

# Material Incentives and Other Potential Factors Associated With Response Rates to Internet Surveys of American Evaluation Association Members: Findings From a Randomized Experiment

American Journal of Evaluation  
2020, Vol. 41 (2) 277-296  
© The Author(s) 2019  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/1098214018818371  
journals.sagepub.com/home/aje



Chris L. S. Coryn<sup>1</sup>, Lyssa W. Becho<sup>1</sup>, Carl D. Westine<sup>2</sup>,  
Pedro F. Mateu<sup>3</sup> , Ruqayyah N. Abu-Obaid<sup>1</sup>,  
Kristin A. Hobson<sup>4</sup>, Daniela C. Schröter<sup>1</sup>, Erica L. Dodds<sup>1</sup>,  
Anne T. Vo<sup>5</sup>, and Mary Ramlow<sup>1</sup>

## Abstract

Internet surveys of American Evaluation Association (AEA) members are a common method for studying evaluation practice. Response rates obtained from Internet surveys of AEA members are, however, frequently very small. To investigate whether or not material incentives increase response rates to Internet surveys of AEA members, a between-subjects three-treatment and one control randomized experiment in which a randomly selected sample of AEA members were randomly assigned to a no-incentive control condition, lottery condition, token incentive condition, or philanthropic donation incentive condition was utilized. The overall response rate to the survey was 39.66% and the response rates for each of the four conditions were control = 36.24%, lottery = 44.39%, token incentive = 43.28%, and philanthropic donation = 34.67%, respectively. The cost-effectiveness of each of the four conditions also was examined, demonstrating that the lottery was the most cost-effective. Other factors potentially influencing response or nonresponse decisions also are discussed.

## Keywords

research on evaluation, survey response rates, survey incentives, American Evaluation Association

---

<sup>1</sup> Western Michigan University, Kalamazoo, MI, USA

<sup>2</sup> University of North Carolina at Charlotte, Charlotte, NC, USA

<sup>3</sup> Universidad del Pacífico, Lima, Peru

<sup>4</sup> Indiana University, Bloomington, IN, USA

<sup>5</sup> University of Southern California, Los Angeles, CA, USA

## Corresponding Author:

Chris L. S. Coryn, Western Michigan University, 1903 West Michigan Avenue, Kalamazoo, MI 49008, USA.

Email: [chris.coryn@wmich.edu](mailto:chris.coryn@wmich.edu)

## Background and Introduction

### *Research on Evaluation (RoE) and American Evaluation Association (AEA) Members*

RoE has increased considerably since Henry and Mark (2003) put forth an agenda for RoE and has more than tripled in the last decade (Coryn et al., 2017). Findings reported in many of these investigations have been derived from self-report surveys of AEA members. Response rates to these surveys have been, on average, only slightly more than 25% (e.g., Azzam & Szanyi, 2011; Fleischer & Christie, 2009; N. Galport & Azzam, 2017; Szanyi, Azzam, & Galen, 2013), with some as low as 15% (Seidling, 2015) or 16% (Lewis, Harrison, Ah Sam, & Brandon, 2015).<sup>1</sup> These estimates do not include AEA Topical Interest Group (TIG) surveys, which sometimes produce greater response rates (e.g., Preskill and Caracelli's [1997] survey of the AEA's Evaluation Use TIG, which resulted in a response rate of 54%), and also exclude surveys sometimes distributed through EvalTalk (the official listserv of AEA).

Surveys of AEA membership are important gateways into practice-related inquiry, but in order to be valuable, these surveys should adhere to sound methodological regimen in order to obtain samples that produce generalizable findings. As shown in Table 1, more than half (57%) of the authors of recently published surveys of AEA members claim to have used probability sampling methods, such as simple random samples, when in fact these descriptions are inaccurate (see notes b–d in Table 1). All consisted of one mode of communication with potential respondents, as well as one mode of survey administration, and only one sent a presurvey notification to potential respondents. Follow-up contacts used in the seven investigations ranged from one to four. In their meta-analysis of response rates to Internet surveys, Cook, Heath, and Thompson (2000) report that presurvey notifications and number of follow-up contacts are among the dominant factors affecting response rates. Moreover, none provided a material incentive to encourage responses, instead, seemingly relying on the theory of leverage salience. Leverage-salience theory posits that that people vary in which aspects of a survey request they find attractive, which unattractive, and which are of no particular relevance to their decision to participate (see Seifert, 2008, for a detailed discussion of leverage-salience theory and survey response).

In addition to published articles reporting findings from surveys of AEA members, many unpublished masters' theses and doctoral dissertations have utilized AEA members as their target population. As part of a three-paper doctoral dissertation, Ozeki (2016), for example, conducted two independent surveys of AEA members and obtained response rates of slightly less than 25% for each sample (24.88% and 24.65%, respectively). If the field of evaluation continues to rely on surveys of AEA members to build an empirical base of practice, it is, therefore, important to understand potential factors for increasing rates of response.

### *Factors That Influence Survey Response*

Much is known about the effects of various types of incentives on rates of response in the context of mail, face-to-face, and telephone surveying, particularly in the context of general population surveys. In essence, response rates increase as incentives increase (Boulianne, 2008; Church, 1993; Singer, Van Hoewyk, Gebler, Raghunathan, & McGonagle, 1999; Yu & Cooper, 1983); though not always linearly (Armstrong, 1975; Fox, Crask, & Kim, 1988). In most populations, redeemable incentives (e.g., gift cards) appear to be the most effective incentive for long questionnaires, while lotteries are more effective for short surveys, monetary incentives increase response rates more than gifts, and prepaid incentives increase rates of response more than promised incentives or lotteries (Deutskens, de Ruyter, Wetzels, & Oosterveld, 2004). Less, however, is known about the effectiveness of incentives when used for Internet surveys and much of what is known is often conflicting (Bosnjak & Tuten, 2003; Fan & Yan, 2010). Internet surveying, contrasted with mail and face-to-

**Table 1.** Characteristics of Recent Surveys of American Evaluation Association (AEA) Members.

Author(s)	Year	Sample Design	Response Rate	Mode of Communication	Mode of Administration	Presurvey Notification <sup>a</sup>	Duration of Administration <sup>a</sup>	Number of Follow-Ups <sup>a</sup>	Material Incentive <sup>a</sup>
Fleischer and Christie	2009	Census	29.81%	E-mail	Internet	No	4 Weeks	1	No
Azzam and Szanyi	2011	Simple random sample <sup>b</sup>	21.70%	E-mail	Internet	No	Unreported	Unreported	No
Szanyi, Azzam, and Galen	2013	Census	28.80%	E-mail	Internet	No	Unreported	1	No
Lewis et al.	2015	Census	16.47%	E-mail	Internet	No	2 Weeks	1	No
Seidling	2015	Simple random sample <sup>c</sup>	15.31%	E-mail	Internet	No	4 Weeks	1	No
Coryn et al.	2016	Simple random sample	43.94%	E-mail	Internet	Yes	6 Weeks	4	No
Galport and Azzam	2017	Simple random sample <sup>d</sup>	20.21%	E-mail	Internet	No	3 Weeks	2	No

<sup>a</sup>Presurvey notification, duration of administration, number of follow-ups, and material incentive are either estimates/inferences (as could be derived from the published studies or as derived directly from queries sent to the study author[s]). <sup>b</sup>As noted by Azzam and Szanyi (2011), "Approximately 1,500 evaluators were randomly selected from the American Evaluation Association (AEA) 2004 directory as the sample" (p. 135). However, no a-priori sample size (as derived from a predetermined rate of sampling error) was reported. The sample, therefore, is better classified as a nonprobability sample. <sup>c</sup>Although Seidling (2015) indicates that "... the sampling frame was immense and appropriate" (p. 89) and that the "... author received from the AEA administrative office a random sample of 1,000 out of the 5,558 AEA members residing in the United States" (p. 90), no a-priori sample size (as derived from a predetermined rate of sampling error) was provided and in addition it is unclear if those selected to participate were truly selected at random (the sample was selected by the AEA, and it is not clear how it was selected). The sample, therefore, is better classified as a nonprobability sample. <sup>d</sup>Although Galport and Azzam (2017) indicate that "... a simple random sample of 1,994 members of the AEA [was taken] ..." (p. 83), no a-priori sample size (as derived from a predetermined rate of sampling error) was provided and in addition it is unclear if those selected to participate were truly selected at random (the sample was selected by the AEA, and it is not clear how it was selected). The sample, therefore, is better classified as a nonprobability sample.

face surveys, poses logistical challenges as regard the use of material incentives. It is widely agreed that if incentives are to increase response, they should be given in advance instead of being made contingent on the completion of a questionnaire, as is often the case with Internet surveys (Dillman, Smyth, & Christian, 2014). Yet providing advance material incentives in the context of Internet surveys is often simply not feasible or extremely cumbersome (Göriz, 2006). Although much is known regarding factors that influence response rates, there is clearly room for further investigation so as to potentially reduce ambiguity and uncertainty as to how to effectively design and administer incentives structures for surveys of AEA members.

### ***Rationale for, Positioning, and Significance of the Current Investigation***

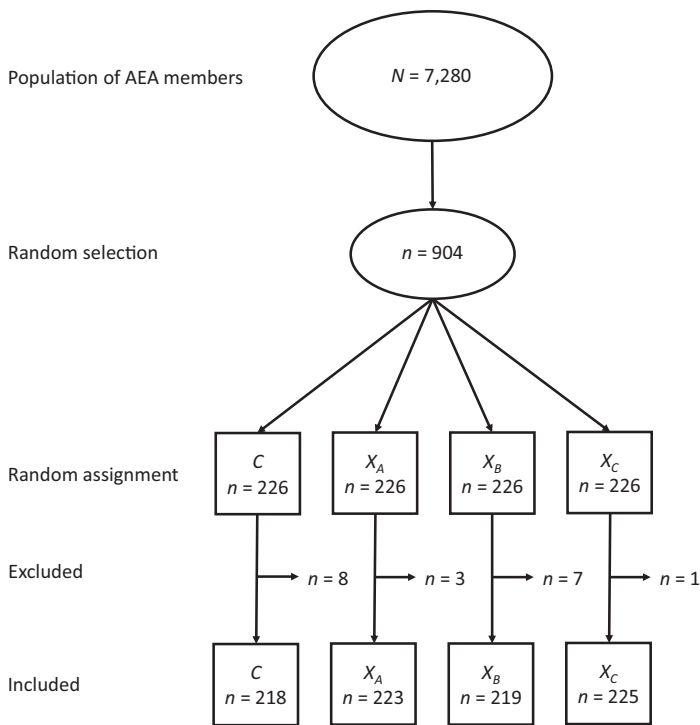
Demonstrably, RoE matters and if findings from RoE are assimilated into evaluation theories, methods, and practices "... evaluators could potentially produce better quality, more useful, and higher impact evaluations" (Coryn et al., 2016, p. 169). In the absence of findings grounded in high-quality empirical evidence, however, the potential influence of RoE is diminished and not likely warranted (Mark, 2007). In many ways, evaluation and the professional practice of evaluation is an "evidence-seeking" endeavor, and inquiry into its theories, methods, and practices should reflect this very same attribute. Much of what is known about evaluation practice has been derived from surveys of AEA members (Coryn et al., 2017; M. Galport & Galport, 2015). These surveys frequently produce less than optimal response rates. These poor response rates, sometimes coupled with less than optimal sampling designs and procedures, severely reduce the representativeness of observed samples and inflate standard errors. As a result, any reasonable generalizations and inferences from these samples to the population which they are intended to reflect as well as the precision of estimates of population parameters are often restricted or simply unknowable (Levy & Lemeshow, 2013). Logically, and of necessity, if RoE is to meaningfully influence evaluation practice, RoE scholars need to investigate potential factors that influence response rates, including the use of material incentives for Internet surveys of AEA membership, which have been demonstrated to increase response rates to surveys in other professional organizations (Rogelberg, 2006). Although the factors that influence survey response are numerous and vary across different populations, AEA members are unique in that they often conduct surveys themselves as part of the professional work and, therefore, they should theoretically also be willing to respond to surveys; in particular, those that pose questions about their practice.

### **Study Objectives and Questions Investigated**

Given the importance of AEA members as a primary source of information for studying evaluation practices, exploring potential methods for and factors associated with increasing response rates to Internet surveys of AEA members was the primary objective of the study. The focal question investigated was:

- (1) Do material incentives increase response rates to Internet surveys of AEA members?
  - a. Are there statistical and/or meaningful differences in response rates to Internet surveys of AEA members when a lottery, token incentive, philanthropic donation, or no incentive are used?
  - b. What traits or characteristics of AEA members, if any, moderate differences in response rates to Internet surveys of AEA members?
  - c. Which of the material incentive types is most cost-effective (i.e., lottery, token incentive, or philanthropic donation)?

Simultaneously, the following secondary questions were addressed as part of the investigation:



**Figure 1.** Flowchart of selection and assignment process.  $X_A$  = lottery,  $X_B$  = token incentive,  $X_C$  = philanthropic donation, and  $C$  = control.

(2) What other factors, if any, positively and/or negatively influence response rates to Internet surveys of AEA members?

Expressed in the form of a null and alternative, directional hypothesis of superiority, the hypothesis related to providing material incentives to AEA members for responding to Internet surveys was:

$$\begin{aligned}
 H_0 &: p_{X_A} = p_{X_B} = p_{X_C} = p_C, \\
 H_A &: p_{X_A} > p_{X_B} > p_{X_C} > p_C,
 \end{aligned}$$

where  $p$  represents a proportion (i.e., a response; the focal outcome),  $X_A$  represents the lottery condition,  $X_B$  represents the token incentive condition,  $X_C$  represents the philanthropic donation incentive condition, and  $C$  represents control (i.e., no incentive). The specific incentives associated with each of the conditions are described in the “Procedures” section. Throughout the remainder of this article,  $X_A$  is referred to as “lottery,”  $X_B$  as “token incentive,”  $X_C$  as “philanthropic donation,” and  $C$  as “control” (with the exception of Figure 1).

## Method

### Design

To investigate whether or not incentives (of varying types) increase response rates to Internet surveys of AEA members, a between-subjects three-treatment and one control randomized experiment (i.e., a four-group experiment) in which a randomly selected sample of AEA members (see

“Sample” section) were randomly assigned to the control, lottery condition, token incentive, or philanthropic donation conditions utilized.

For the design, statistical power was estimated to reject the null hypothesis ( $H_0$ ) under the following alternative hypothesis ( $H_A$ ): (1) for the control the response rate would be  $p = .30$ , (2) for philanthropic donation, the response rate would be  $p = .40$ , (3) for token incentive, the response rate would be  $p = .50$ , and (4) for the lottery, the response rate would be  $p = .60$ . Under this distribution, the predetermined effect sizes of  $p = .30$ ,  $.40$ ,  $.50$ , and  $.60$ , with statistical power ( $\pi = 1 - \beta$ ) =  $.80$ , produced a total necessary sample size of  $n = 904$  to detect any statistically and meaningful differences in the effect (i.e., minimum detectable effect size) associated with the various incentives, with  $n = 226$  randomly assigned to each of the four conditions (see “Sample” section).

### Sample

In November 2016, following an application procedure with and approval from the AEA Executive Board, e-mail addresses and secondary information relevant to the study (e.g., gender, ethnicity) of all AEA members (as of November 28, 2016) were obtained. The member database provided by AEA had a total of  $N = 7,289$  unique records. Excluding nine of the study authors, who were registered members of AEA at the time of the investigation, the usable sampling frame was  $N = 7,280$ . Based on the power analysis (see “Design” section), a simple random sample of  $n = 904$  was drawn from the usable sampling frame. The selected random sample of  $n = 904$  was then randomly assigned to one of the four conditions in equal group sizes of  $n = 226$ .<sup>2</sup>

Shown in Table 2 are the traits/characteristics of the AEA membership overall and the traits/characteristics of the selected random sample that was randomly assigned to each of the four conditions. Congruence between the overall AEA membership and the randomly selected sample that was randomly assigned to each of the four conditions was ascertained from univariate  $z$  tests of equality of proportions. As shown in Table 2, and other than a few exceptions, the simple random sample selected and randomly assigned to each of the four conditions reasonably corresponds to the overall AEA member population.

Although four groups of  $n = 226$  were randomly assigned to each condition, the delivery rates (via e-mail) for each of the conditions varied. No replacements were made in the selected sample or conditions for undeliverable e-mails or selected respondents who asked to be removed from AEA’s research list during the administration of the study. The final, usable sample sizes for each condition were control = 218, lottery = 223, token incentive = 219, and philanthropic donation = 225, respectively, and  $n = 885$  over all conditions as shown in Table 3 and as illustrated in Figure 1.

### Instrumentation

The questionnaire distributed to the sample(s) was intentionally brief, simple, and designed only to elicit information pertinent to the study. Demographic traits/characteristics were provided in the AEA database and, therefore, not asked of respondents (see Table 2). All selected participants received the same questionnaire, no matter which condition they were randomly placed in (and, the objective of the investigation was not item responses, per se, but whether an incentive motivated any response whatsoever).

In total, the survey questionnaire consisted of 7 items, of which 4 were closed-response (including one which consisted of a matrix of 18 factors that were to be rated as to the extent to which they positively or negatively influence whether or not one decides to complete a survey questionnaire) and three were open (free)-response items.<sup>3</sup>

The lower-bound estimate of reliability over the 18 ordinal-level items was determined using ordinal  $\alpha$  and  $\Omega$  (see Gugiu, Coryn, & Applegate, 2010), assuming a congeneric (i.e.,

**Table 2.** American Evaluation Association (AEA) Member Population, Control, and Treatment Conditions Selected Sample Traits/Characteristics and Tests of Equality.

Trait/Characteristic <sup>a</sup>	Condition									
	Control		Lottery		Token Incentive		Philanthropic Donation		Test of Equality, $p^b$	Test of Equality, $p^b$
	AEA Member Population (N = 7,280), %	Selected Sample (n = 226), %	Test of Equality, $p^b$	Selected Sample (n = 226), %	Test of Equality, $p^b$	Selected Sample (n = 226), %	Test of Equality, $p^b$	Selected Sample (n = 226), %		
Gender										
Male	24.05	27.88	.14	17.70	.04*	27.43	.24	22.12	.50	
Female	61.58	57.96	.27	66.81	.11	61.06	.87	62.39	.71	
Ethnicity										
African American, Black	7.82	7.96	.94	7.96	.94	7.08	.87	7.96	.94	
American Indian, Native American, Alaska Native	1.00	1.33	.26	1.33	.26	0.88	.86	1.77	.08	
Asian	6.04	7.08	.36	6.19	.93	5.75	.86	4.87	.65	
Caribbean Islander	0.66	0.44	.67	0.44	.67	0.88	.70	0.44	.67	
European American, White	53.46	52.65	.81	53.54	.88	56.64	.28	51.33	.62	
Latino or Hispanic	4.52	6.19	.24	3.54	.69	4.42	.94	5.75	.39	
Middle Eastern or Arab	0.69	0.44	.65	0.00	.21	1.33	.06	0.88	.74	
Native Hawaiian or Pacific Islander	0.30	0.44	.71	0.00	.41	0.00	.41	0.00	.41	
Other	3.48	3.10	.97	2.21	.30	2.65	.50	6.19	.03*	
Highest level of education										
Doctorate	41.48	35.40	.09	37.61	.24	42.04	.77	42.48	.67	
Masters	42.61	49.56	.03*	45.58	.31	41.59	.76	42.04	.97	
Bachelors	5.67	3.98	.28	6.64	.37	6.64	.37	5.31	.96	
Other	0.63	0.44	.72	0.44	.72	0.44	.72	1.77	.00*	
Country										
United States	77.73	77.88	.84	81.42	.14	79.65	.40	73.45	.13	
Other	14.05	13.72	.96	10.62	.20	13.27	.74	17.70	.08	
Primary work setting										
College/University	28.01	27.88	.92	24.34	.28	30.53	.41	29.65	.50	
Nonprofit organization	21.02	18.58	.37	21.68	.81	20.35	.81	21.24	.81	
Private business	20.95	20.80	.92	26.11	.04*	20.80	.92	20.80	.92	
Federal agency	5.25	3.98	.39	6.19	.54	6.19	.54	5.75	.74	
State agency	2.95	2.21	.52	1.33	.30	2.65	.79	4.87	.04*	

(continued)

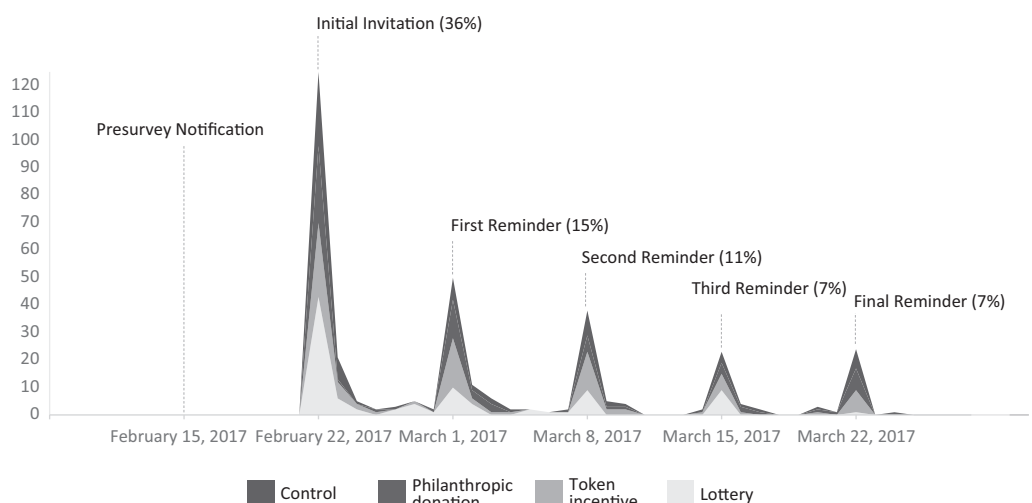




**Table 3.** Delivery Rates of Survey Questionnaire.

Condition <sup>a</sup>	Undeliverable, <sup>b</sup> <i>n</i> (%)	Removed, <sup>c</sup> <i>n</i> (%)	Total Usable Sample, <sup>d</sup> <i>n</i> (%)
Control	3 (1.33)	5 (2.21)	218 (96.46)
Lottery	2 (0.88)	1 (0.44)	223 (98.67)
Token incentive	3 (1.33)	4 (1.77)	219 (96.90)
Philanthropic donation	0 (0.00)	1 (0.44)	225 (99.56)
Total	8 (0.88)	11 (1.22)	885 (97.90)

<sup>a</sup>Percentages do not always total 100% due to rounding error. <sup>b</sup>Selected respondents having undeliverable e-mail addresses at time of survey administration. <sup>c</sup>Selected respondents who asked to be removed from American Evaluation Association's research list during survey administration. <sup>d</sup>Represented as the percentage of *n* = 226 selected for each condition and *n* = 904 for the total sample; the final, total usable samples were used for all analyses reported in the "Results" section.



**Figure 2.** Response patterns for the control and three incentive conditions over survey administration period. The y-axis is stacked and represents the cumulative number of responses over all conditions.

unidimensional) measurement model and accounting for the ordinal-level scaling of the items. For the obtained sample over all conditions, ordinal  $\alpha = .86$  and ordinal  $\Omega = .90$ .

### Procedure

The survey(s) of AEA members was administered from February 22, 2017 (initial invitation) to March 25, 2017 (survey closure) using the Qualtrics survey system. A presurvey notification e-mail message inviting the selected AEA members and informing them of the study and its purposes was sent 1 week prior to administration of the actual survey on February 15, 2017. Weekly reminders (four in all) were sent to those who had not yet completed the survey questionnaire. The pattern of responses for all four groups over the administration period is shown in Figure 2. On the date of the initial invitation, 36.34% of the total responses were received. Each of the four reminders produced response rates ranging from 14.53% (first reminder), 11.05% (second reminder), to approximately 7% (6.69% and 6.98% for the third and final reminders, respectively; see Figure 2).

All communications sent to each of the four groups were identical, except for the following: in communications to those who were assigned to the lottery they were informed that “As a token of

our gratitude for participating in this study, you will be entered in a lottery to have an opportunity to receive a tablet (e.g., Apple iPad, Microsoft Surface) of your choice valued at up to US\$800,” those assigned to the token incentive were informed that “As a token of our gratitude for participating in this study, you will receive a US\$5 Amazon.com gift card,” and those assigned to the philanthropic donation were informed that “As a token of our gratitude for participating in this study, we will donate US\$5 to the AEA Graduate Education Diversity Internship (GEDI) program on your behalf.” The control received no mention of a tangible incentive, other than “This study can only be successful with the generous help of our fellow evaluators,” which was also communicated in the messages sent to the other three conditions. The fact that alternative incentives existed or were offered was intentionally withheld.

Throughout the planning, design, and administration of the surveys, the principles of Dillman, Smyth, and Christian’s (2014) tailored design method were carefully applied in an effort to increase both the quality and quantity of responses. This included application of social exchange theory, which is premised on (1) reducing the costs of responding, (2) increasing the benefits of responding, and (3) establishing trust with (potential) respondents. However, one of the primary features associated with the tailored design method, personalization, could not be applied. The AEA member database used was deidentified, thus preventing the use of potential respondents’ names or titles in communications. Even so, this was not a major factor in this particular investigation (see “Results” section).

*Institutional review.* The study was reviewed and approved by the Western Michigan University Human Subjects Institutional Review Board. Prior to participating in the study, the randomly selected sample of AEA members were asked to read an electronic informed consent.<sup>4</sup>

### *Data Processing and Analysis*

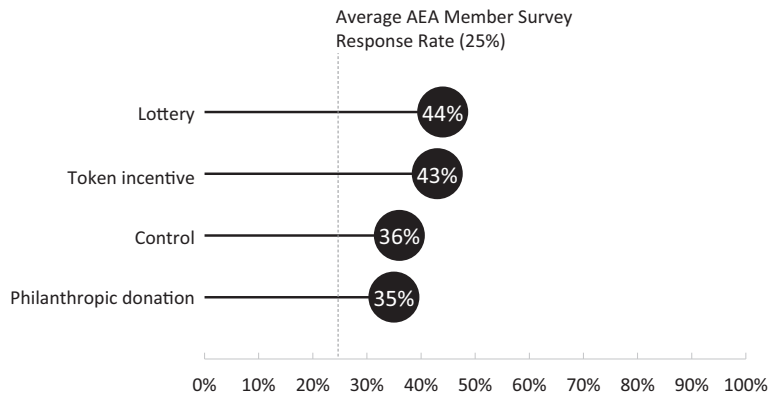
Closed-response data were downloaded from the Qualtrics survey system as comma separated values files and then imported into SAS 9.4 for processing and analysis. Where relevant, bounds on errors of estimation for statistical estimates of population parameters were calculated. For estimating rate of response, a “response” was recorded for participants who provided answers to at least 4 of the 7 items in the questionnaire, treating the matrix of 18 factors as a single item. Over all conditions, item nonresponse for all closed-response items (including the matrix of 18 factors) ranged from as low as 0.29% to a maximum of 4.65%, with no differential rates of item nonresponse between the four conditions. Missing data analysis also indicated that missing responses were missing at random and, therefore, there was no evidence of survey fatigue (i.e., completing early items and not completing later items in the questionnaire). No missing data imputations were performed.

Two of the study authors independently coded the open (free)-response items, following collaborative construction of an emergent coding scheme derived from an initial screening of responses. The independent coding procedure yielded reliability coefficients of  $p_o = .83$  for exact agreement and  $\kappa = .80$  accounting for the probability of chance agreement. Following the independent coding procedure, the two coders worked collaboratively to build consensus on the final coding.

## **Results**

### *Main Effect of Incentives and Moderators of Response Rates*

The overall response rate to the survey(s) was 39.66% ( $n = 351$ ) and, as illustrated in Figure 3, the response rates for each of the four conditions were control = 36.24% ( $n = 79$ ), lottery = 44.39% ( $n = 99$ ), token incentive = 43.28% ( $n = 95$ ), and philanthropic donation = 34.67% ( $n = 78$ ),



**Figure 3.** Response rates for the control and three incentive conditions.

**Table 4.** Response Rates and Tests of Equivalence of Proportions for Obtained Sample.

Condition	Selected Sample	Usable Sample	Obtained Sample	Response Rate, %	<i>p</i>
Control	226	218	79	36.24	
Lottery	226	223	99	44.39	.08
Token incentive	226	219	95	43.38	.13
Philanthropic donation	226	225	78	34.67	.73

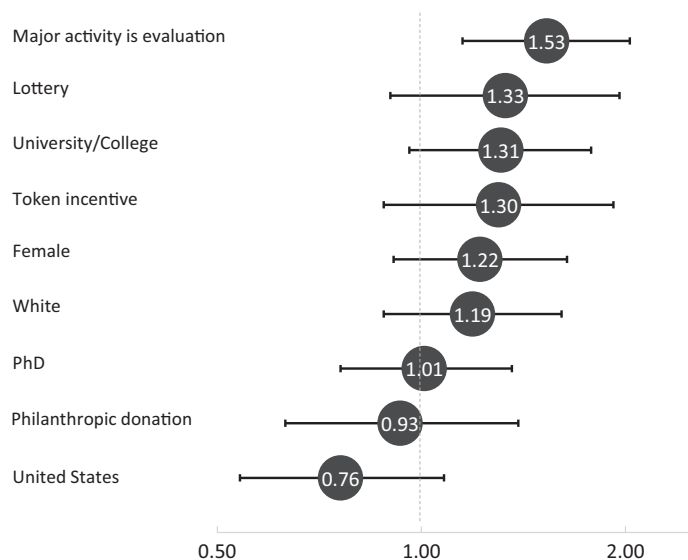
respectively. As shown in Table 4, however, there was not a statistically meaningful difference between the proportions of responses over the four conditions ( $\chi^2 = 3.99$ ,  $df = 3$ ,  $p = .26$ ), nor between each condition and the control.

To further examine the effect of the incentive conditions and potential moderators of response rates (not considered causal), a logistic regression model was fit to the data to estimate the predicted odds of a response. For the model, the criterion was response or nonresponse (where *response* = 1 and *nonresponse* = 0). Seven predictors were used in the model, of which six were taken directly from the AEA membership information provided for the study (see Table 2). The predictors of response used were condition (with the control as the referent; including pairwise comparisons between the control and the other three conditions), gender (with female as the referent), ethnicity (collapsed to European American, White and other with European American, White as the referent), highest level of education (collapsed to doctorate [PhD] and masters/bachelors/other with doctorate as the referent), country (collapsed to United States and other with United States as the referent), primary work setting (collapsed to college/university and other with college/university as the referent), and major activity (collapsed to evaluation and other with evaluation as the referent). In total,  $n = 885$  (see Table 3) unique records with sufficient information were included in the analysis. Overall model fit was moderate, with  $R^2 = .30$  and max-rescaled  $R^2 = .40$  (see Table 5 for fit statistics).

As shown in Table 5, and illustrated in Figure 4, major activity (i.e., evaluation; odds ratio [OR] = 1.53, lower limit [LL] = 1.15, upper limit [UL] = 2.03), the lottery (i.e., for up to US\$800 for a tablet [e.g., Apple iPad, Microsoft Surface]; OR = 1.33, LL = 0.90, UL = 1.96), primary work setting (i.e., college/university; OR = 1.31, LL = 0.96, UL = 1.78), and the token incentive (i.e., US\$5 Amazon gift card; OR = 1.30, LL = 0.88, UL = 1.92) were the factors most associated with an increase in the likelihood of responding. Only major activity (i.e., evaluation), however, was statistically significant ( $p < .01$ ).

**Table 5.** Moderators of Response Rates.

Predictor	$\beta$	SE $\beta$	Wald's $\chi^2$	df	p	OR	95% CI
Intercept	−.88	.21	16.72	1	.42		
Lottery	.28	.20	2.01	1	.16	1.33	[0.90, 1.96]
Token incentive	.26	.20	1.74	1	.19	1.30	[0.88, 1.92]
Philanthropic donation	−.07	.20	0.12	1	.73	0.93	[0.63, 1.39]
Gender	.20	.15	1.74	1	.19	1.22	[0.91, 1.64]
Ethnicity	.17	.15	1.28	1	.26	1.19	[0.88, 1.61]
Highest Level of education	.01	.15	0.01	1	.94	1.01	[0.76, 1.36]
Country	−.27	.18	2.29	1	.13	0.76	[0.54, 1.08]
Primary work setting	.27	.16	2.89	1	.09	1.31	[0.96, 1.78]
Major Activity	.42	.15	8.48	1	<.01	1.53	[1.15, 2.03]
Tests			$\chi^2$	df	p		
Overall model fit							
Likelihood ratio test			24.46	9	<.01		
Score test			24.21	9	<.01		
Wald's test			23.67	9	<.01		



**Figure 4.** Forest plot of odds ratios and 95% confidence intervals of predictors of survey response.

Although major activity (i.e., evaluation) increases the likelihood of responding ( $OR = 1.53$ ,  $LL = 1.15$ ,  $UL = 2.03$ ), as shown in Table 6, there was not a statistically meaningful difference between the proportions of responses over the four conditions ( $\chi^2 = 3.36$ ,  $df = 3$ ,  $p = .34$ ) nor between each condition and the control for the subsample of AEA members who self-identify their major activity as evaluation. The overall response rate for those who identify their major activity as evaluation was 44.68% as compared to 39.66% for the total sample (i.e., +5.02% greater than the total sample; see Tables 4 and 6), in particular for the control (+11.59%), and

**Table 6.** Response Rates and Tests of Equivalence of Proportions for Subsample of American Evaluation Association (AEA) Members Who Self-Identify Their Major Activity as Evaluation.<sup>a</sup>

Condition	Selected Sample	Usable Sample	Obtained Sample	Response Rate, %	<i>p</i>
Control	96	92	44	47.83	
Lottery	107	107	48	44.86	.68
Token incentive	121	118	57	48.31	.95
Philanthropic donation	107	106	40	37.74	.15

<sup>a</sup>Due to space limitations, the subsample of AEA members who self-identify their major activity as evaluation is not disaggregated in other analyses given that the differences between the overall sample and the subsample are, generally, very small.

**Table 7.** Cost-Effectiveness Estimates for the Three Incentive Conditions.

Condition	Response Rate	Total Expended Cost	Per Unit Direct Cost	Per Unit Administrative Cost <sup>a</sup>	Total per Unit Cost	Cost-Effectiveness Ratio <sup>b</sup>
Lottery	44.39% ( <i>n</i> = 99)	\$800.00	\$8.08	\$0.07	\$8.15	\$18.36:1
Token incentive	43.28% ( <i>n</i> = 95)	\$475.00	\$5.00	\$1.09	\$6.09	\$14.07:1
Philanthropic donation	34.67% ( <i>n</i> = 78)	\$390.00	\$5.00	\$0.17	\$5.17	\$14.91:1

Note. Currency in USD.

<sup>a</sup>Administrative costs include only time and effort involved in the delivery of incentives and do not include, for example, costs associated with obtaining university authorization to use or distribute incentives. <sup>b</sup>Cost-effectiveness ratios for each the conditions were estimated by dividing total per unit cost (the sum of per unit direct cost and per unit administrative cost) by response rate (i.e., effect size).

to a lesser extent the lottery (+0.47%), token incentive (+5.03%), and philanthropic donation (+3.07%).

### Cost-Effectiveness of Incentives

The total anticipated costs per condition assuming a 100% response rate were lottery = US\$800.00, token incentive = US\$1,130.00, and philanthropic donation = US\$1,130.00, with a total projected cost of US\$3,060.00 over all conditions (where control = US\$0.00). Given that the response rates for the lottery and token incentive conditions are statistically equivalent and produced the largest effect sizes (i.e., rate of response), then the lottery is the most cost-effective (see Table 7). Considering that the cost per unit of the token incentive and philanthropic donation is dependent upon their effect sizes and that the lottery has a fixed cost, the lottery might, therefore, be considered more cost-effective, given that cost-effectiveness is not dependent upon response rate and that per unit administrative costs are much lower than those associated with the token incentive and philanthropic donation.

### Other Potential Factors Influencing Response Rates

A majority (80.23%  $\pm$  4.11%; see Table 8) of AEA members consider surveys of members either important or very important as a method for examining evaluation practice, and as shown in Table 9, more than two thirds (68.02%  $\pm$  4.82%) have also been previously asked to participate in surveys of AEA members. Nearly one third (29.49%  $\pm$  0.09%) who indicated being asked to complete a survey of AEA members in the past and who indicated that they do not typically respond to surveys of AEA members responded to the survey used in this investigation (see Table 9). Less than one fourth

**Table 8.** Importance of Surveying American Evaluation Association Members to Examine Evaluation Practice.

Condition <sup>a</sup>	Very Unimportant, %	Unimportant, %	Important, %	Very Important, %
Control	6.49	14.29	67.53	10.39
Lottery	3.16	13.68	71.58	10.53
Token incentive	7.45	13.83	63.83	13.83
Philanthropic donation	5.13	11.54	65.38	17.95
Total	5.52	13.37	67.15	13.08

<sup>a</sup>Percentages do not always total 100% due to rounding error.

**Table 9.** Participation in American Evaluation Association (AEA) Member Surveys.

Condition <sup>a</sup>	Ever Been Asked to Participate in a Survey of AEA Members, %			Typically Participate When Asked, <sup>b</sup> %	
	Yes	No	Uncertain	Yes	No
Control	61.04	25.57	10.39	68.09	31.91
Lottery	77.89	11.58	10.53	70.27	28.38
Token incentive	70.21	17.02	11.70	65.15	34.85
Philanthropic donation	60.26	21.79	17.95	78.72	21.28
Total	68.02	19.19	12.50	70.09	29.49

<sup>a</sup>Percentages do not always total 100% due to rounding error. <sup>b</sup>Of those who indicated “yes” to having ever been asked to participate in a survey of AEA members.

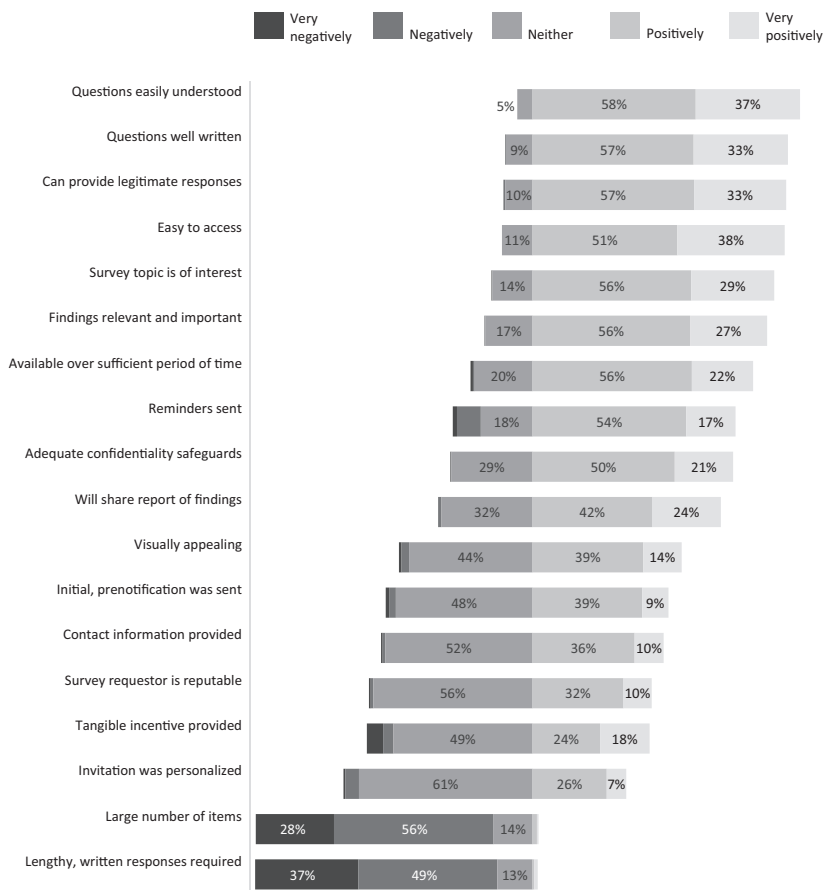
**Table 10.** Frequency of Skimming or Reviewing a Survey Prior to Deciding to Participate.

Condition <sup>a</sup>	Never, %	Very Rarely, %	Rarely, %	Occasionally, %	Very Frequently, %	Always, %
Control	24.68	10.39	15.58	25.97	16.88	5.19
Lottery	21.51	11.83	16.13	25.81	17.20	7.53
Token incentive	20.21	23.40	11.70	26.60	14.89	2.13
Philanthropic donation	21.79	19.23	11.54	28.21	12.82	6.41
Total	21.93	16.37	13.74	26.61	15.50	5.26

<sup>a</sup>Percentages do not always total 100% due to rounding error.

(20.76%  $\pm$  4.20%; see Table 10) of AEA members (very frequently or always, combined) skim or review a survey prior to deciding whether or not to participate.

As shown in Figure 5, questions easily understood (94.56%  $\pm$  2.39%), questions well written (90.30%  $\pm$  3.12%), survey easy to access (89.16%  $\pm$  3.27%), can provide legitimate responses (89.67%  $\pm$  3.22%), survey topic is of interest (85.46%  $\pm$  3.68%), and findings relevant and important (82.93%  $\pm$  3.85%) are among the potential factors that most positively influence (positively and very positively, combined) whether AEA members decide whether or not to participate in a survey of members. The factors that most negatively influence (very negatively and negatively, combined) whether AEA members participate in surveys of members are lengthy, written responses required (85.67%  $\pm$  3.71%) and large number of items (84.08%  $\pm$  3.84%). Whether a tangible incentive is provided (the primary impetus for this investigation) was neither negatively nor positively endorsed by nearly half of AEA members (49.24%  $\pm$  5.27%).



**Figure 5.** Factors that influence whether American Evaluation Association members decide to respond to member surveys.

Thematic analysis of the open (free)-response items—intended to elicit information about respondents’ motivations for responding to the survey, what types of incentives, if any, are most likely to encourage them to respond to surveys of AEA members, and any other factors that influenced their decision to respond or not—revealed several additional factors related to responding not specifically enumerated in the closed-response items. Approximately one fourth of AEA members noted specific characteristics of the survey and the survey procedure, including multiple e-mail reminders, the inclusion of a material incentive, and short completion time as influences for responding. Reminder e-mails were the most frequently mentioned reason for responding. For some, reminders came at an appropriate or opportune time. As one respondent remarked, “I got a reminder at a time when I had a few minutes free.” Others, however, felt the reminders were “nagging” and, therefore, responded in order “. . . to stop the reminders!” Many respondents paired their comments of reminders with feelings of guilt or professional responsibility. As one respondent wrote, “. . . the reminder e-mail made me feel guilty for not responding as an evaluator” and another stated “. . . [I responded because of] . . . repeated follow-ups. I had good intentions, but earlier e-mails came at a time when I was busy with other work. I do survey data collection too, so I always feel that I should complete surveys given that it is something I ask others to do.”

A desire to contribute to the profession and practice of evaluation was the second most cited reason for responding, closely followed by characteristics of the respondent. Many respondents noted their desire to contribute to expanding the evaluation knowledge base and "...help to improve evaluation practice." Others noted their desire to accumulate "survey karma" (e.g., "...[I am] hoping that if I have a survey request one day, others will be responsive to me."). Various respondents mentioned their personal interest in the study topic as a reason for responding, some noting the "meta" nature of taking a survey on habits of taking surveys, "As an evaluator and creator of surveys myself, I appreciate the benefit of having a good response rate...it is quite a meta-topic—doing a survey about surveys, so it would be very difficult to pass this opportunity over."

Many also indicated their affiliation with AEA as a primary motivation for completing the survey, noting their trust and respect for the organization. As one respondent remarked, "If the survey comes from my professional association, I feel bound to respond because I know the association must feel the survey is important to our field."

Although one third of respondents indicated financial motives for responding to the survey request, twice as many cited intrinsic motivations. Respondents frequently mentioned the relevancy of the topic to their work or personal interests as incentives to respond. Many simply stated that topics of interest motivate them, while others noted the "relevance of the topic to [their] work" and "subjective importance of [the] topic." One respondent stated, the "... greatest incentive is believing that the survey has a good purpose and the information is important to collect" whereas another wrote, "... a topic that I find interesting and that I think I can give good feedback on is the most important incentive for me." Relevance or ability to adequately respond was a recurring theme, and some respondents found motivation "... when it [the survey] is a topic that [they] have an opinion about or experience with." These types of responses were frequently coupled with observations on the importance of a high-quality survey requiring only a short-time commitment. One respondent simply stated, "... the incentive does not make any difference. The design and quality of the questions do," while another emphasized the importance of "... thoughtful questions" and stated that they "... often quit in the middle of surveys because they're so poorly conceptualized."

A major, often overlooked, motivating force mentioned by many respondents was receiving a summary of results or written report of findings. While not originally promised in communications with selected participants, a two-sided, single-page summary of the primary findings of the investigation was disseminated to all respondents and nonrespondents given the prominence of this motivating factor.

## Discussion

Over all conditions, the response rate to the survey(s) was greater (39.66%) than the average rate of response reported in recent surveys of AEA members (25.18%); in particular for the lottery (44.39%) and the token incentive (43.28%). Considering only the subsample who self-identified their major activity as evaluation, the overall response rate increases to 44.68% (and to 48.31% for those who were offered the lottery) and they are 1.53 times more likely to respond to surveys of AEA members than members who do not identify their primary activity as evaluation.

Not offering or providing a material incentive is clearly cost-effective, yet yields only a modest return in terms of responses relative to the two material incentives offered (though essentially equal to the philanthropic donation response rate). Compared to the token incentive and philanthropic donation, the lottery yielded the largest response rate relative to total costs. Even in the absence of a material incentive (i.e., the control), the response rate for the subsample of AEA members who self-identify their major activity as evaluation was nearly twice (47.83%) the average rate of response reported in recent surveys of AEA members (25.18%).



Although the primary objective of this investigation was to determine what type of incentive, if any, increases response rates to Internet surveys of AEA members, slightly less than half ( $49.24\% \pm 5.27\%$ ) of respondents indicated that a material incentive is a primary motive for whether or not they decide to respond. Other factors (e.g., questions easily understood, questions well written, survey easy to access, can provide legitimate responses, survey topic is of interest, findings relevant and important) are often equally or more important to a large majority of AEA members. In addition, the majority of surveys of AEA members are administered for between 2 and 4 weeks (see Table 1). As illustrated in Figure 2, however, meaningful numbers of responses continue to occur beyond 4 weeks, suggesting that longer periods of administration might be a useful, simple means for increasing rates of response.

### *Implications*

Internet surveys are likely to remain a common method for eliciting information from AEA members (and for use in evaluation practice, more generally) for the foreseeable future, whether or not material incentives are used. Members appear to be encouraged and more likely to provide responses when a large lottery incentive is offered, in particular when coupled with a social exchange approach to soliciting participation.

If the target population of an investigation is a particular subpopulation (e.g., those who identify their major activity as evaluation), then the relevant subpopulation should be intentionally extracted and isolated from the larger population of all members so that ineligible units (i.e., those who do not identify their primary activity as evaluation) are not included in the sampling frame. Doing so will likely result in increases in the representativeness of obtained samples as well as rates of response.

Although not applied in this investigation, mixed-mode surveys (e.g., Internet and mail), although more expensive, often produce higher response rates than Internet surveys alone (Converse, Wolfe, Huang, & Oswald, 2008; Greenlaw & Brown-Welty, 2009). For list-based samples whose postal addresses are known, such as AEA's membership, mixed modes of administration should be considered.

Although more technically demanding, nonresponse bias analyses, such as poststratification weighting or other methods of nonresponse adjustment, should potentially be performed as a means of bias reduction; especially when response rates are low, as nonresponse bias tends to increase as response rates decrease. Particularly, such methods should be considered if nonrespondents systematically differ from respondents and are a sufficiently large proportion of the population that such differences effect sample estimates. Even so, such techniques require either sample-based or non-sample-based variables useful for statistical adjustments to nonresponse.<sup>5</sup>

Recent surveys of AEA members as well as those sometimes distributed through EvalTalk frequently consist of relatively large numbers of complex items and often require lengthy written responses, both of which discourage responses among AEA membership. Simultaneously, the suggested completion times described in solicitations for these surveys are sometimes erroneous or misleading. Many of these surveys are seemingly conducted for completing masters' theses and doctoral dissertations. If so, these students should be cognizant of how lengthy a survey questionnaire truly is; in particular, if one is not only concerned with the response rate but also response quality. These students should also consider using a material incentive, such as a lottery, in an effort to increase responses. An US\$800 lottery is a modest investment for a masters' thesis or doctoral dissertation.

### *Limitations*

During administration of the survey(s), it became anecdotally known that not all participants read the statement regarding the assigned incentive and, therefore, responded to the survey for reasons other than the randomly assigned incentive or no-incentive condition. Although unlikely in this instance, such failure to receive treatment could be problematic as it could bias inferences about

the effectiveness of a treatment and could be considered as an implementation failure (Shadish, Cook, & Campbell, 2002). Additionally, not all AEA members likely use Amazon.com, and thus, the US\$5 Amazon.com gift card was not influential to some of those who were assigned to the token incentive condition.

The survey used in this investigation was intentionally brief and survey length is an important consideration in whether or not AEA members decide to respond. Other surveys of AEA members are frequently far longer in terms of the overall number of items as well as anticipated completion time. Therefore, possibly reducing the generalizability of findings reported.

### **Future Research**

Additional research is needed to improve survey methodology within the evaluation community. With respect to survey administration, multiple e-mail messages can be sent to potential respondents for little or no cost, and the impact of multiple contacts on response decisions is a relevant subject of investigation in the administration of Internet surveys (Sheehan & McMillan, 1999).

It is also important to more carefully investigate respondents' and nonrespondents' perceptions of the costs and benefits of survey participation. The goal of such research is to develop empirically based efforts to improve the survey experience. Given the experimental design, the benefits of survey participation other than a material incentive can be assumed constant across groups, hence more attention to the perceived benefit of such incentives is needed.

As demonstrated in the cost-analysis, the lottery incentive is the only condition for which a respondent has a probabilistic expected value and therefore offers an opportunity for research on response behavior in greater depth. Furthermore, because the effect size is dependent upon participants' perceived levels of expected value, research into better understanding this phenomenon is warranted. The expected value to a participant of the lottery is imprecise due to the nature of response rates impacting the reward pool. In this instance, the expected value to a participant ranged from US\$3.59, in the instance that all 223 responses in the usable sample were obtained, to US\$800 in the instance of only a single response. With a response rate of 44.39%, the resulting expected value of the lottery condition was US\$8.08, but market imperfections dictating extrinsic value and preferences (i.e., intrinsic value) for rewards should also be factored into this expected value. Additionally, given the lower average response rate of recent surveys of a similar population, the perceived expected value to a respondent at the time of response may have been higher. Conceptually, an optimal value of the lottery exists as compared to the next best alternative, the token incentive, and further research is needed to describe and ultimately model this value.

Given the increasingly internationalization of professional evaluation, the practices of members of professional associations and societies other than, or in addition to, members of AEA should be systematic investigated. These associations and societies might include, but certainly are not limited to, the African Evaluation Association, Australasian Evaluation Society, Canadian Evaluation Society, and European Evaluation Society, among many others. Otherwise, what is known about evaluation practice predominately reflects only Western practices.


### **Declaration of Conflicting Interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### **Funding**

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Interdisciplinary PhD in Evaluation (IDPE) program at Western Michigan University (WMU).

## ORCID iD

Pedro F. Mateu  <https://orcid.org/0000-0001-8759-9244>

## Notes

1. These response rates are somewhat lower than, yet comparable to those often obtained from surveys of members of other types of professional organizations (Rogelberg, 2006).
2. A simple random sample (estimated using a bound on the error of estimation of  $\pm 5\%$  and a conservative population proportion of  $p = .25$ , i.e., assuming a 25% response rate; the average for surveys of American Evaluation Association [AEA] members) of the  $N = 7,280$  usable records produced a necessary sample size of  $n = 288$  to address the focal research question. Given that the required sample size resulting from the power analysis was greater than that required for a simple random sample not considering statistical power, the larger of the two sampling options was used.
3. The complete survey questionnaire is available from the first author upon request.
4. In addition to requesting informed consent, participants also received the following information in the initial and subsequent e-mail notifications: "You are receiving this e-mail as a member of the American Evaluation Association (AEA). This research request was reviewed by a Research Request Task Force consisting of tenured AEA members. If you have concerns about the survey and would like to express them to the AEA leadership, please e-mail [info@eval.org](mailto:info@eval.org). Any concerns raised will be shared, confidentially, with the Executive Committee of the association. AEA allows its membership list to be used infrequently for research that focuses on the field of evaluation. If you would like to opt out of AEA's research list, please send an e-mail request to [info@eval.org](mailto:info@eval.org). Please note that we encourage you to consider remaining on the list as such research strengthens and furthers the field's knowledge base."
5. Nonresponse adjustments were not applied in the analyses reported here given the statistical match between the obtained sample and target population (i.e., poststratification weighting was statistically unwarranted).

## References

- Armstrong, J. S. (1975). Monetary incentives in mail surveys. *Public Opinion Quarterly*, 39, 111–116.
- Azzam, T., & Szanyi, M. (2011). Designing evaluations: A study examining preferred evaluation designs. *Studies in Educational Evaluation*, 37, 134–143.
- Bosnjak, M., & Tuten, T. L. (2003). Prepaid and promised incentives in Web surveys: An experiment. *Social Science Computer Review*, 21, 208–217.
- Boulianne, S. (2008). Incentives. In P. J. Lavrakas (Ed.), *Encyclopedia of survey research methods* (Vol. 1, pp. 326–331). Thousand Oaks, CA: Sage.
- Church, A. H. (1993). Estimating the effect of incentives on mail survey rates: A meta-analysis. *Public Opinion Quarterly*, 57, 62–79.
- Converse, P. D., Wolfe, E., Huang, X., & Oswald, F. (2008). Response rates for mixed-mode surveys using mail and e-mail/web. *American Journal of Evaluation*, 29, 99–107.
- Cook, C., Heath, F., & Thompson, R. L. (2000). A meta-analysis of response rates in web- or internet-based surveys. *Educational and Psychological Measurement*, 60, 821–836.
- Coryn, C. L. S., Ozeki, S., Wilson, L. N., Greenman II, G. D., Schröter, D. C., Hobson, K. A., . . . Vo, A. T. (2016). Does research on evaluation matter? Findings from a survey of American evaluation association members and prominent evaluation theorists and scholars. *American Journal of Evaluation*, 37, 159–173.
- Coryn, C. L. S., Wilson, L. N., Westine, C. D., Hobson, K. A., Ozeki, S., Fiekowsky, E. L., . . . Schröter, D. C. (2017). A decade of research on evaluation: A systematic review of research on evaluation published between 2005 and 2014. *American Journal of Evaluation*, 38, 329–347.
- Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed-mode surveys: The tailored design method* (4th ed.). Hoboken, NJ: Wiley.

- Deutsdens, E., de Ruyter, K., Wetzels, M., & Oosterveld, P. (2004). Response rate and response quality of Internet-based surveys: An experimental study. *Marketing Letters*, 15, 21–36.
- Fan, W., & Yan, Z. (2010). Factors affecting response rates of the web survey: A systematic review. *Computers in Human Behavior*, 26, 132–139.
- Fleischer, D. N., & Christie, C. A. (2009). Evaluation use: Results from a survey of U.S. American Evaluation Association members. *American Journal of Evaluation*, 30, 158–175.
- Fox, R. J., Crask, M. R., & Kim, J. (1988). Mail survey response rate—A meta-analysis of selected techniques for inducing response. *Public Opinion Quarterly*, 52, 467–491.
- Galport, N., & Azzam, T. (2017). Evaluator training needs and competencies: A gap analysis. *American Journal of Evaluation*, 38, 80–100.
- Galport, M., & Galport, N. (2015). Methodological trends in research on evaluation. *New Directions for Evaluation*, 148, 17–29.
- Göriz, A. S. (2006). Incentives in Web studies: Methodological issues and a review. *International Journal of Internet Science*, 1, 58–70.
- Greenlaw, C., & Brown-Welty, S. (2009). A comparison of web-based and paper-based survey methods: Testing assumptions of survey mode and response cost. *Evaluation Review*, 33, 464–480.
- Gugiu, P. C., Coryn, C. L. S., & Applegate, E. B. (2010). Structure and measurement properties of the patient assessment of chronic illness care instrument. *Journal of Evaluation in Clinical Practice*, 16, 509–516.
- Henry, G. T., & Mark, M. M. (2003). Toward an agenda for research on evaluation. *New Directions for Evaluation*, 97.
- Levy, P. S., & Lemeshow, S. (2013). *Sampling of populations: Methods and applications* (4th ed.). Hoboken, NJ: Wiley.
- Lewis, N. R., Harrison, G. M., Ah Sam, A. F., & Brandon, P. R. (2015). Evaluators' perspectives on research on evaluation. *New Directions for Evaluation*, 148, 89–102.
- Mark, M. M. (2007). Building a better evidence base for evaluation theory: Beyond general calls to a framework of types of research on evaluation. In N. L. Smith & P. R. Brandon (Eds.), *Fundamental issues in evaluation* (pp. 111–134). New York, NY: Guilford.
- Ozeki, S. (2016). *Three empirical investigations into the logic of evaluation and valuing practices* (Unpublished doctoral dissertation). Western Michigan University, Kalamazoo.
- Preskill, H., & Caracelli, V. (1997). Current and developing conceptions of use: Evaluation Use TIG survey results. *American Journal of Evaluation*, 18, 209–225.
- Rogelberg, S. G. (2006). Understanding nonresponse and facilitating response to organizational surveys. In A. I. Kraut (Ed.), *Getting action from organizational surveys: New concepts, methods, and applications* (pp. 312–325). San Francisco, CA: Jossey-Bass.
- Seidling, M. B. (2015). Evaluator certification and credentialing revisited: A survey of American Evaluation Association members in the United States. *New Directions for Evaluation*, 145, 87–102.
- Seifert, T. (2008). Leverage-salience theory. In P. J. Lavrakas (Ed.), *Encyclopedia of survey research methods* (Vol. 1, pp. 423–425). Thousand Oaks, CA: Sage.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.
- Sheehan, K., & McMillan, S. (1999). Response variation in e-mail surveys: An exploration. *Journal of Advertising Research*, 39, 45–54.
- Singer, E., Van Hoewyk, J., Gebler, N., Raghunathan, T., & McGonagle, K. (1999). The effect of incentives on response rates in interviewer-mediated surveys. *Journal of Official Statistics*, 15, 217–230.
- Szanyi, M., Azzam, T., & Galen, M. (2013). Research on evaluation: A needs assessment. *Canadian Journal of Program Evaluation*, 27, 39–64.
- Yu, J., & Cooper, H. (1983). A quantitative review of research design effects on response rates to questionnaires. *Journal of Marketing Research*, 20, 36–44.