) based inference with stratified samples when the aggregate shares are known. *Journal of Econometrics*, *165*(2), 258–265.

van Teijlingen, E., & Hundley, V. (2001). The importance of pilot studies. *Social Research Update*, (35), 1–4. Retrieved from http://sru.soc.surrey.ac.uk/SRU35.pdf

van Teijlingen, E., Rennie, A.-M., Hundley, V., & Graham, W. (2001). The importance of conducting and reporting pilot studies: The example of the Scottish Births Survey. *Journal of Advanced Nursing*, *34*(3), 289–295.

Wikman, A., & Wärneryd, B. (1990). Measurement errors in survey questions: Explaining response variability. *Social Indicators Research*, *22*(2), 199–212.

Wing, C., Simon, K., & Bello-Gomez, R. (2018). Designing difference in difference studies: Best practices for public health policy research. *Annual Review of Public Health*, *39*(1), 453–469.

Wiseman, F. (1972). Methodological bias in public opinion surveys. *The Public Opinion Quarterly*, *36*(1), 105–108.