Information Entropy as a Quality Control Tool in Survey Research

Dan Friesner
North Dakota State University

Carl S. Bozman
Gonzaga University

Matthew McPherson
Gonzaga University

Faith Valente
North Idaho College

Anqing Zhang
George Washington University
Children's National Medical Center

This manuscript assesses whether information entropy can be used to test for flaws within survey research. Under the study's null hypothesis, a set of well-designed survey items should not exhibit systematic differences in the quantity of information – as measured by information entropy - provided by specific groups of respondents. The study was conducted within the context of a major amateur sporting event in 2018. Customer satisfaction was assessed using a survey whose core questions have been assessed repeatedly over time, and which contains two previously validated constructs. One construct (the List of Values) is unchanged from its original form. Another construct (the Basic Empathy Scale) contains two original survey items, with six additional survey items added to that construct (using the same response scale). Regression analysis indicates that the well-designed set of survey items exhibit no statistically significant differences in information entropy. The poorly designed survey items exhibit statistically significant differences in information entropy, suggesting that information entropy can be a useful quality control tool in survey design and assessment.

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

INTRODUCTION

Surveys represent one of the most common means to elicit self-reported information from a population. This is especially true in circumstances where the information is not directly observable (for example, eliciting respondent attitudes, beliefs, or values), or where self-reporting is the only viable means (for example, quantifying an individual's capacity for empathy) to collect such information (Batra et al., 2001; Beatty et al., 1985; Bradley et al., 2015; Innamorati et al., 2019; Kahle et al., 1986; Spreng et al., 2009). Unfortunately, while almost anyone can design a survey, it is challenging to create a survey that elicits accurate, precise, and informed responses. Small mistakes in survey design, survey administration, item wording, or response scale design can lead to "unfortunate practices that diminish the usefulness of the results" (Dillman, 2000, p. 269). As Dillman (2000, p. 268) further notes, these mistakes are exacerbated when designing and administering surveys to collect immediate feedback following an event, such as an experience or the purchase of a service. The survey research literature has developed an array of best practices to guide survey development, as well as a battery of empirical techniques to assess the validity and reliability of surveys, once constructed (Churchill, Jr., 1979). Dillman (2000) provides an overview of survey design and administration techniques, while Hair et al. (2006) provide an overview of many commonly used empirical techniques to assess survey responses, once collected.

A critical, yet under-explored, aspect of survey research is "survey iteration." The development and refinement of a survey is an evolutionary process (Churchill, Jr., 1979). At each step in that process, specific methodologies and techniques exist to assess the appropriateness (i.e., validity, reliability, etc.) of that particular element of a survey's design, administration, or analysis of survey results. Most marketing research textbooks (for example, see Smith and Albaum, (2005)) provide an overview of these methodologies and techniques. A major principle woven throughout survey methodologies is that a survey's design, implementation, and methods of analysis must satisfy *all* generally accepted assessment benchmarks, both conceptual and empirical, to be considered "appropriate" for use (Fain, 2009, pp. 120-123). This makes the creation and implementation of surveys both challenging and time consuming.

One possible iterative error may occur when researchers attempt to adapt existing (and previously validated) survey items, scales, or elements of survey administration. Such adaptations may include (but are not limited to) changing response scales, minor changes to the wording of existing survey items (especially to those with lower literacy levels or when translating the survey to other languages), adding new items to the survey, or changing the delivery form of the survey (i.e., paper surveys to web-based surveys). In such cases, the literature's approach is to consider the revised/iterated survey a *new* survey, and the researcher must completely re-assess the revised survey, in its entirety, to ensure that it retains all of the properties of the previous iteration of the survey (Nieswiadomy, 2012, p. 168). This is both cumbersome and daunting to individuals who are not well-versed in all aspects of survey design, administration, and analysis. However, it is vitally important because a change in the properties of the survey, inclusive of its administration, fundamentally change the *information* collected by the researcher. Most experts suggest using pilot studies as a simple and effective means to collect a limited amount of data and assess the properties (and, by extension, the appropriateness) of the revised survey prior to applying the survey within the context of a complete, adequately powered, study design (Dillman, 2000, pp. 146-147; Nieswiadomy, 2012, p. 168).

Focusing on the information contained in a set of responses (possibly collected via a pilot study) provides an efficient and sufficient (but not necessary) means to identify possible flaws in survey adaptation or re-design. If the information collected from a survey changes across iterations or adaptations of a survey, or different types of respondents provide fundamentally different information within the same iteration of a survey, then the survey is unlikely to exhibit a full set of properties (reliability, validity, etc.) necessary to justify its use in academic or professional research. Concomitantly, if the information collected does not vary across survey iterations/adaptations or across respondents within the same survey, then it is worthwhile to invest additional time and effort to fully explore the properties of the survey.

Dahl and Osteras (2010) are careful to distinguish between the *quantity* of information and the *quality* of information contained in survey responses. The *quantity* of information refers to a comparison of the distribution of observed survey responses compared to the researcher's theoretical expectation for this distribution. If the theoretical expectation matches the empirically observed distribution, the data contain zero quantity of information, because the same distribution would be observed in the absence of the data collection process. In such cases, there is little to be gained from collecting the data. For data sets with a positive quantity of information, the *quality* of information refers to the underlying trends and interrelationships that exist within and between variables in the data set (i.e., means, variances, correlations, etc.). Dahl and Osteras (2010) note that the vast majority of tools available to assess survey design characterize the *quality* of information provided by respondents and incorporate that information (whether retrospectively or prospectively) into all aspects of the survey research process (design, data collection and data analysis). But aside from survey response rates, few tools are available to assess the *quantity* of information. This is problematic, because there is no reason to assess the quality of information in data if it contains little or no quantity of information. One notable exception is the concept of *information entropy*, which explicitly measures the quantity of information contained in a given set of survey responses (Dahl & Osteras, 2010; Golan, 2006; Golan, Judge, & Miller, 1996; Golan, Judge, & Perloff, 1996; Janes, 1957; Janes, 1982; Shannon, 1948; Shannon, 1949). The entropy measure can be assessed in absolute terms (i.e., the quantity of information captured in a single survey item) or in relative terms (i.e., cross-entropy, or the relative gains in the quantity of information contained in responses from one survey item to the next). It is also relatively simple to calculate using common spreadsheet computer programs, and is sufficiently flexible to characterize information across respondents completing the same survey (i.e., cross-sectionally), across a collection of survey items measuring the same underlying construct, across repetitions of the same survey item(s) over time, or any combination of these three dimensions. If specific groups of survey responses generate fundamentally different levels of information entropy than other groups of respondents, then differences in the quantity of information provided by each of these groups also exist in the data. In such cases, it is likely that a possible flaw (or "quality control" issue) exists in the survey's design or implementation, which must be addressed. Concomitantly, if no discernable differences in information entropy exist across different groups of respondents, then it is worthwhile to pursue more extensive, traditional assessments of the survey's information quality, inclusive of validity and reliability. Taken collectively, information entropy may provide a simple and effective quality control tool to help guide "survey iteration," and more generally the adaptation of surveys to alternative uses, settings, or time frames.

The purpose of this manuscript is to conduct an initial, exploratory case study to assess whether information entropy can be used as a simple, initial quality control test for flaws within survey research. The study is conducted within the context of a natural experiment. We use a single administration of a customer satisfaction survey that been assessed repeatedly over time using scientifically rigorous survey techniques, and likely has high test-retest validity (Bozman et al., 2010; Kurpis et al., 2010). The survey contains two previously validated constructs, with multiple items in each construct. One construct is unchanged from its original form. In the other construct, the original survey items and response scales are unchanged, but several additional survey items (using the same response scale) are added to that construct. The quantity of information (as measured by information entropy) is calculated for each construct. We apply regression analysis using each entropy measure as a dependent variable, and various respondent characteristics as independent variables to assess the null hypothesis of no difference in information entropy exist across different types of respondents. As noted previously, for a well-designed construct (our first construct), the null hypothesis should not be rejected for each and every independent variable. For a poorly designed construct (our second construct) the null hypothesis should be rejected for one or more regressors.

In the next section, the basic concepts of information entropy (including assumptions, definitions, and hypotheses) as drawn from the information theory literature and used to characterize the quantity of information provided by survey respondents, are described. In the third section, we describe the survey's implementation and the details surrounding the natural experiment, namely the 2018 Hoopfest basketball tournament. Variable names, definitions, and data collection processes are described in the fourth section. The fifth section presents the results of the information entropy analysis. We conclude the paper by discussing how the results inform survey research methodology, by identifying major study limitations, and suggesting directions for future research in this area.

METHODOLOGY

Formulating Information Entropy Measures

The concept of entropy requires an assumption before it can be applied to the evaluation of survey responses. This simple assumption is the prior belief of ignorance. In terms of the survey design, this indicates that a survey is designed and administered so that the designer/researcher has no prior expectations concerning the distribution of the response for a given survey item or a given respondent (Jaynes, 1957; Janes, 1982). In other words, the distribution of survey responses is expected to be uniform. A uniformly distributed response minimizes 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 (Friesner et al., 2016; Smith & Albaum, 2005). Movement away from a uniform distribution also implies that respondents are providing a unique quantity of information in their responses that cannot be obtained through "statistical chance," or more specifically the assignment of responses based on a uniform distribution.

As a basic concept in information theory, entropy can be effectively applied to survey items with multiple choices responses (Cox, 1980; Shannon, 1948). Let $k = 1, \dots, K$ be the possible responses and $p_1, p_2, \dots, p_k$ be the probability of the responses that fall into category k for a given sample of survey data. Under the framework described above, the information entropy, or average amount of information in the response space $H(p)$ can be defined as:

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

The entropy function expressed in (1) is interpreted as both as a measure of uncertainty and as a measure of the amount (or quantity) of information. It is straightforward to show that the maximum of the entropy function is obtained when responses are uniformly distributed; that is $p_1 = p_2 = \cdots = p_k = 1/k$. When entropy is at its maximum, the unique quantity of information provided by respondents is minimized. Movement away from maximum entropy concomitantly implies that responses contain greater quantities of unique information.

Extensions of the basic information entropy measure described by (1) have been proposed in literature (Dahl & Osteras, 2010; Friesner et al., 2013; Friesner et al., 2016; Friesner et al., 2021; Friesner et al., 2022) to address practical considerations. For instance, a typical survey might ask an individual to respond to several survey items. Each item in the survey might ask participants to select response from a series of K mutually exclusive and collectively exhaustive categories. One possible method to analyze the information content in this type of survey is to aggregate the responses for all survey respondents, and analyze the distribution of empirical probabilities $\mathrm{p_k}s$ for that question, and possibly across a series of survey items. Schibik et al. (2012), for example, examine the quantity of information contained in student rating of instruction. In this study, 118 students rated their professors over 7 survey items. For each survey item, students rated their professor on a 1 to 5 scale, with a score of 1 indicating strongly disagree and 5 indicating strongly agree. A relative frequency distribution of responses was calculated for a given survey item based on the 188 student responses. The relative frequencies were combined using (1) to calculate the information entropy of each survey question. Comparing entropy values across survey items allowed for a relative assessment of which survey items contain a greater quantity of information than others, and by extension provided inferences about which student ratings of instruction questions are more informative than others.

Friesner et al. (2022) apply information entropy to determine whether the same survey items, when administered repeatedly over time to similar populations, provide varying quantities of information. They found that the quantity of information increased slightly over repeated survey administrations. This indicates that valid and reliable surveys, when administered repeatedly, should exhibit "dynamic stability."

Friesner et al. (2021) develop a methodology to evaluate whether the inclusion or exclusion of a survey item in a construct changes the quantity of information provided by respondents. This gives researchers an additional tool to help ensure that any set of survey items comprising a construct are designed efficiently (i.e., a smaller number of survey items can be used to accurately and precisely characterize a construct). A limitation of this manuscript is that it assumes that all survey respondents interpret survey items and response scales in a similar fashion. They make no attempt to account for whether specific groups of respondents interpret and respond to survey items differently from other groups of respondents. The current manuscript represents an initial attempt to address this limitation.

An alternative approach was posited by Friesner et al. (2016), who (within the context of a single survey administration) examined the distribution of responses to a series of survey questions (which presumably measure the same underlying construct) provided by a single respondent. This facilitates relative assessments about which groups of respondents provide relatively more or less information in their responses for that construct. If different groups of respondents provide statistically different quantities of information, then it may be necessary to adjust sampling designs and/or methods of data analysis to account for these differences.

To express this idea within the context of information entropy, let $i = 1, \dots n$ denote the observations/respondents, $l = 1, \dots L$ denote the number of survey items in a given scale, and $k = 1, \dots, K$ denote the number of possible options that a respondent can select when answering a survey item (i.e., yes/no questions yield $k = 2$). Define $p_{ik} = \frac{\sum_{l=1}^{L} D_{kli}}{L_i}$ for each $k = 1, \dots K$, where D is a binary indicator that gives a value of 1 if respondent i gives response k for survey item 1, and a zero otherwise. Lastly, let L_i represent the total number of survey items answered by i^{th} respondent. Given these definitions, the extended entropy measure can be calculated over n respondents in a given sample as:

$$H_i(p) = -\sum_{k=1}^{K} p_{ki} log_2(p_{ki}) \tag{2}$$

The expression in (2) can be interpreted as a percentage of information captured in the scale by each respondent. Under the null hypothesis of no difference in the quantity of information, analysis of variance (whether parametric or nonparametric) or maximum-likelihood-based regression analysis can be used to assess differences in information entropy across groups of respondents. Friesner et al. (2016) specified a reduced form, linear in parameters Tobit regression[1] with the following form:

$$h_i(p) = \left(\frac{H_i(p)}{max(H_i(p))}\right) = \alpha + \sum_{q=1}^{Q} \beta^q X_i^q + \varepsilon_i \tag{3}$$

where X_i^q is one of the Q variables characterizing a respondent's incentives for $q = 1, \dots Q$; α denotes the intercept; β^q denote the slope parameters for the corresponding Q variables; ε_i is the random error term. Note that normalizing the entropy measure (i.e., dividing the entropy measure by its theoretical maximum) to the unit interval is not required, but is done to facilitate the estimation of a Tobit model with very clear and consistent points of censoring for any set of survey items to which the entropy calculation is applied.

Under the null hypothesis, the slope parameters β^q should be individually (H_0: $\beta^q = 0$ for each $q = 1, \dots Q$) or jointly (H_0: $\beta^i = \beta^j = 0,$ for some $i \neq j$) zero at an acceptable (i.e., 5 percent) level of significance. Tests on the individual slope parameter can be implemented using the standard t-test, while joint hypothesis tests can be evaluated using a chi-square test. As noted earlier, greater entropy levels implies a lower quantity of information, and vice versa. Hence, negative coefficient values indicate that a specific type of respondent provides a greater quantity of information, holding the other specified regressors constant (and vice versa).

Information Entropy as a Quality Control Tool in Survey Design

The Friesner et al. (2016) study was predicated on the assumption that all of the survey items used to create the entropy measure were appropriate (i.e., valid and reliable) measures of the same underlying construct. If that is accurate, then it is entirely reasonable to assume that differences in entropy across different respondent groups could be attributed to issues in the survey's administration and/or sampling design. However, the *complement* to this argument is *also* logically consistent. Assuming that a survey employs an appropriate experimental design, sampling design, and plan of analysis, then changes in the wording of survey items, or changes in response scales (or both), may cause different groups of respondents to interpret those survey items differently, and by extension respond differently to those survey items. In fact, the latter is *more fundamental* to survey research precisely because the design of survey items and/or response scales occur prior to, and often inform, elements of survey administration. More importantly, the stability reliability (or test-retest reliability) of survey instruments may change over time (Nieswiadomy, 2012, pp. 169-170), especially in situations where survey may be applied to different populations, or when major changes in population characteristics occur. This necessitates continual reassessment and adaption of the survey, which in terms requires continuously assessing the survey's appropriateness in an efficient and effective manner.

Fortunately, the juxtaposition of assumptions does not change the basic empirical methodology employed by Friesner et al. (2016), and we retain their methodology for the current analysis. More specifically, we assume that all facets of a survey's initial design, implementation, and analysis are appropriate. We further assume that the study's population is unchanged. Holding all else constant, changes are made to the survey's items or response scales (or both). Given these assumptions, we operate under the following null and alternative hypotheses:

$\boldsymbol{H_0}$: *No mean differences in information entropy exist across different groups of respondents.*

$\boldsymbol{H_A}$: *Mean differences in information entropy exist across different groups of respondents.*

Essentially, the null hypothesis posits that the revised survey did not change the quantity of information provided by specific groups respondents relative to other groups of respondents. As such it *may* be appropriate, and provide valid and reliable responses. Rejecting the null hypothesis indicates that the survey revisions produced information differences across respondents, and that the survey has a quality control issue (which may be a survey item with a low quantity of information which should be eliminated from the survey, an item whose quantity of information is positive, but which varies by respondent group, a survey administration error, or some combination of these issues) which must be addressed prior to its administration in a full-scale research study.

Operationally, testing the null hypothesis proceeds as described previously. Information entropy is calculated via equation (2). A reduced form, linear in parameters Tobit model as specified in (3) can be estimated, and t-tests and/or chi-square tests can be used to evaluate whether statistically significant differences in information entropy exist across respondents. All tests employ a 5 percent significance level, although statistical significance at the 10 percent level is reported for the interested reader.

The Natural Experiment

The setting for the natural experiment is the 2018 Hoopfest Basketball Tournament, which is the largest 3-on-3 amateur basketball tournament in the world. Held on the streets of Spokane, Washington, the tournament hosted more than 6,000 teams and 225,000 people (https://www.spokanehoopfest.net/). The event has been held each year since 1990, and over the past decade, annual tournament attendance routinely approaches or exceeds 200,000 (Schnell, 2014). As a sporting event, assessing customer experience is a vital component of the tournament's success, and Hoopfest organizers have worked with experts to develop and implement a (paper-based) satisfaction survey that has been assessed repeatedly using established methodologies and found to provide valid and reliable results (Bozman et al, 2010; Kurpis et al., 2010). A full set of information about the survey, inclusive of inclusion criteria, can be found in Bozman et al. (2010), Kurpis et al. (2010), and Friesner et al. (2016).

The core items in the survey, as well as its methods of administration, are unchanged over time. Taken collectively, these core items ask respondents to report on their basic demographic characteristics, economic expenditures during the tournament, and those psychological factors that might influence satisfaction derived from attending the event. The survey has also evolved over time to mirror changes in the event. For example, the Hoopfest staff developed a mobile app for the event, and a survey item was developed to assess use of the app. In 2006, Hoopfest staff added the well-known list of values (LOV) construct to the survey (Batra et al., 2001; Beatty et al., 1985; Homer & Khale, 1988; Kahle, 1983; Kahle et al., 1986). The LOV is comprised of nine items that characterize how social affiliation values impact daily decisions. The items are states in a sufficiently simple and general manner that virtually any respondent who met the study's inclusion criteria could provide informative responses to those items. Respondents are asked to rate each LOV item on a nine-point scale, with a value of one indicating "very unimportant" and nine indicating "very important." The 2013 version of the survey also added a validated two-item construct - the Basic Empathy Scale (Carre, et al., 2013; Jolliffe & Farrington, 2006) - that asked respondents to describe their ability to empathize with others. As with the LOV, the empathy items were worded in a manner that was accessible to the vast majority of survey respondents who met the study's inclusion criteria. Both items use the same, five-point response scale, with values of one indicating "strongly disagree" and 5 indicating "strongly agree."

In 2018, Hoopfest organizers adapted the survey in a manner that created a natural experiment. More specifically, six survey items were added to the empathy construct. Each of these items used the same five-point response scale as the original empathy items. Two of these items ("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") address cognitive aspects of empathy. An additional two items ("I feel happy when I see smiles on other people's faces" and "I am sad when I observe someone in distress") address affective aspects of empathy. The final two items ("I get agitated when I see someone in distress" and "I help others when I see they need help") address behavioral aspects of empathy.

While the six survey items were thoughtfully designed, the items were not subjected to the full set of reliability and validity tests typically recommended by the literature. Concomitantly, the LOV was included in the survey, but was unchanged from the literature and previous iterations of the survey. Survey administration techniques were also largely unchanged. Thus, the LOV likely retains most, if not all, of its properties since little has changed in its use. So, if one calculates an information entropy measure based on equation (2) for the LOV items, and uses it as the basis for estimating a Tobit regression consistent with equation (3), it is unlikely that statistically significant differences exist in information entropy (and, by extension, the quantity of information provided) across groups of survey participants. The study's null hypothesis is unlikely to be rejected.

On the other hand, the properties of the eight empathy items, taken collectively, may differ substantially from those properties established by the literature for the two-question empathy measure, especially if those survey items are poorly designed and exhibit quality control issues. Calculating an entropy measure based on equation (2) for each respondent across the eight empathy items and used in a regression analysis consistent with equation (3), may lead to significant differences in entropy across respondent groups, especially if those survey items are poorly designed. Therefore, it is more likely that the study's null hypothesis will be rejected for one or more coefficient estimates in the information entropy regression based on the eight empathy items.[2]

DATA

As noted above, the current study analyzes the results of the 2018 survey. Hoopfest personnel collected and electronically coded the data, de-identified the data, and subsequently provided the de-identified data to the study's authors. Because the data is de-identified and the researchers have no ability to link responses to individuals, the research is not considered human subjects research and intuitional review board approval is not required for this study. Given a population of over several hundred thousand individuals, 5 percent sampling error, and a conservative (i.e., 50/50) effect size, Dillman (2000, pp. 207) suggests that a sample size of approximately 385 individuals is sufficient to provide statistically meaningful estimates of the population parameters. Hoopfest organizers took a conservative approach and decided to randomly identify 500 individuals and invite them to participate in the survey. Of these 500 individuals, 437 agreed to participate in the survey, yielding a gross response rate of 87%. After eliminating respondents who did not respond to each survey item analyzed in this study, we are left with a working sample of 336 observations, which represents a net survey completion rate of 67%. Further details about the survey administration process can be found in Bozman et al. (2010), Kurpis et al. (2010), and Friesner et al. (2016).

The survey allowed for the construction of the two entropy variables: one for the nine LOV items (ENTROPY) and one for the eight empathy items (ENTROPY2). Each entropy variable is further normalized (by dividing the entropy variable by its theoretical maximum) to the unit interval (NENTROPY and NENTROPY2, respectively).

Respondents also provided information on several meaningful demographic, satisfaction, and economic variables, which were used to create several respondent-specific covariates. Demographic variables included respondent age in years (AGE) and a binary variable identify female respondents (FEMALE). Additionally binary variables were created to identifying the reason the individual attended Hoopfest, including playing in tournament (PLAY), as a spectator (WATCH), as a volunteer (VOLUN), or in another capacity (OTHROLE). Binary variables were also created to identify the amount of time that respondents spent at the tournament. HRS05 identifies individuals who attended Hoopfest for between 0-5 hours, HRS611 identifies those who spent 6-11 hours at Hoopfest, and HRS identifies those who spent 12 or more hours at the Hoopfest tournament. Lastly, since many people attend Hoopfest as a part of a group of family unit, the variable NPEOPLE was created to quantify the number of people who attended Hoopfest with the respondent.

Information was collected on economic activities and expenditures associated with Hoopfest. The binary variable OVERNGT identifies individuals who stayed away from their home for at least one night to attend Hoopfest. For those individuals, binary variables were also created to identify those who stayed overnight with extended family members (FAMILY), in a hotel (HOTEL), at a campground (CAMPGR), or in other accommodations (OTHACC). Individuals who purchased lodging accommodations was recorded using a binary variable (LODGPAY), as were self-reported lodging expenditures (LODGING). A binary variable (NFOOD) was used to identify respondents who purchased food at the tournament. The variable MEALS captures the number of meals purchased, while FOODCOST characterizes self-reported total food expenditures (number of meals x price per meal) at the tournament. Lastly, PURCH is a binary variable identifying participants who made other, relevant purchases at the tournament, while PURCH gives the self-reported expenditures on those items.[3]

Third, information was collected on the behavioral characteristics of respondents. Three binary variables indicate respondents who are very satisfied (VERYSAT) with the Hoopfest tournament, moderately satisfied (MODSAT), and less than moderately satisfied (OTHSAT). Three binary variables also capture those respondents who report that they will definitely attend Hoopfest next year, (DEFATT), probably attend (PROBATT), or all other responses (OTRATT). Lastly, a binary variable was used to identify respondents who downloaded the Hoopfest mobile app to their phone (MOBILEAPP).

RESULTS

Descriptive statistics are provided in Table 1. The mean entropy list of values item is 0.994 with the standard deviation of 0.775. When normalized, the mean level of entropy is 31.4% (or 0.314) of its maximum possible value, with a standard deviation of 24.5% (or 0.245). The mean level of entropy emotional connection items is 1.169 with the standard deviation of 0.617. When expressed as a proportion of total possible entropy, the mean is 0.292, with a standard deviation of 0.154.

The mean age of the respondents was 34.5 years of age and half (50%) of them were female. More than half (57.4%) of the respondents attended Hoopfest were identified as viewers, while 38.7% of the respondents participated as Hoopfest players. The remaining, relatively small proportion of the respondents participated as volunteers (3.3%) or in other ways (0.6%). Approximately 67.6% of the respondents participated Hoopfest between 6 and 11 hours, while 25.0% of them attended less than 5 hours. A small percentage (7.4%) of respondents attended for 12 hours or more.

More than half (62.2%) of the participated respondents lived within the areas of the host city, Spokane, WA, with the remaining 16.7% stayed overnight at a hotel, 19.9% stayed overnight with family, and 0.3%

stayed at a campground. The mean total lodging expenditures was $82.40, the mean total food expenditures was $96.90, and the mean total expenditures on other items was $ 57.90.

Among all the respondents in the sample, 71.7% of them expressed that they were very satisfied with the Hoopfest event, and 62.5% stated that they definitely intended to attend future Hoopfest tournaments. With regard to mobile application, 62.2% of the respondents indicated they downloaded it either from the Apple Store or Google Play.

TABLE 1
DESCRIPTIVE STATISTICS

Variable	**Description**	**Mean**	**Std. Dev.**
Dependent Variables			
ENTROPY	Entropy Calculation - List of Values Items	0.994	0.775
NENTROPY	Normalized (Proportional) Entropy Calculation - List of Values Items	0.314	0.245
ENTROPY2	Entropy Calculation – Empathy Items	1.169	0.617
NENTROPY2	Normalized (Proportional) Entropy Calculation - Empathy Items	0.292	0.154
Covariates			
AGE	Respondent Age in Years	34.497	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	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	215.261
LODGPAY	Binary Variable Identifying Respondents Who Pay for Lodging	0.188	
MEALS	Number of Meals Purchased	4.179	5.475
NFOOD	Binary Variable Identifying Respondents Who Purchased Food at Hoopfest	0.890	
FOODCOST	Total Food Expenditures	96.929	246.363
PURCH	Total Expenditures on Other Items	57.932	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 Tobit regression analysis for the normalized LOV entropy variable (NENTROPY). The chi-square statistic's probability value of 0.563 indicates that the set of regressors included into this model fails to explain a significant number of variations regarding the dependent variable, and the study's null hypothesis cannot be rejected. In other words, when taken collectively, the various groups of respondents as characterized jointly by the regressors, do not provide statistically significantly different quantities of information. Aside from the model's intercept and Tobit disturbance term, only one individual coefficient estimate – for the binary variable identifying those individuals who stayed overnight with family while attending the tournament - is statistically different from zero at the 5 percent level of significance (coefficient estimate = 0.116; p-value=0.022). Because the joint effects of the covariates are insignificant, we give primary consideration to the chi-square test results.

TABLE 2
TOBIT ANALYSIS OF LIST OF VALUES (LOV) ENTROPY MEASURE

Dependent Variable:	**NENTROPY**				
Regressor	**Coeff.**	**Std. Error**	**t-Statistic**	**P-value**	
Intercept	0.225	0.095	2.360	0.018	**
AGE	0.001	0.002	0.840	0.403	
FEMALE	0.062	0.042	1.480	0.140	
PLAY	-0.011	0.048	-0.230	0.820	
HRS05	0.003	0.046	0.070	0.948	
HRS12U	-0.062	0.076	-0.820	0.410	
NPEOPLE	0.002	0.005	0.340	0.732	
FAMILY	0.116	0.051	2.280	0.022	**
LODGPAY	-0.447	0.279	-1.600	0.109	
ln(LODGING)	0.090	0.048	1.870	0.061	*
MEALS	0.002	0.004	0.420	0.674	
NFOOD	0.026	0.107	0.240	0.811	
ln(FOODCOST)	-0.009	0.024	-0.360	0.719	
ln(PURCH)	-0.015	0.027	-0.570	0.569	
PURCHDV	0.019	0.113	0.170	0.864	
MODSAT	-0.030	0.046	-0.670	0.506	
OTHSAT	0.022	0.178	0.120	0.902	
PROBATT	-0.001	0.045	-0.030	0.979	
OTHATT	-0.039	0.084	-0.460	0.644	
MOBILEAPP	-0.068	0.041	-1.640	0.101	
Tobit Disturbance Term	0.319	0.016	20.180	<0.001	**
Log-Likelihood Function			-160.387		
Restricted Log-Likelihood Function			-169.085		
Chi-Square Test Statistic Value			17.39722	0.563	
Degrees of Freedom			19		
Number of Observations			336		

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 results of the Tobit regression analysis for the entropy variable based on eight empathy items (NENTROPY2). The results for Table 3 contrast starkly with the previous regression results. In Table 3, the chi-square test statistic's probability value is less than 0.001, indicating the regressors explained a significant proportion of variations in the empathy measure. Within the context of the natural experiment, this implies that, when taken jointly, specific groups of respondents provide fundamentally different quantities of information in their responses to the empathy items. The study's null hypothesis is rejected.

Analysis of the individual coefficient estimates in Table 3 provides inferences about which types of respondents are providing fundamentally different quantities of information for the empathy items. Seven coefficient estimates are statistically different from zero at the 5% level of significance or better. Respondents who attended Hoopfest 5 hours or less (HRS05) potentially provided less informative responses (as indicated by entropy values that are closer to the theoretical maximum value) than those who attended between 6 and 11 hours (coefficient=3.61; p-value<0.001), holding the other specified regressors in the model constant. Those respondents who participated 12 hours or more (HRS12U) were more likely to provide greater quantities of information – again, as expressed by entropy values that are further from the maximum theoretical value (coefficient estimate= - 2.05; p-value=0.041), holding all other regressors in the model constant.

Those individuals who paid for lodging (LODGEPAY) likely provided more informative responses than those who did not pay for lodging (coefficient estimate= -0.294; p-value=0.035), holding the other regressors in the model constant. Among those respondents who stayed with lodging, the greater the lodging expenditures (ln(LODGING)), the more likely they provided more quantities of information (coefficient estimate=0.054; p-value=0.025). Similar differences in the quantity of information are observed from respondents who purchase other items (PURCHDV) (coefficient estimate= - 0.112; p-value=0.048), holding the effects of the other specified covariates constant.

Regarding behavioral characteristics, those respondents who report being moderately satisfied (MODSAT) provided significantly more informative responses (coefficient estimate= - 0.045; p-value=0.05) compared to the omitted category (VERYSAT) and holding all other model regressors constant. Lastly, those respondents who expressed a probable intention to attend Hoopfest next year (PROBATT) provided significantly higher quantities of information (coefficient estimate= - 0.06; p-value=0.007) relative to the omitted category (VERYSAT).

TABLE 3
TOBIT ANALYSIS OF EMPATHY ENTROPY MEASURE

Dependent Variable:	**NENTROPY2**				
Regressor	**Coeff.**	**Std. Error**	**t-Statistic**	**P-value**	
Intercept	0.238	0.048	4.980	<0.001	**
AGE	0.000	0.001	0.270	0.786	
FEMALE	0.041	0.021	1.930	0.054	*
PLAY	-0.023	0.024	-0.960	0.340	
HRS05	0.082	0.023	3.610	<0.001	**
HRS12U	-0.078	0.038	-2.050	0.041	**
NPEOPLE	0.003	0.002	1.260	0.209	
FAMILY	0.045	0.025	1.750	0.079	*
LODGPAY	-0.294	0.140	-2.110	0.035	**
ln(LODGING)	0.054	0.024	2.240	0.025	**
MEALS	0.002	0.002	1.000	0.318	
NFOOD	0.059	0.053	1.100	0.271	
ln(FOODCOST)	-0.010	0.012	-0.780	0.435	
ln(PURCH)	0.023	0.013	1.750	0.080	*
PURCHDV	-0.112	0.056	-1.980	0.048	**
MODSAT	-0.045	0.023	-1.960	0.050	**
OTHSAT	-0.157	0.091	-1.730	0.083	*
PROBATT	-0.060	0.022	-2.690	0.007	**
OTHATT	-0.011	0.042	-0.260	0.798	
MOBILEAPP	-0.002	0.021	-0.100	0.920	
Tobit Disturbance Term	0.163	0.007	22.950	<0.001	**
Log-Likelihood Function			57.866		
Restricted Log-Likelihood Function			28.639		
Chi-Square Test Statistic Value			58.454	<0.001	**
Degrees of Freedom			19		
Number of Observations			336		

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

DISCUSSION AND CONCLUSIONS

The results of this study are twofold. First, data culled from a natural experiment suggest that information entropy can be used as an initial quality control tool in survey design and administration. Well-designed survey items do not lead to different groups of respondents providing fundamentally different quantities of information in their responses. Poorly designed survey items do lead to statistically significant differences in the quantity of information provided by different groups of respondents. Information entropy is relatively easy to calculate using Excel, and can be subjected to various forms of analysis of variance or regression analysis, making it accessible to wide range of individuals (not just survey design specialists, psychometricians, etc.) who may design surveys as a part of their daily job duties. Individuals who are developing and pilot testing surveys are therefore encouraged to use information entropy as a “first-pass” screening device to save time and effort. If differences in the quantity of information provided by respondents exist, then the survey designer has reason to suspect that the survey has a flaw that must be addressed. If no differences in the quantity of information exist across respondents, then it is worthwhile to proceed to other, more advanced assessment techniques to establish the survey’s reliability and validity.

A second conclusion that can be drawn from this study applies to the Hoopfest tournament, which forms the context for our study. Large entertainment events such as Hoopfest rely heavily on economies of scale to generate economic profit. They also require several months of planning to execute successfully. Consumer behavior theories unambiguously predict that as participants’ tastes and references align with a product or service, and as familiarity with a product or service grows, they are able to more accurately and precisely gauge the utility of that product or service (Hanna & Wozniak, 2001, pp. 101-129). If the individual’s experiences positively align with her/his tastes and preferences (and by extension, his/her utility) relative to other alternatives, the more likely the individual will (perhaps repeatedly) purchase the good or service. As the number of participants increases (if the event is initially successful), the more varied are the perceptions of the experience. This wider array of preferences must be appropriately characterized and incorporated into the next event planning process months in advance of the next event. This places a significant onus on Hoopfest’s organizers to ensure that the survey is designed appropriately. Moreover, this study found that customers gave significantly different quantities of information to those poorly designed questions based on economic factors, demographic factors, and behavioral factors. And as perceptions of the experience becomes more varied (and each of those varied perceptions is based on demographic, economic and behaviors factors), it becomes disproportionately more difficult for the survey designer to assess and address survey flaws. This, in turn, makes it increasingly difficult to draw meaningful conclusions from those survey items.

While the current manuscript provides some interesting findings, it is an exploratory case study, and its results should be viewed with caution. One major limitation of this study is that, while we provide evidence suggesting that information entropy may be useful as an initial quality control tool in survey assessment, we have not provided a formal framework to explain how entropy informs and supports other elements of survey assessment. That is, if differences in information entropy exist across respondents, it is reasonable to conclude that a flaw is present. However, information entropy does not identify the specific flaw that must be corrected. Concomitantly, if no differences in information entropy exist, then it is still necessary to use other methods to assess a survey’s reliability, validity, and general appropriateness. We merely provide an empirical illustration showing that entropy might be useful as a survey assessment tool. Establishing formal methodological links between the use of information entropy and other methods of assessment would substantially improve the literature’s understanding of the utility of information entropy as a tool in survey assessment.

A second limitation is that the data are drawn from a single case study or a very large sporting event. This case study had the luxury of (quite reasonably) assuming that the survey was (with the exception of the six new empathy items) designed and administered reasonably well, and that the sample size was sufficient to conduct a meaningful statistical analysis. Many survey assessments are conducted within the context of pilot studies, which use smaller sample sizes. It is unclear in these circumstances whether sufficient statistical power exists in the data to employ information entropy in a manner utilized in the current manuscript.

Related to this point, a third limitation of the study is that it uses a customer satisfaction survey drawn from a large sporting event. Many surveys do not assess customer satisfaction, or address customer satisfaction for very different products and/or services than a sporting event. Future research is necessary to replicate the current study within the context of those surveys and markets to ensure that our conclusions are generalizable to these broader contexts.

Lastly, the empirical analysis uses covariates drawn from an existing survey, and adopts an econometric method (i.e., Tobit regression with a reduced form, linear in parameters response function) that was used in a previous study. If the survey does not provide a full set of covariates, the regression analysis will suffer from omitted variable bias. If the response function is truly nonlinear, or if the assumption of a censored normal distribution (which underlies the Tobit model) is inappropriate, our empirical results will be biased. Future research is necessary to identify a broader and more generalizable means to assess differences in information entropy across groups of respondents, in order to make full use of information entropy as a survey quality control tool.

ENDNOTES

1. While a Tobit regression is a natural choice to estimate (3), it is not the only viable regression method. One could also estimate nonparametric regression analysis, including (but not limited to) quantile regression methods. This manuscript utilizes the Tobit model for two reasons. The first is to facilitate consistency and comparability with Friesner, Valente, and Bozman (2016). The second is that the use of nonparametric methods creates information loss, as the focus is on ranking ordering of the dependent variable across observations, not the magnitude of the dependent variable itself.
2. Note that as the number of survey items increases (holding all else constant), the shape and values of the entropy function (including its maximum) change. The distribution also more closely approximates a more continuous distribution. With only two survey items, there are only three possible (whether normalized or non-normalized) entropy values: 0, ¼, and 1. With eight survey items (and five possible responses), the number of possible normalized entropy values increases nonlinearly. Thus, it is inappropriate to directly compare mean differences between the information entropy measure for the two-item empathy construct and its eight-item analog.
3. To reduce the presence of heteroscedasticity in the regression results, the natural logarithm transformation is applied to the LODGING, FOODCOST and PURCH variables.

REFERENCES

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.

Churchill, G. (1979). A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research*, *16*(1), 64–73.

Cox, E.P. (1980). The optimal number of response alternatives for a scale: A review. *Journal of Marketing Research*, *17*(4), 407–422.
Dahl, F., & Osteras, N. (2010). Quantifying information content in survey data by entropy. *Entropy*, *12*(2), 161–163.
Dillman, D.A. (2000). *Mail And Internet Surveys: The Tailored Design Method.* New York, NY: John Wiley and Sons.
Fain, J. (2004). *Reading, Understanding, And Applying Nursing Research* (3rd ed.). Philadelphia, PA: F. A. Davis Company.
Friesner, D., Bozman, C., McPherson, M., & Valente, F. (2022). On the dynamic stability of the information entropy measure in survey research. *Journal of Business and Economic Perspectives*, *49*(1), 97–117.
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., Khayum, M., & Schibik, T. (2013). Characteristics of the information content in business sentiment surveys. *American Journal of Business*, *28*(1), 19–37.
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.
Hair, J., Black, W., Babin, B., Anderson, R., & Tatham, R. *Multivariate data analysis* (6th Ed.). Upper Saddle River, NJ: Pearson/Prentice Hall.
Hanna, N., & Wozniak, R. (2001). *Consumer Behavior: An Applied Approach.* Upper Saddle River, NJ: Prentice Hall.
Homer, P.M., & Kahle, L.R. (1988). A structural equation analysis of the value-attitude-behavior hierarchy. *Journal of Personality and Social Psychology*, *54*(3), 638–646.
Innamorati, M., Ebisch, S., Gallese, V., & Saggino, A. (2019). A bidirectional measure of empathy: Empathic experience scale. *PLOS One*, *14*(4), e0216164. https://doi.org/10.1371/journal.pone.0216164
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.
Jolliffe, D., & Farrington, D.P. (2006). Development and validation of the basic empathy scale. *Journal of Adolescence*, *29*, 589–611.
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.
Nieswiadony, R. (2012). *Foundations Of Nursing Research.* Upper Saddle River, NJ: Pearson.
Schibik, T., Khayum, M., & Friesner, D. (2012). The quantity of unique information in student ratings of instruction. *Atlantic Economic Journal*, *40*(2), 221–223.
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.
Shannon, C.E., & Weaver, W. (1949). *The Mathematical Theory of Communication*. Urbana: University of Illinois Press.
Smith, S.M., & Albaum, G.S. (2005). *Fundamentals of marketing research.* Thousand Oaks, CA: Sage.
Spreng, R.N., McKinnon, M., Mar, R., & Levine, B. (2009). The Toronto empathy questionnaire: Scale development and initial validation of a factor analytic solution to multiple empathy measures. *Journal of Personality Assessment*, *9*(1), 62–71.