



Dissertations

Graduate College

4-2016

Crowd Sourced Product Reviews: A Study of Evaluation Standards Used

Alexander W. Manga
Western Michigan University, mangaa@ferris.edu

Follow this and additional works at: <https://scholarworks.wmich.edu/dissertations>



Part of the E-Commerce Commons, Educational Assessment, Evaluation, and Research Commons, and the Sales and Merchandising Commons

Recommended Citation

Manga, Alexander W., "Crowd Sourced Product Reviews: A Study of Evaluation Standards Used" (2016).
Dissertations. 1412.

<https://scholarworks.wmich.edu/dissertations/1412>

This Dissertation-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Dissertations by an authorized administrator of ScholarWorks at WMU. For more information, please contact wmu-scholarworks@wmich.edu.



CROWD SOURCED PRODUCT REVIEWS: A STUDY OF EVALUATION
STANDARDS USED

by

Alexander W. Manga

A dissertation submitted to the Graduate College
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
Interdisciplinary Ph.D. in Evaluation
Western Michigan University
April 2016

Doctoral Committee:

Daniela Schroeter, Ph.D., Chair
Chris Coryn, Ph.D.
Kuanchin Chen, Ph.D.

CROWD SOURCED PRODUCT REVIEWS: A STUDY OF EVALUATION STANDARDS USED

Alexander W. Manga, Ph.D.

Western Michigan University, 2016

People are using online product reviews and evaluations more and more (Chandler et al., 2013). As the usage of online reviews persists and more smart phone applications are created, the demand for online product review continues to increase; yet, there is no indication of the quality of these reviews. In fact, some online reviews have been found to be fraudulent and misleading. Many online product reviews come from Internet-based crowdsource organizations. Few studies have explored evaluation practices among these organizations, and as a result, it is unclear what, if any, evaluation standards are used by crowdsource reviewers, particularly those found on open, self-serve sites such as MTurk. The purpose of this study, therefore, was to determine (a) what, if any, evaluation standards are used by crowdsource organizations and their requesters, and (b) to what extent these standards adhere to the Joint Committee on Standards for Education Evaluation (JCSEE) Program Evaluation Standards (Yarbrough et al., 2011).

Descriptive, survey data was collected from 454 MTurk product reviewers. Findings indicate these product reviewers do not appear to use any standards. The MTurk product reviewers that participated in this survey are using personal, experience-based opinions as a basis for their online reviews. The literature tells us, however, these

opinions are not reliable, as they change with the providers experience and knowledge of the product.

Results further indicate participants appear to not be very procedural. Document management seemed to be reviewer dependent. Moreover, open-ended follow-up questions reveal that when asked if they used more technical review designs, the majority of participants answered “often,” while simultaneously indicating their reviews were based on personal experience. This result was very conflictive with survey results, and further points to a misperception that MTurk product reviewers are providing reliable online product reviews.

© 2016 Alexander W. Manga

ACKNOWLEDGMENTS

I want to thank all those that helped me in this endeavor, the many friends that were there for me when I needed it.

To my sons, Michael and Christian, you both sacrificed the most for this achievement. I was gone two to three nights a week for the better part of three years. I sincerely love you both with all my heart and soul, and hope one day you'll understand it was worth it. Thank you for not giving up on me.

Thank you to Pedro Mateu, Kristin Hobson, Chris Coryn, Daniela Schroeter, K.C. Chen, and Gabi Vinci Manga. Thank you all for being the team that allowed this win to happen.

To Dr. Tess Gordon, thank you for pushing me to the finish! You are an incredible human being of which I am very fond of and admire. Thank you for all of your support and wisdom.

Alexander W. Manga

TABLE OF CONTENTS

ACKNOWLEDGMENTS	ii
LIST OF TABLES	viii
LIST OF FIGURES	x
CHAPTER	
I. INTRODUCTION	1
Background of the Study	2
Crowdsourcing	3
MTurk, How it is Researched, and How it is Used for Research	5
Product Evaluation	9
Statement of the Problem	14
Purpose of the Study	15
Research Questions	15
Definition of Terms	16
II. REVIEW OF LITERATURE	18
The History of Product Evaluation	18
Good Housekeeping	19
Consumers Research	20
The Federal Trade Commission	22
Underwriters Laboratory	22
The Current Situation of Product Evaluation	27

Table of Contents—Continued

CHAPTER

Online Product Reviews	30
Market Sites.....	31
Producers or Distributors	31
Review Sites.....	32
Social Media.....	33
Crowdsourcing	33
MTURK: The First Open to the Public Crowdsourcing Model Called Open Sourcing	36
Advances and Continued Areas for Growth in Crowdsourcing.....	39
Problems Associated With Online Reviews and Crowdsourcing	40
Post-purchase Consumer Product Evaluation Approaches	42
Evaluation by Alternative Product Comparison.....	42
Goal-based Performance Evaluations	44
Key Evaluation Checklist.....	45
Goal Free Approach	45
Connoisseur Product Evaluation	46
Scriven’s Critical Review of Consumer Reports	49
The Joint Committee on Standards for Education Evaluation	52
Other Standards Organizations.....	55
ISO 9001 Standards.....	55
ANSI Standards.....	56

Table of Contents—Continued

CHAPTER		
	ASTM.....	57
	The United States Government Accountability Office Standards	58
	AEA Guiding Principles.....	58
	Summary	59
III.	METHOD.....	61
	Purpose of the Study.....	61
	Research Questions	62
	Research Design	62
	Non-probability Sample	65
	Instrumentation.....	67
	Data Collection Procedures	68
	Setting up the Crowdsourcing Project	68
	Recruitment of Reviewers.....	69
	Informed Consent Process.....	69
	Data Analysis	70
	Secondary Analysis: Thematic Synthesis.....	70
IV.	RESULTS.....	72
	Participant Demographics	72
	Participant Crowdsourcing Evaluation Experience	73
	Participant Range of Pay Per Review	75

Table of Contents—Continued

CHAPTER	
	Crowdsource Workers’ Use of the Utility, Feasibility, Propriety, and Accuracy Domains of the JCSEE Program Evaluation Standards 76
	Utility 77
	Feasibility 82
	Cross-tabulations 89
V.	CONCLUSION 99
	Context of the Study 99
	Review of Methodology 104
	Summary of Empirical Findings 106
	Implications 109
	Contributions 112
	Limitations of the Study 113
	Research Opportunities 114
	REFERENCES 115
	APPENDICES 124
	A. Human Subjects Institutional Review Board Approval 124
	B. HSIRB Application 126
	C. Survey of Evaluation Standards Questionnaire 139
	D. Contract with MTurk 144
	E. Reviewer Agreement 146
	F. Amazon Mechanical Turk Participation Agreement 148

Table of Contents—Continued

APPENDICES

G. Request for Participation	159
H. Consent Letter	161
I. Thematic Analysis	163
J. Thematic Mapping.....	182

LIST OF TABLES

1.	JCSEE Program Evaluation Standards	12
2.	Historical Review of Good Housekeeping and Consumers Union Product Evaluation Activities	25
3.	Summary of Research Methodology	71
4.	Participant Crowdsourcing Evaluation Experience	74
5.	Participant Range of Pay Per Review	75
6.	Frequencies and Percentages of Crowdsourcing Workers' Use of the Utility, Domain of the JCSEE Program Evaluation Standards	76
7.	Frequencies and Percentages of Crowdsourcing Workers' Use of the Feasibility Domain of the JCSEE Program Evaluation Standards	83
8.	Frequencies and Percentages of Crowdsourcing Workers' Use of the Propriety Domain of the JCSEE Program Evaluation Standards	85
9.	Frequencies and Percentages of Crowdsourcing Workers' Use of Accuracy Domain of the JCSEE Program Evaluation Standards	87
10.	Cross-tabulation of Were Your Reviews Solely Based on Past Experience and My Reviews Were Reliable	90
11.	Cross-tabulation of My Reviews Were Reliable and Do You Have Formal Training for Conducting Product Reviews	92
12.	Cross-tabulation of Were Your Reviews Solely Based on Past Experience and I Used Practical Procedures	93
13.	Cross-tabulation of Do You Have Formal Training for Conducting Product Reviews and What is Your Role in Crowdsourcing	95

List of Tables—Continued

14.	Cross-tabulation of Were Your Reviews Solely Based on Past Experience and I Used More Technical Reviews Designs, When Appropriate, for the Specific Review	97
15.	Summary of Adherence to JCSEE Standards	107

LIST OF FIGURES

1. Basic branches of product evaluation.	30
2. Scriven’s fundamental product evaluation approaches.....	47
3. Percentages of crowdsource workers’ use of the Utility domain of the JCSEE program evaluation standards.	78
4. Percentages of crowdsource workers’ use of the Feasibility domain of the JCSEE Program Evaluation Standards.	83
5. Percentages of crowdsource workers’ use of the Propriety domain of the JCSEE program evaluation standards.	85
6. Percentages of crowdsource workers’ use of the Accuracy domain of the JCSEE program evaluation standards.	88

CHAPTER 1

INTRODUCTION

At the start of the new millennium, businesses were beginning to see the value of attracting customers to their websites. Either for the transfer of general information or e-commerce, virtual business was becoming real and big. The dot-com bubble burst just a few years prior, leaving techies scrambling to acquire new technologies and strategies to take advantage of a growing stream of online activity. Concurrently, businesses were also learning that Internet users with similar interests were likely to congregate, stick to, and become members of websites of preference. In essence, businesses were learning they could acquire a captive market online with basic and simplistic value propositions based on need, speed, and convenience.

Crowdsourcing is one specific business strategy that emerged out of the growing online market of the early 2000s. “Generally, crowdsourcing is when a company uses the Internet to outsource activities previously performed by employees within the organization” (Howe, 2006). MTurk is a successful open source approach to crowdsource business that uses workers to perform a variety of tasks ranging from completing surveys, participating in experiments, looking at pictures, and reviewing data or scripts (Buhrmester, Kwang, & Gosling, 2011). One task MTurk frequently performs in particular is online product evaluation (product reviews). According to Le, Edmonds, Hester, and Biewald (2010), product evaluation using crowdsourcing may be superior to traditional methods of evaluation because it facilitates feedback from a variety of viewpoints. Yet, little is known about crowdsourced evaluations, or the standards by which they are conducted. The purpose of this study was to learn about crowdsource

evaluation standards. More specifically, this research sought to determine what standards crowdsource evaluation workers are using and if these standards adhere to the Joint Committee on Standards for Educational Evaluation (JCSEE) program evaluation standards (Yarbrough, Shulha, Hopson, & Caruthers, 2011).

Background of the Study

e.Lilly appears to have kicked off the original concept of crowdsourcing. In 2001, Alpheus Bingham, Vice President of e.R&D at e.Lilly, announced a new e-business venture called InnoCentive LLC. The purpose of InnoCentive was to use the Internet to create and enhance open-source scientific research and development. As explained by Darren Carroll, Chief Executive Officer at InnoCentive, “We are at the heart and soul of what the Internet is all about. InnoCentive represents a return to the Internet’s roots...an open-source approach to scientific collaboration and innovation. We are seeking access to that particular mind that is uniquely prepared for solving a specific scientific problem” (<https://investor.lilly.com/releasedetail2.cfm?releaseid=52031>). In this way, the venture was pioneering and changed the industry. This crowd effect ramped up during the early to mid-2000s; however, it was InnoCentive’s approach that captured the attention of many, as it would use its audience to solve complex problems. This was new and quite compelling to the online community.

During this time, share sites became another way to capture a unique and rare competency using the Internet. iStockphoto was a picture exchange site that allowed its members to submit and license photos for use at little to no cost. This attracted a diverse population of buyers of photostock. According to Howe (2006), writers and members of the media needing photos no longer had to pay more than \$150.00 a photo from a

freelancer; they could get high quality photos from iStockphoto for \$1.00 each. The scarcity effect on pricing in the picture industry was now gone.

Howe (2006) noted that companies such as iStockphoto were once shunned by the stock industry because of their capacity to diminish photographers' wealth. In February 2006, Getty Images, the largest photo agency of its time, purchased iStockphoto for \$50 million. This acquisition may have been a win-win for the stock industry in general. iStockphoto quickly became a go-to source for many larger corporate buyers such as IBM and the United Way, and also offered an easy exchange and pay system that gave individuals more money for their photos than previously experienced.

Crowdsourcing

Companies such as InnoCentive and iStockphoto quickly became leaders among e-business ventures. What made InnoCentive and iStockphoto different from other companies was their goal of harnessing labor. This was unique and compelling. It also seemed like a more durable and sustainable source of growth. By 2005, the model of harnessing your crowd or market space became a business strategy and key value proposition that other companies performing consultation on e-commerce began to utilize. In fact, much of the literature about e-commerce at that time was focused on coring and capturing market space.

The new Internet model initiated by companies like InnoCentive and iStockphoto resembled outsourcing, which had also become a growing hallmark in American manufacturing. Outsourcing is the act of a company going outside of its own company to acquire products or labor to satisfy production or customer needs. In June 2006, Jeffrey Howe coined the term *crowdsourcing* in Wired Magazine. According to Howe (2006),

crowdsourcing can be defined as “the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined, network of people in the form of an open call.” Overall, the crowd effect has been very positive. It has driven the cost of labor down significantly. The volume of workers on crowdsourcing sites cut the time needed to procure and complete work to a fraction. In an era where the economy was still reverberating from the events on September 11 and barely any GNP or measureable economic growth, this was very appealing.

According to Howe (2006), the advantages of crowdsourcing include reductions in labor costs and the time it takes to complete tasks. Prior to the availability of crowdsourcing, requesting companies had to fully employ laborers. This was expensive and contrary to the new millennial business model logic, which sought to decrease labor expense and increase productivity in a climate of world competition and pricing. Moreover, labor costs often varied based on availability and level of talent needed. The new model of crowdsourcing broke barriers of scarcity by allowing workers to come and go as they pleased, and to take on smaller parts of work at a time vs. an entire project contract. The new model also helped to attenuate scarcity of labor by using an open source model that attracted very large volumes of people at one time.

How crowdsourcing works is not nearly as complex as it may seem. A company identifies a list of tasks or work currently being accomplished in-house that it now wants to outsource. It then uses its own site or portal to organize workers and post and deploy available work or tasks. Workers then undertake the work or tasks with the expectation of being paid based on identified pay rates. Submitted work is paid for if it is acceptable. If the submission is not acceptable, the worker may or may not have the opportunity to redo

work for further consideration. In some cases, more than one worker may work on the same work or task, and the company will choose only one best submission (Morris & Picard, 2012).

Jeff Bezzo, Founder and CEO of Amazon.com, was so impressed with the advantages of using crowd labor, he directed Peter Cohen, Director of Corporate Development Amazon Web Services, to create a model within Amazon to solve its own problems with e-commerce webpages that could not be resolved systematically using computers (<https://www.mturk.com/mturk/help?helpPage=overview>). In November 2005, Cohen announced the release of a beta version of MTurk, a worker exchange site that would allow basically trained individuals to perform human intelligence tasks (HIT) in order to pan pages of code, products, product descriptions, and specifications to identify and cull redundancy. Amazon's MTurk has a three-pronged strategic approach designed to bring together community, technology, and compensation (www.amazon.com).

MTurk, How it is Researched, and How it is Used for Research

The name MTurk is as compelling and interesting as the service itself. Named after a 17th century hoax, MTurk represents the use of human intelligence through a mechanical device. The Turk, created by a Hungarian nobleman Wolfgang Von Kempelen, was essentially a wooden box with a human figure on top that would challenge passersby to a game of chess as it toured Europe. Well known mentalists of the time, Benjamin Franklin and Napoleon, were both beaten by the Turk. Yet, it was not the mechanical human figure atop the box that was winning chess on the streets in Europe in the 1760s—it was the chess master hidden in the box below (Howe, 2006).

Created to solve internal webpage problems at Amazon, MTurk was quickly viewed as a source for open services and uses as its worker base exploded through 2006. Cohen allowed other companies to come into the MTurk site as *requesters* and take advantage of the extraordinary low pay rate labor pool known as *workers*. MTurk acts as a self-serve portal for both requesters and workers. It is fairly simple to use. One only has to enter the website and establish an account as a requester or worker. Workers can browse available tasks on the MTurk pages, determine if they are qualified, take a qualification test, or simply submit to the available work. Workers generally choose work based on requirements, qualifications, and pay available. Workers are rated based on their performance on past work. If a worker has a poor rating they may not be contracted or paid (Buhrmester et al., 2011). Amazon charges at least a 10% commission and additional fees, and workers are paid either automatically or manually from a requester account that was funded using the requester's credit card (<https://www.mturk.com>).

MTurk has become a focal point for researchers in recent years. The ability to acquire sample data at lower costs and higher speeds has opened a panacea of opportunity for research. Without question, companies performing survey research have been waiting for an opportunity like this for many years. The information gathered from consumers has marketing research companies and academics re-writing strategies to extract innovative ideas, conduct market studies, and research consumer behavior in general. The crowdsourcing literature yields many examples of academic and research institutions studying crowdsourcing for purposes of future research. There are different ways researchers can approach MTurk for data and subjects. This is extremely important as

access to reliable and useable subjects and a well-planned experiment designs are critical to research success. As Rubin (2001) noted:

Arguably, the most important feature of experiments is that we must decide on the way data will be collected before observing the outcome data. If we could try hundreds of designs and for each see the resultant answer, we could capitalize on random variation in answers and choose the design that generated the answer we wanted! The lack of availability of outcome data when designing experiments is a tremendous stimulus for “honesty” in experiments and can be in well-designed observational studies as well. (p. 169)

Random assignment used for experimental design is often considered impractical from the standpoint of time and cost. As an alternative, researchers often develop quasi-experiments. The primary difference between random and quasi-experiments is that in quasi-experiments, subjects are not randomly selected. This calls to attention internal validity issues, namely, selection bias. According to Stuart and Rubin (2008), to get closer to random assignment experimentation using a quasi-experimental method, social scientists try to match both treatment and control group characteristics to obtain unbiased estimates of the effects of interest. Matching methods are increasingly being used to closely replicate the properties of randomization in experiments using observational data. In theory, matching selects subsamples of the treated and control groups that are only randomly different from one another. This ensures that subsamples are balanced with observed covariates. A second method of matching is a two-step process. In step one, subjects are chosen without reference to any values or outcome data. The second step then estimates outcome data estimating treatment affects. Matching methods can be

considered a legit manner to ensure a researcher is selecting the most appropriate data for an observational study.

Azzam and Jacobson (2013) conducted research to explore the viability of using online crowdsourcing for creating matched group comparisons. This study used MTurk to request workers to complete a survey that measured college students' levels of academic motivation and satisfaction in a quasi-experimental manner ($N = 500$). A matched-comparison group design yielded no significant difference between treatment and comparison groups, indicating that they were able to duplicate the findings of the randomized control group. Crowdsourcing, more specifically MTurk, sampling did have the potential to reliably create matched group comparisons.

In psychometric research conducted by Buhrmester and colleagues (2011), it was concluded that participants could be recruited easily and inexpensively ($N = 500$). Overall, participation rates were sensitive to compensation and time commitment. The range of compensation for these studies was from \$.02 to \$.50. Quality was not affected by pay. Further, mean alphas were good to excellent ($\alpha = .73$ to $.93$). Two different tests were conducted to determine a test re-test reliability rating. Again, results were very high ($r = .80$ and $.94$). These results are very comparable to results found using traditional methods.

It is possible that crowdsourcing was conceived by the needs of research. InnoCentive developed their version of crowdsourcing in 2001. The rationale behind the development of their worker base was to assist in product development and innovation and to answer scientific questions that required research. Morris and Picard (2012) used crowdsourcing to develop and outline a system to create a tool that could be used in

behavioral sciences for therapy. This tool would solicit emotional feedback from crowdsource workers. Goals for this study were to use crowdsourced systems for: (a) empathizing, (b) detecting cognitive distortions, and (c) crafting relevant cognitive reappraisals. Morris and Picard (2012) found that MTurk workers from the United States have little trouble generating sympathetic responses when instructed to do so. As the authors noted:

Our results support the hypothesis that, with guidance, crowdworkers will help craft empathetic reappraisals for strangers. By contrast, when told to simply make a person feel better, crowdworkers are less likely to be empathetic and offer reappraisals... We conclude that MTurk workers can reliably identify cognitive distortions within short, one-to-three sentence descriptions. With minimal instructions, MTurk workers seemed to understand the concept. (Morris & Picard, 2012, pp. 5, 7)

Product Evaluation

As mentioned, crowdsourcing has also been used for purposes of evaluation. With the ability to inexpensively access consumer opinions and experiences, it is no surprise individuals and companies use crowdsourcing to obtain product, marketing, and concept evaluations that can be very powerful in decision-making in today's business climate. Consumer packaged goods companies have developed a layered strategy to get information out of the market. From using their own social media to MTurk, these companies spend a lot of money and time evaluating products at conceptual stages in the marketing pipeline to determine if moving forward is appropriate. In general, other types of product evaluations take place that have also become very impactful in consumerism.

Because crowdsourcing facilitates feedback from a wide variety of viewpoints, it may be superior to a more traditional, highly structured judgment task (Le et al., 2010). Further, consumers learn more from reviews of a given product than they do from their own past experiences (Zhao, Yang, Narayan, & Zhao, 2013).

According to Stufflebeam and Coryn (2014), “Evaluation is a process for giving attestations to such matters as reliability, effectiveness, cost-effectiveness, efficiency, safety, ease of use, and probity” (p. 3). Product evaluations are performed in many arenas. Consumer Reports has been one of the leading product evaluators for many years. A hallmark and standard for many consumers, the goal of the Consumer Reports is to provide evaluations that determine if products are safe, effective, and reliable (Consumer Reports, 2015). However, Scriven (1994) identified several challenges in Consumer Reports’ product evaluations that seem to point directly to industry accepted evaluation standards not being applied. Consumer Reports does use a formal product evaluation process or method, but it is unclear how standards and criteria for such evaluation are established. Moreover, there are many common product evaluation shortcomings found in practitioner methods that do not make use of accepted standards (Scriven, 1994).

In addition to the challenges described above, there are also many different types of evaluation methods. In many instances, online product reviews are representative of an expertise-oriented evaluation approach. The contribution of the expert or professional can come in an array of ways including personal experience attestation, as a consultant to a contract holder of an evaluation, or as a participant on a team of evaluators. Another approach to evaluation is criticism and connoisseurship. This method also uses experts in a given area to provide a substantive evaluation based on appropriate criteria, and in

many cases, make judgments of value as well. According to Stufflebeam and Shinkfield (2007), “The methodology of the criticism and connoisseurship includes critics’ systematic use of their perceptual sensitivities, past experiences, refined insights, and abilities to communicate their assessments” (p. 184). Is this the type of evaluations we are seeing occur in crowdsource product reviews and evaluations? It may be more important to determine what, if any, evaluation standards are being deployed in crowdsource evaluations. Creating a survey instrument that employs a well-known and accepted set of evaluation standards to measure if, and to what degree formal evaluation standards are being used is a sensible approach to inform this body of research.

The Joint Committee on Standards for Educational Evaluation (JCSEE) program evaluation standards are well known in the education and program evaluation industry as *the* hallmark for evaluation standards (Yarbrough et al., 2011). They are not only used in program evaluation context, but extend to other contexts where formal evaluation is applied (Yarbrough et al., 2011). The JCSEE program evaluation standards provide a framework for evaluation standards in five distinct parts: (a) Utility, (b) Feasibility, (c) Propriety, (d) Accuracy, and (e) Accountability. Accordingly, these evaluation standards can easily be applied to general product evaluations. A key benefit of using the program evaluation standards is they were developed by a committee of professionals from evaluation, higher education, and education counseling, education measurement, education research, and program evaluation disciplines, with an additional 400 outside contributors and an independent evaluation panel designed to review and provide perspectives and validation to the legitimacy and usefulness of these standards. Simply,

there are no established and validated standards in use today that are as accepted and applicable as the JCSEE program evaluation standards.

The JCSEE program evaluation standards identify and define evaluation quality (Yarbrough et al., 2011). Yet, it is important to remember that standards are not laws, but are voluntary, consensus statements developed with extensive stakeholder input and then discussed, revised, and approved (Yarbrough et al., 2011). The JCSEE program evaluation standards seem to brand an evaluation with “quality and accuracy first” by ensuring the evaluation is centered properly in Utility, Feasibility, Propriety, Accuracy, and Accountability domains. Table 1 below identifies and describes the evaluation standards that were applied in this study.

Table 1

JCSEE Program Evaluation Standards

Standard	Definition
Utility	
Evaluator credibility	Evaluations should be conducted by qualified people who establish and maintain credibility in the evaluation context.
Attention to stakeholders	Evaluations should devote attention to the full range of individuals and groups affected by the evaluation.
Negotiated Purposes	Evaluation purposes should be identified and revisited based on the needs of the stakeholders.
Explicit values	Evaluations should clarify and specify the individual and cultural values underpinning the evaluation purposes, processes, and judgments.
Relevant information	Evaluation information should serve the identified and emergent needs of the intended users.
Meaningful process and products	Evaluation activities, descriptions, findings, and judgments should encourage use.
Timely and appropriate reporting	Evaluations should attend in a timely and ongoing way to the reporting and dissemination needs of stakeholders.

Table 1—Continued

Standard	Definition
Feasibility	
Project management	Evaluations should use effective evaluation strategies
Practical procedures	Evaluations should be practical and responsive on how the program works.
Contextual viability	Evaluations should recognize, monitor, and balance the cultural and political interests and needs of individuals and groups.
Resource use	Evaluations should use resources effectively and efficiently.
Propriety	
Responsive and inclusive orientation	Evaluations should be responsive to their stakeholders and communities
Formal agreements	Evaluation agreements should be negotiated to make obligations explicit and take into account the needs, expectations, and cultural contexts of clients and other stakeholders.
Human rights and respect	Evaluations should be designed and conducted to protect human and legal rights and maintain the dignity of participants and other stakeholders
Clarity and fairness	Evaluations should be understandable and fair in addressing stakeholder needs and purposes.
Transparency and disclosure	Evaluations should provide complete descriptions of findings, limitations, and conclusions to all stakeholders, unless doing so would violate legal and propriety obligations.
Conflicts of interests	Evaluations should openly and honestly identify and address real or perceived conflicts of interests that may compromise the evaluation
Fiscal responsibility	Evaluations should account for all expended resources and comply with sound fiscal procedures and processes.
Accuracy	
Justified conclusion and decisions	Evaluation conclusions and decisions should be explicitly justified on the cultures and contexts where they have consequences.
Valid information	Evaluation information should serve the intended purposes and support valid interpretations.
Reliable information	Evaluation procedures should yield sufficiently dependable and consistent information for the intended uses.
Information management	Evaluations should employ systematic information collection, review, verification, and storage methods.

Table 1—Continued

Standard	Definition
Accuracy	
Sound designs and analyses	Evaluations should employ technically adequate designs and analysis that are appropriate for the evaluation purposes.
Explicit evaluation reasoning	Evaluation reasoning leading from information and analysis findings, interpretations, conclusions, and judgments should be clearly and completely documented.
Communication and reporting	Evaluation communicating should have adequate scope and guard against misconceptions, biases, distortions, and errors.
Accountability	
Evaluation documentation	Evaluations should fully document their negotiated purposes and implemented designs, procedures, data, and outcomes.
Internal metaevaluation	Evaluators should use these and other applications standards to examine the accountability of the evaluation design, procedures employed, information collected, and outcomes.
External metaevaluation	Program evaluation sponsors, clients, evaluators, and other stakeholders should encourage the conduct of external metaevaluation using these and other applicable standards.

Source: Yarbrough, D. B., Shulha, L. M., Hopson, R. K., & Caruthers, F. A. (2011). *The program evaluation standards: A guide for evaluators and evaluation users* (3rd ed.). Thousand Oaks, CA: Sage.

Statement of the Problem

More and more people are using online product reviews and evaluations (Shapiro, Chandler, & Mueller, 2013); however, there is no indication of the quality of these reviews. Few studies have explored evaluation practices among crowdsource organizations. As a result, it is unclear what, if any, evaluation standards are used by crowdsource reviewers, particularly those found on open, self-serve sites such as MTurk. Without a clear understanding of these standards, there is little foundation from which crowdsource sites can explicate and refine strategies used in online product reviews and evaluations. Additionally, an absence of clearly identified standards potentially subjects

consumers to wide variations in the trustworthiness and reliability of evaluations provided by crowdsourced reviewers. Therefore, more research is needed concerning evaluation practices in crowdsourced organizations, which will give crowdsourced reviewers a basis from which they can define and improve the standards currently being implemented in their industry.

Purpose of the Study

The purpose of this study was to learn about crowdsourced evaluation standards and methods. Specifically, this study sought to determine: (a) what, if any, evaluation standards are being used by crowdsourced organizations and their requesters, and (b) to what extent these standards adhere to accepted evaluation standards used by professional evaluators in the evaluation discipline, precisely, the JCSEE program evaluation standards. Consumers may be interested in online or crowdsourced evaluations because they tend to be aggregated from consumers themselves; however, it is not the intent of this study to determine why consumers are using online and crowdsourced evaluations and reviews, but rather to discover what these evaluations represent and to what standards they are being completed.

Research Questions

Using data was collected from a web-based survey of workers and managers of the self-serve crowdsourced company MTurk, in order to explore the following research questions:

1. Are evaluation standards used in crowdsourced product evaluations? If not, why?

2. To what extent do crowdsource workers and management adhere to the five domains (i.e., Utility, Feasibility, Propriety, Accuracy, and Accountability) of the JCSEE program evaluation standards (Yarbrough et al., 2011)?

Definition of Terms

AMT: Amazon's Mechanical Turk (MTurk). See *MTurk* definition below.

Connoisseurship: Also referred to as *criticism*, connoisseurship is the art of appreciation. It is not necessarily a liking or preference for that which is observed, but rather an ability to notice, "to recognize differences that are subtle but significant in a particular qualitative display" (Alvin, 2004, p. 198). The connoisseur's perceptual acuity results largely from knowledge of what to look for (i.e., advanced organizers or critical guideposts) gained through extensive previous experience, education, and reflection on that experience.

Crowdsourcing: Crowdsourcing is where a problem or task is broadcast into a crowd of potential participants for solution, is used for an increasingly wide variety of tasks on the web. Of particular mention is the growing use of crowdsourcing in the context of evaluations or judgments (Dasgupta & Ghosh, 2013, p. 1).

Evaluation: Evaluation is both a process and a product. As a process, evaluation is the procedure through which an evaluator determines merit, worth, or significance. An evaluation is a product of that process. Professional evaluation is done in a systematic and objective way with a degree of expertise that requires extensive specific training or learning (Scriven, 2007, p. 3).

Evaluation criteria: Measureable dimensions and features of inquiry pertinent to the goals of an evaluation (Yarbrough et al., 2011).

Evaluation standards: Evaluation standards define evaluation quality and guide evaluators and evaluation users in the pursuit of evaluation quality (Yarbrough et al., 2011, p. xxii).

HIT: According to the MTurk website, an HIT (i.e., human intelligence task) is a question that needs an answer. It represents a single, self-contained task that a worker can work on, submit an answer, and collect a reward for completing (<https://www.mturk.com/mturk/help?helpPage=overview>).

MTurk: The Amazon Mechanical Turk (MTurk) is a crowdsourcing Internet marketplace that enables individuals or businesses (known as *requesters*) to coordinate the use of human intelligence to perform tasks that computers are currently unable to do (<https://www.mturk.com/mturk/help?helpPage=overview>).

Joint Committee on Standards for Education Evaluation (JCSEE): The Joint Committee represents a coalition of major professional associations formed in 1975 to help improve the quality of standardized evaluation (Yarbrough et al., 2011).

CHAPTER II

REVIEW OF LITERATURE

As stated in Chapter I of this dissertation, more and more people are using online product reviews and evaluations (Chandler et al., 2013); however, there is no indication of the quality of these reviews. Few studies have explored evaluation practices among crowdsource organizations, and as a result, it is unclear what, if any, evaluation standards are used by crowdsource reviewers, particularly those found on open, self-serve sites such as MTurk. The purpose of this study, therefore, was to determine: (a) what, if any, evaluation standards are being used by crowdsource organizations and their requesters, and (b) to what extent these standards adhere to the Joint Committee on Standards for Education Evaluation (JCSEE) program evaluation standards (Yarbrough et al., 2011). Chapter II reviews literature relevant to this purpose. Included in this review is a definition of the term *product evaluation*, and a discussion of the history of product evaluation activities, online product reviews, crowdsourcing, MTurk, professional evaluation, and problems associated with online reviews and crowdsourcing.

The History of Product Evaluation

The proverbial economic landscape in the United States shifted by the mid to late 1800s. With the advent of rail, producers had a more national market to serve. Accordingly, producers began to innovate, develop, produce, and deliver products across the continent. One outcome of this new business model was the loss of a personal relationship with the local producer, merchant, or distributor. This opened the door for fraud and misrepresentation. Furthermore, prior to 1927, there was not a government agency or advocate dedicated to the consumer to helping them choose between producers

and advertisers (Williams, 1995). A lack of information transfer and laws to protect consumers left the door wide open for consumers to be swindled and misled (Strach & Russell, 2003). As Strach and Russell (2003) noted, “Information about reliability, safety, and other characteristics of consumer goods had become vitally important to even working-class families, but neither the courts nor regulation held companies to their advertised word” (p. 152). Eventually, however, forms of consumer product evaluation did become available to the public. Major events relevant to the history of consumer product evaluation are discussed in the sections below.

Good Housekeeping

Simultaneous to the shift in market strategy by producers was an explosive emergence of monthly magazines fueled and funded by producer advertising dollars. By the late 1800s producers were establishing brands and spreading their messages, and consumers were able to see what could be purchased beyond their local borders from the following journals: *Good Housekeeping* (1885), *McCall's* (1870), *Popular Science* (1872), and *Cosmopolitan* (1886), to name a few (Strach & Russell, 2003). Clark W. Bryan was the founder and publisher of *Good Housekeeping* (Strach & Russell, 2003). According to Strach & Russell (2003), Bryan's strategy was to differentiate himself from the other monthly journals. In the beginning, much of *Good Housekeeping* was dedicated to food issues. Bryan also left editorial space to go after shady manufacturers and false advertisers. Eventually, with the passing of Bryan in 1898, the magazine was purchased by Phelps Publishing Company in 1900. In 1902, it introduced the Experiment Station, and began testing the reliability of materials and equipment designed for the household consumer. According to Strach and Russell (2003):

Producers paid a fee for the testing; the products winning approval were listed in each issue. Soon the magazine only accepted advertising from producers in this group, and regularly published an inflexible contract between the publisher and each subscriber. Popularized as the Ironclad contract, it included a money back guarantee (or more accurately guaranty) of the reliability of the advertisements. (p. 150)

By 1910, 200 products were evaluated and qualified to carry the Good Housekeeping Seal of Approval. Through 1941, the Good Housekeeping seal was possibly the most recognized guide in consumer products. Shortly thereafter, Good Housekeeping leveraged itself further by guaranteeing a replacement or refund for any product advertised in the magazine. Over the years, some changes had to be made to both the seal and the expressed warranty, but without a doubt, this was an early and quintessential example of product evaluation driven by innate concern for consumers. Unlike the other women's magazines, however, *Good Housekeeping* further established its claim of expertise through its campaign to test and guarantee advertised products with an in-house seal of approval. Thus, the magazine presented itself as an authority both in its article and advertisement copy (Chuppa, 2005).

Consumers Research

Another, perhaps more influential, example of independent product evaluation came from Stuart Chase and Frederick J. Schlink's (1927) book *Your Money's Worth*. This book was written to reveal fraud and misrepresentation by manufacturers that were misleading the general public about product construction or performance by providing independent evaluations of a product's ability (Williams, 1995). As described by

Williams (1995), “This book exposed the excesses of advertising and the hazards of shoddy products, and called for independent product testing” (sec. Preface, para. 4).

In the same year his book was published, Frederick J. Schlink founded the Consumers Club in New York City in 1927. In 1929 with a subscription list of 5,000 and a \$10,000 grant, the Consumers Club was leveraged to become Consumers Research (later to become Consumers Union, then Consumer Reports), a facility dedicated to testing consumer products and publishing the results in a bi-monthly publication called the *General Bulletin* (Williams, 1995). Consumers Research had one primary goal—protect the consumer (<http://consumersunion.org/about/mission/>). Specifically, its intent was to protect consumers from companies selling products with little to no value, or that were dangerous to use. Consumers Research was a major influence in the development of the Consumer Advisory Board under Franklin D. Roosevelt. The goal of the advisory board at that time was to prevent monopolies, and also protect the consumer against amount of product data back to the consumer using a monthly magazine called *Consumer Reports* (<http://consumersunion.org/about/mission/>).

Most individuals have consulted a *Consumer Reports* magazine issue from time to time before purchasing a new technology. *Consumer Reports* assesses many appliances, cars, tools, and other products in its National Testing and Research Center in Yonkers, NY. The testing center is the largest nonprofit educational and consumer product-testing center in the world where the testing of various brands of products takes place.

Consumers Union researchers assess the various models of a product to determine which product is the best value, the most effective, or some other criterion (Kelley, 2010).

The Federal Trade Commission

The Federal Trade Commission was created in 1914. According to Waller, Brady, and Acosta (2008), the FTC has two principal goals:

1. To protect consumers by preventing fraud, deception, and unfair business practices in the marketplace, and
2. To maintain competition by preventing anticompetitive business practices (p. 3).

Seemingly, the FTC was created to protect consumers and from fraud, deception, and unfair business practices. However, in those earliest years, the buzz and major role of the FTC was to thwart anti-trust activity, monopolistic ventures, price fixing, and other collusive acts. It does not appear from literature that much substantive time was spent on consumer product evaluation and advertising fraud (Hofstadter, 1965).

Underwriters Laboratory

A final example of independent product evaluation that offers perhaps one of the most rigorous and respected seals of approval is Underwriters Laboratories. Officially organized in 1893 by William Henry Merrill, this company was started as a result of fires breaking out at the Chicago World Fair in 1893. Merrill, then an electrical inspector, was called in to determine how to deal with thousands of Edison bulbs and wiring spaghetti found throughout the fair grounds and their associated risks of fire. Needing a place to set up shop and determine with some degree of confidence the real issues surrounding the problems, Merrill used money from the fire insurance underwriters to fund a local laboratory with basic testing equipment. As described by fundinguniverse.com:

Merrill and his staff of two completed 75 tests; in the first five years, they performed 1,000 tests-checking sockets, switches, wires, and a variety of supposedly noncombustible materials. In 1898 they published the first list of “approved fittings and electrical devices,” including flexible electrical cord and a snap light fixture. The approved products also received distinctive labels, indicating they had been inspected and certified free from reasonable safety hazards by Underwriters Laboratories. Thus was born the UL Mark.

(<http://www.fundinguniverse.com/company-histories/underwriters-laboratories-inc-history/>, Section A Testing Laboratory: 1894-1904)

Underwriters Laboratory went on to test much more than electrical and fire safety components, including safety matches, airplanes, gasoline pumps, ladders, X-ray machines, automotive fuel systems, and headlights (<http://ul.com/aboutul/history/>).

Again, as described by [fundinguniverse.com](http://www.fundinguniverse.com):

In 1998, more than 14 billion UL Marks appeared on new products worldwide. The UL staff has developed more than 600 Standards for Safety, 80 percent of which are approved as American National Standards. Testing and service fees from clients support the independent, not-for-profit organization.

(<http://www.fundinguniverse.com/company-histories/underwriters-laboratories-inc-history/>, para. 1)

Airplanes, for example, are granted airworthiness certificates for both private and commercial aircraft. This form of evaluation is primarily used to establish a list of test criteria, develop standards for the test process itself, and to determine if a product does

what it is intended to do more in terms of fitness and purpose, rather than how well it performs its purpose compared to another brand or product.

These companies—Consumers Union, Better Housekeeping, and Underwriters Laboratories—are the hallmark of product evaluation in the United States. Their primary purpose is to protect the consumer from unsafe products, but they also serve to debunk fraudulent claims about product capability, materials used, and features promoted. Throughout the history of these organizations, elaborate laboratories have been erected and staffed with engineers and scientists to ensure products are evaluated with an approach centered on effective measurement, not opinion. Furthermore, both Consumers Union and Better Housekeeping deliver monthly magazines to provide readers with as much information about latest products tested. Better Housekeeping also tests foods and uses its branding to guarantee results are accurate. In its early history, this was huge leverage for the Good Housekeeping team, as they were willing to more than their reputation on the line and back up their claims with money back guarantees. The purpose of the strategy was to show they were in it solely for the consumer. Underwriters Laboratories took a much deeper dive into testing and evaluating products. Not only did they do these activities for the consumer, but they also became a consultant to the U.S. government, including the Better Business Bureau.

Table 2 provides a historical review of product evaluation activities conducted by Better Housekeeping and Consumers Union. As shown in Table 2, this review clearly identifies a progressive pattern of both product evaluation achievement and the development of testing standards. Underwriters Laboratories was excluded from the table because they have a slightly different operating model. Although, their very existence

was to test products to ensure consumer safety, they were not necessarily consumer-driven, as they were owned and underwritten by an underwriting company and businesses of interest. However, Underwriters Laboratories, as documented earlier, had many historical product and vehicle evaluations that both helped consumers and informed producers of the shortcomings of their products.

Table 2

Historical Review of Good Housekeeping and Consumers Union Product Evaluation Activities

Decade	Good Housekeeping	Consumers Union
1900s	<ul style="list-style-type: none"> • 1900: Phelps Publishing Company purchases The Good Housekeeping magazine. • 1902: Good Housekeeping begins testing products. • 1909: The Good Housekeeping Research Institute (GHRI) is built to test and research products. By the end of 1910, nearly 200 products could carry the Better Housekeeping Seal of Approval. 	
1920s	<ul style="list-style-type: none"> • 1922: Good Housekeeping releases a British version. 	<ul style="list-style-type: none"> • 1927: Consumers Research is founded.
1940s		<ul style="list-style-type: none"> • 1941: Consumers Union builds a sound proof room to test radios. • 1942: Consumers Union changes the name of its magazine to Consumer Reports. • 1946: Circulation of Consumer Reports is 100,000.

Table 2—Continued

Decade	Good Housekeeping	Consumers Union
1950s		<ul style="list-style-type: none"> • 1950: Circulation of Consumer Reports is 400,000. Consumer demand of products is off the chart. • 1953: Consumers Reports publishes the first of a series of reports on the dangers of tar and nicotine in cigarettes. This monumental is it was the only source of this information at the time. • 1954: Consumers Reports tests its first color TV set. • 1956: Consumer Reports starts testing seatbelts.
1960s	<ul style="list-style-type: none"> • 1961: Good Housekeeping urges readers to fit car seats with seatbelts front and back, seven years before it became law. 	<ul style="list-style-type: none"> • 1965: The first Toyota Corolla shows up on American shores. Consumers Reports tests it and gives it a “thumbs up.” • 1967: Ralph Nader writes, <i>Unsafe at Any Speed</i>. He later joined the Consumers Union.
1970s	<ul style="list-style-type: none"> • 1970: Good Housekeeping propels the Care Labeling Act, which requires manufacturers to include washing instructions on their garments labels. 	<ul style="list-style-type: none"> • 1970: The National Commission on Product Safety recommends the establishment of the Consumer Product Safety Council.
1980s		<ul style="list-style-type: none"> • 1980: Consumers Union starts a TV Department. • 1987: The Auto Test Department moves to a 327-acre site in East Haddam, CT.
1990s		<ul style="list-style-type: none"> • 1992: Consumers Union is one of the largest magazines in circulation at 5,000,000. • 1992: Consumers Union becomes available through AOL. • 1995: Consumers Union produces its first multimedia CD-ROM, on automobiles.

Table 2—Continued

Decade	Good Housekeeping	Consumers Union
2000s	<ul style="list-style-type: none"> • 2000: Good Housekeeping tests kids’ bicycle helmets on the market and a large percentage fail the Consumer Product Safety standards. • 2005: GHRI reveals lead in children’s toys and jewelry. 	<ul style="list-style-type: none"> • 2002: Consumers Union has over 800,000 subscribers to its website. • 2010: Consumers Union launches a mobile app, Consumer Reports Mobile, that allows product ratings and rankings to be available easily using the mobile app. This mobile app allows the consumer to scan the product in the store and get reviews and ratings on the product instantly.

Sources: Consumers Union website, Underwriters Laboratories website, Consumers Research website.

The Current Situation of Product Evaluation

The science of evaluation is multifaceted. As Scriven (1994) noted, “Evaluation is a transdiscipline and developing its methodology involves the same dilemma that has been faced across two millennia by logic, another great transdiscipline” (p. 47). Evaluation is as much a philosophy as it is a method. Like personalities, no two evaluations are exactly the same –they have different people, situations, and outcomes. There is not a set of generally accepted evaluation principles such as those found in other disciplines like accounting, which has the Generally Accepted Accounting Principles (GAAP), by which all accountants *must* adhere to by law (<http://www.fasab.gov/accounting-standards/authoritative-source-of-gaap/>). Instead, evaluations can be as simple as a comparison between a standardized color plaque and a visual automotive plastic part to determine if the part has the correct color hue, or as complex as determining if healthcare programs among seniors are effective in lowering healthcare costs while improving quality of life for critically ill patients.

In general, an evaluation can be performed using many different approaches perform a complete, systematic, and authentic evaluation (JCSEE, 1994, p. 3). Fournier (1994) posits that the general logic of evaluation consists of common reasoning and rules that can guide and evaluation practitioner through a logical process of evaluation. Below is a basic outline of this logic model from Fournier (1994):

1. *Establishing criteria of merit.* On what dimensions must the evaluand do well?
2. *Constructing standards.* How well should the evaluand perform?
3. *Measuring performance and comparing with standards.* How well did the evaluand perform?
4. *Synthesizing and integrating data into a judgement of merit or worth.* What is the merit or worth of the evaluand? (Fournier, 1994, p. 16).

The general logic can be found across various instances of the evaluation inquiry process. For example, the numerous evaluation approaches developed by theorists vary from one another in many details; yet, the researcher finds that they do share this common logic (Fournier, 1994).

Overall, the definition of evaluation is codependent upon its context and application (Scriven, 1993). As mentioned earlier, there are many dimensions in an evaluation. Some evaluation approaches become inadequate and are discarded by practitioners; thus, illuminating the fact that evaluation is a fluid and dynamic profession in terms of methods and approaches. A small, yet important example comes from the field of education. Education programs are among those in which evaluation has a high stakes, prominent role. The emphasis on education programming was high in the mid to

late-1960s with the government rollout of new Elementary and Secondary Education Act of 1965. Simultaneously, there was a need for more accurate and complete program evaluation by the U.S. government. Lee Cronbach, an educational psychologist, penned an article that criticized prevailing evaluation techniques, which were heavily reliant on norm-referenced test result comparisons between experimental and control groups (Stufflebeam & Coryn, 2014). One byproduct of this evaluation focus, therefore, was older evaluation approaches being called into question, and in some cases, replaced by new more adequate approaches (Stufflebeam, 1968).

While the education example above provides an illustration of the changing nature of evaluation, this dissertation is interested in post-purchase consumer product evaluation. As mentioned, consumer product evaluation has entered a new era of online product reviews. Traditionally, as Scriven (1994) noted, “Product evaluation is probably the best-developed and oldest practice within evaluation, although explicit discussion of its methodology has not received as much attention as in the case of program evaluation” (p. 46). The following sections review some of the many different dimensions of consumer product evaluation, including online reviews such as those found on Amazon.com; independent analysis such as *Consumer Reports* or *Better Housekeeping*; government agency oversight such as the Consumer Product Safety Council and related organizations such as Underwriters Laboratories. Figure 1 summarizes the basic branches of product evaluation, providing context for the following section.

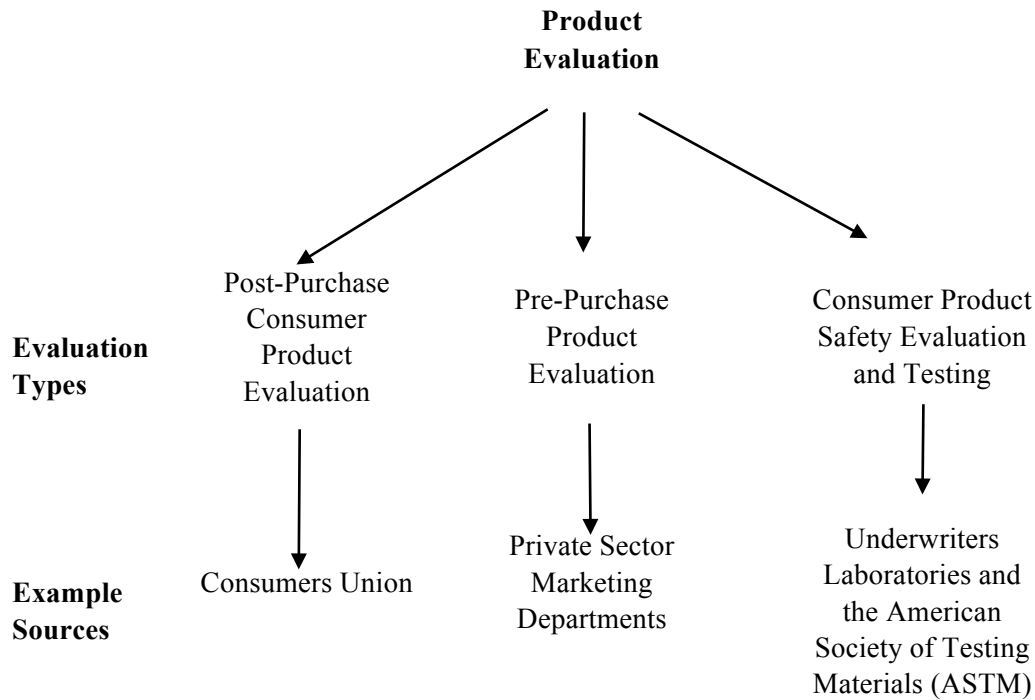


Figure 1. Basic branches of product evaluation.

Online Product Reviews

With the coming of the Technology Age, the world of product evaluation experienced an entirely new way to get feedback from consumers. Web-based reviews (i.e., online reviews) allow certain product evaluations to take place today with amazing speed and ability to influence nearly three generations of online and mobile users. Retailers, for example, are using online reviews to increase sales. They know consumers use the Internet in increasing numbers if not at home, then on the go using mobile apps. UNCTAD (2002) determined through survey research that online product reviews were “important” or “extremely important” in the buying decisions of half of consumers who visited retailer sites with consumer postings. For this reason, many websites are now designed with pages that allow consumers to register and provide online reviews for products (e.g., vehicles, batteries, sensors) and services (e.g., restaurants, oil changes,

beauty salons). These review applications and portals appear primarily in four different online contexts: (a) market sites, (b) producers or distributors, (c) review sites, and (d) social media.

Market Sites

Market sites are designed to support or inform consumers about product and technology within a specific market segment. For example, PCMAG.com is a site specifically created to share information about all things related to personal computers. PCMAG.com and other similar sites are natural hosts for product reviews. On these sites, cross-sections of consumers who use these products “hangout” and read about new products or share information that informs their community.

Producers or Distributors

Websites created by producers or distributors such as Ford and Amazon also use product review applications. These sites in general receive heavy traffic from the consumer stream to begin with, and as expected, the addition of a product review portal on these sites has only increased traffic and sales. Amazon.com is a great example of a distributor site with online reviews. As Jiang and Wang (2008) noted, “Ever since Amazon.com published its first consumer book review in 1995, an estimated 43% of online retailers now offer consumer reviews or ratings on their websites” (p. 1). Customers can set up an account with Amazon and provide feedback on both the products they have purchased and also their level of satisfaction with the purchase experience. Amazon has arguably the largest product review site on the Internet, and puts considerable resources into preventing false reviews; nevertheless, apparently the power of these reviews has caused both consumers and producers to take fake reviews to

another level. In August 2012, *The New York Times*, for example, exposed publisher John Locke for purchasing 300 reviews for \$6,000 from a company owned by Todd Jason Rutherford (Streitfeld, 2012). As explained by *Forbes* magazine,

Locke couldn't have pulled off this fraud without some complicity from Amazon.com, however. Amazon.com has access to loads of information that is not available to the consumer (such as the IP addresses of reviewers and their purchase histories) that could help them identify fraudulent reviews and ban reviewers with a history of fraud. Sudden bursts of reviews (as happened with Locke) should also have sent up some red flags in Seattle. There's simply no excuse for letting Locke get away with these sorts of shenanigans.

(<http://www.forbes.com/sites/suwcharmananderson/2012/08/28/fake-reviews-amazons-rotten-core/>)

Review Sites

Review sites are designed to specifically for reviews only. Several online review sites have been established for restaurants, for example. These sites include yelp.com, urbanspoon.com, and yomoto.com. Of these restaurant review sites, Yelp seems to have the highest level of filtering and screening of false reviews, which seems to be a growing concern in the use of online reviews. Yelp claims it uses specific criteria (i.e., IP address, user history, and experience) to ensure that reviews are real and authentic. Interestingly, Yelp also offers a suite of products that can be purchased by companies to help improve their ratings.

Social Media

The use of social media to introduce new products and deploy mobile apps has become an amazing tool. Webmasters (i.e., managers of websites) are continuously looking for ways to get consumers to their sites. They know the faster and easier methods for gathering information and data transfer social media provides will bring more people to their sites and keep them there longer. Furthermore, mobile apps with the integration of smartphones have now become one of the largest growth technological segments, up 120 percent in 2012 over 2011 (Nielson, 2013). Use of these apps only furthers the ease through which companies and consumers can provide product reviews.

Overall, it is clear the Internet provides a significant proportion of today's product reviews and evaluations. Cisco Internet Business Solutions Group (2013) noted:

5,000 shoppers across five countries were asked to indicate the three most important sources of information they use for making buying decisions. Online ratings and reviews on retailer websites (52%) were included among the top three sources of information most frequently by respondents—ahead of advice from friends and family members (49%) and advice from store employees (12%).

Coupled with mobile apps like those found with CNET, PCMag, Yelp, and Consumer Reports, which make it faster and easier to get product reviews and information to consumers, it is, for the most part, easier than ever to access product reviews through the Internet, which should only accelerate online product review usage.

Crowdsourcing

According to Howe (2006), *crowdsourcing* occurs when a company uses the Internet to outsource activities that may have previously been or could be performed by

employees within the organization. For Example, in 2001, Alpheus Bingham, Vice President of e.R&D at e.Lilly, announced a new e-business venture called InnoCentive LLC. The purpose of InnoCentive was to use the Internet to create and enhance open-source scientific research and development. As described by the company:

We are at the heart and soul of what the Internet is all about. InnoCentive represents a return to the Internet's roots...an open-source approach to scientific collaboration and innovation. We are seeking access to that particular mind that is uniquely prepared for solving a specific scientific problem.

(<https://investor.lilly.com/releasedetail2.cfm?releaseid=52031>)

In this way, the venture was pioneering and changed the industry. This crowd effect ramped up during the early to mid-2000s; however, it was InnoCentive's approach that captured the attention of many, as it would use its audience to solve complex problems. This was new and quite compelling to the online community.

Prior to the availability of crowdsourcing, requesting companies had to fully employ laborers. This was expensive and contrary to the new millennial business model logic, which sought to decrease producer labor expense and increase productivity in a climate of world competition and pricing. Moreover, labor costs often varied based on availability and level of talent needed. The model of crowdsourcing broke barriers of scarcity by allowing workers to come and go as they pleased, and to take on smaller parts of work at a time vs. an entire project contract. This model also helped to attenuate scarcity of labor by using an open source model that attracted very large volumes of people at one time.

How crowdsourcing works is not nearly as complex as it may seem. A company identifies a list of tasks or work currently being accomplished in-house that it now wants to outsource. It then uses its own site or portal to organize workers and post and deploy available work or tasks. Workers then undertake the work or tasks with the expectation of being paid based on identified pay rates. Submitted work is paid for if it is acceptable. If the submission is not acceptable, the worker may or may not have the opportunity to redo work for further consideration. In some cases, more than one worker may work on the same work or task, and the company will choose only one best submission. Overall, as Hienerth and Rier (2012) argued:

The relevance of crowd-based evaluation is high, as some leading firms already utilize the crowd (e.g., InnoCentive, Apple, Facebook, Threadless, Kickstarter or Android). Rising numbers of ideas, concepts and solutions to be handled forces firms to consider crowds as an alternative source for evaluation.

With the ability to inexpensively access consumer opinions and experiences, consumer packaged goods companies have developed a layered strategy to get information out of the market. From using social media, their own websites, and crowdsourcing companies like crowdsource.org, these companies spend a lot of money and time evaluating products at conceptual stages in the marketing pipeline to determine if moving forward is appropriate or if they should take initial concepts to phase two design and development with the help of “crowd” input. This information is fast, accessible, and inexpensive. In general, multiple types of product evaluation take place that are very impactful in consumerism. Because crowdsourcing facilitates feedback from a wide variety of viewpoints, it may be superior to a more traditional, highly structured

judgment task (Biewald et al., 2010). Further, consumers may actually learn more from reviews of a given product than they do from their own past experiences (Zhao et al., 2013).

Open source crowdsourcing is open to the public, where both *workers* (person performing the work) and *requesters* (person wanting work performed) can enter a site (like MTurk), not have a single conversation, and meet each other's needs and expectations by mutually performing specific tasks and operations consistent with the crowdsourcing sites operational scope and work requested by the requester. Further, both the worker and the requestor have specific pre-requisites that may need to be met prior to engagement. In general, a worker wants to make a certain pay rate per hour, be able to accomplish the tasks asked to keep his or her ratings up, and get paid. A requester wants a worker with good ratings (ratings are derived by requester feedback) and possibly a host of other filters like demographics, work experience, pay rate, etc. All of these conditions and parameters can be entered into the sites programming, and the output is a posting on a board the workers use to read about opportunities and determine if the opportunity is for them or not.

MTURK: The First Open to the Public Crowdsourcing Model Called Open Sourcing

MTurk is a wholly owned subsidiary of Amazon. It is an open source crowdsourcing site; that is, it is open to the public. This site was created by and is maintained by Amazon in the United States. Jeff Bezo, Founder and CEO of Amazon.com, was so impressed with the advantages of using crowd labor he directed Peter Cohen, Director of Corporate Development Amazon Web Services, to create a model within Amazon to solve its own problems with e-commerce webpages that could

not be resolved systematically using computers

(<https://www.mturk.com/mturk/help?helpPage=overview>). In November 2005, Cohen announced the release of a beta version of MTurk, a worker exchange site that would allow basically trained individuals to perform human intelligence tasks (HIT) in order to scan pages of code, products, product descriptions, and specifications to identify and cull redundancy. Amazon's MTurk has a three-pronged strategic approach designed to bring together community, technology, and compensation (www.amazon.com).

Named after a 17th century hoax, MTurk represents the use of human intelligence through a mechanical device. The Turk, created by a Hungarian nobleman Wolfgang Von Kempelen, was essentially a wooden box with a human figure on top that would challenge people passing by to a game of chess as it toured Europe. Well known mentalists of the time, Benjamin Franklin and Napoleon, were both beaten by the Turk. Yet, it wasn't the mechanical human figure atop the box that was winning chess on the streets in Europe in the 1760s—it was the chess master hidden in the box below (Howe, 2006).

Created to solve internal webpage problems at Amazon, MTurk was quickly viewed as a source for open services and uses as its worker base exploded through 2006. Cohen allowed other companies to come into the MTurk site and take advantage of the extraordinary low pay rate labor pool known. MTurk acts as a self-serve portal for both *requesters* and *workers*. It is fairly simple to use. One only has to enter the website and establish an account as a requester or worker. Requestors can ask workers to perform a variety of tasks from completing surveys, participating in experiments, looking at pictures, reviewing data, scripts, etc. (Buhrmester et al., 2011). Workers can browse

available tasks on the MTurk pages, determine if they are qualified, take a qualification test, or simply submit to the available work. Workers generally choose work based on qualifications and pay available. Workers are rated based on their performance on past work. If a worker has a poor rating they will not be used or paid (Buhrmester et al., 2011). Amazon charges a 10 percent commission, and workers are paid either automatically or manually from a requester account that was funded using the requester's credit card (<https://www.mturk.com>).

MTurk has become a focal point for researchers in recent years. Paolacci and Chandler (2014) have conducted a great deal of research concerning MTurk workers. In general, researchers can use MTurk for any study that is suitable to be conducted online (Paolacci & Chandler, 2014). The ability to acquire sample data at lower costs and higher speeds has opened a panacea of opportunity for research. Without question, companies performing survey research have been waiting for an opportunity like this for many years. The information gathered from consumers has marketing research companies and academics re-writing strategies to extract innovative ideas, conduct market studies, and research consumer behavior in general. The crowdsourcing literature yields many examples of academic and research institutions studying crowdsourcing for purposes of future research. There are different ways researchers can approach MTurk for data and subjects. This is extremely important, as access to reliable and useable subjects and a well-planned experiment designs are critical to research success. Although data quality can be defined in several ways, research assessing MTurk on dimensions universally relevant to researchers supports the idea that worker samples are reliable (Paolacci & Chandler, 2014).

Advances and Continued Areas for Growth in Crowdsourcing

As indicated earlier, InnoCentive began to use the crowd for problem solving and task processing. Today, InnoCentive has taken that even further. They are now quite possibly the largest open innovation crowdsource location on the web with over 250,000 members. People are free to enter the site and register. During registration, they submit their credentials and interest and are exposed to an array of opportunities, including problem-solving, collaborating, innovation, or just hanging out and learning. It is a dynamic culture that is easy to access and participate in. Moreover, InnoCentive is an iconic American invention that once again will help shape the way the world operates. In this site, collaborators can also register for contests that challenge innovation, talent, and creativity.

Beyond advances in crowdsourcing itself, research on crowdsourcing is growing as well. Yet, while the number of studies on crowdsourcing is growing, the literature base on this topic is still relatively under-established. Zhao and Zhu (2014) found that only 55 crowdsource academic articles were retrievable for a crowdsource study they conducted. Furthermore, Zhao and Zhu (2014) found there was not a particular theoretical orientation common among crowdsourcing studies, which may be another sign of immaturity in this area of research.

Crowdsourced research samples are unique. If a participant possesses the qualifications required, he or she may or may not elect to participate based on several factors. Recent work postings are more likely to get responses (Chilton, Horton, Miller, & Azenkot, 2010). Compensation plays a role in participation as well. Workers are looking at and reviewing work that pays more first (Horton & Chilton, 2010). In

psychometric research conducted by Buhrmester and colleagues (2011), it was concluded that participants could be recruited easily and inexpensively ($N = 500$). Overall, participation rates were sensitive to compensation and time commitment. The range of compensation for these studies was from \$.02 to \$.50. Quality was not affected by pay. Further, mean alphas were good to excellent ($\alpha = .73$ to $.93$). Two different tests were conducted to determine a test re-test reliability rating. Again, results were very high ($r = .80$ and $.94$). These results are very comparable to results found using traditional methods.

Problems Associated With Online Reviews and Crowdsourcing

Quality of reviews is a primary problem associated with online reviews and crowdsourcing. It appears that the use of online reviews increases producer sales. Forman, Chris, Ghose, and Wiesenfeld (2008), for example, determined more attention or reviews is better, not necessarily dependent on positive reviews, but volume. This scenario is problematic for evaluation accuracy. Sales increase with more reviews; however, reviews may be quick and inaccurate, especially if done anonymously on sites that do not require an account to write a review. As Jacobsen (2015) noted, “I find clear evidence that consumer ratings are swayed by the reviews of experts, increasing their ratings in response to positive expert reviews and decreasing them in response to negative expert review.” In addition, there are inconsistent results coming from online reviews compared to product evaluations. “For models that have extremely high or extremely low quality, reported ratings differ substantially from the ratings reported by *Consumer Reports*” (Chen, Fay, & Wang, 2003). The reason for this difference is unclear, but the fact it exists is concerning. It may be that a product evaluation with rigor determined that

a product was very reliable, or the opposite, where an online reviewer with little empirical knowledge could rate the product diametrically opposite the product evaluator.

Another problem associated with online product reviews is objectivity. What about sites such as cnet.com, pcmag.com, engadget.com, and kbb.com? None of these sites are producers, however, producers through advertisements support all of them (Chen et al., 2003). Furthermore, consumer review sites often are either sponsored by retailers, or accept referral fees or advertising from manufacturers and retailers, all of which raises an issue of objectivity (Chen et al., 2003). Is it possible companies can or will game the system? Mayzlin (2002) posited that retailers do, in fact, use websites to promote and even sponsor forum chat to increase product awareness.

A third problem with online product reviews is small sample size. Overall ratings can be significantly affected when there is a small sample pool. As Chen et al. (2003) noted, “The primary benefit of increasing the number of reviews occurs when the sample size is so low that it can be easily manipulated by one or a few biased reviews” Conversely, when there is a large sample pool and ratings which are negative, they have very little effect on the overall product rating (Chen et al., 2003). When products have a superior product rating, overall product ratings will be high. Nevertheless, compared to traditional survey methods, online reviews seem to be more accurate. Furthermore, for the user, they are less expensive and easier to access (Chen et al., 2003).

A final problem with online product reviews is a lack of veracity. “I was in need of teeth whitening and my friend referred me to Southland Dental,” begins a thumbs-up for a clinic in Sherman Oaks, California. Then, there’s a description of the whitening procedure favored by Southland and this closer: “Pain or no pain, it was very much worth

it. I can't stop staring at my bright smile in the mirror." What sounds like a positive review is not a review at all:

This reads like a rave on Yelp, but it is actually a sample from a help-wanted ad on another site — specifically, Mechanical Turk, a Web site owned by Amazon.com and a place where companies invite “Mechanical Turk workers” — thousands are registered, worldwide — to complete what could be described as microtasks. Each task pays a tiny sum. In the case of Southland Dental, workers were asked to write a fake, five-star review and post it to Southland's Yelp page, for which they would earn 25 cents. (Segal, 2011, http://www.nytimes.com/2011/05/22/your-money/22haggler.html?_r=0, para. 4)

Post-purchase Consumer Product Evaluation Approaches

Each individual or company has their own procedure for evaluating a product. There is very little refereed evaluation literature about consumer product evaluation methodology. Michael Scriven appears to be the primary contributor and source of leading consumer product evaluation methodology in the United States. In this dissertation I attempted to look closely at key approaches Scriven mentions in his 1994 “Product Evaluation –The State of the Art” submission to *Evaluation Practice*. These approaches seem to stem from the following evaluation methodologies or models: alternative product comparisons, goal-based performance evaluations, Michael Scriven's Key Evaluation Checklist, the goal free approach, and connoisseur product evaluation.

Evaluation by Alternative Product Comparison

Consumer buying behavior is complex. As a result, there is a multitude of theory and idealism approaching the questions of how and why consumers make their product

choices. Abelson and Levi (1985) posited the consumer wants a product that will comparatively give him or her the best value. They further posited consumers are rational and will seek a process that effectively determines the truth, and consumers will cognitively and systematically rank and assign weight to purchase decision criteria (Abelson & Levi, 1985).

Methodology for consumer product comparison is rather straight forward, and consists of identifying pricing, physical, operational, and outcome specific criteria. More specifically, the products must be able to have the same basic characteristics and share the same expected performance outcomes (Scriven, 2007). With respect to product evaluation, Scriven (1994) arrives at the very same elements to inform a post-purchase product evaluation as those used in comparisons to other products. Specifically, Scriven (1994) described this process as, "...identify and validate criteria of merit, determine performance on those criteria, and combine the two according to some valid principle of integration" (p. 42). Further, Scriven (1994) sought to evaluate products by way of comparison to establish a much closer determination of strengths and weaknesses of key product attributes:

Since the consumer is almost always facing a choice between products, this improves utility; but also increases validity, since weaknesses and strengths emerge much more clearly in a horse race than in the stable yard. The overall result is a wonderfully useful, although complex, set of resources –if you know how to get to them and use them. (p. 46)

Overall, based on this framework, the primary tasks of a product evaluation can be summarized as identifying criteria and establishing a standard to judge those criteria.

Goal-based Performance Evaluations

Authors van Osselaer and Janiszewski (2012) describe a *goal* as a motivational cognitive concept that encourages expected behaviors. Goals can represent desired outcomes. These desired outcomes are related to some collection of needs or benefits to the consumer. In describing the goal-based choice model, van Osselaer and Janiszewski (2012) noted:

The goal-based choice model represents three types of concepts as nodes in a connectionist network: means (products), goals (benefits), and an outcome node that represents the overall evaluation. Similar to other concepts in declarative memory (Anderson et al. 2004), the activation of goals and means nodes consists of two parts, base-level activation and incoming activation from other nodes. In the case of goal nodes, incoming activation can come from means nodes or from situational cues. In the case of means nodes, incoming activation can come from goal nodes or situational cues. The evaluation node is a simple output node whose value depends on (1) predictive associations between the means and the goals and (2) the activation of the goals. The model operates by decision episode. That is, the model divides the flow of behavior into chunks consisting of a single decision or choice. Our main focus is on the decisions consumers make and how they learn from the consumption experiences that result from those decisions. (p. 262)

As an example of goal-based evaluations, Scriven (1994) mentioned several points for the improvement of American auto-companies to compete better with off-shore competitors. Specifically, Scriven (1994) noted that improvements could be made as it relates to price, reliability, convenience, and performance. These items were what

Scriven (1994) thought were the main failure points of U.S. carmakers. Improvement on these and other criteria can be viewed by an evaluator as the goals or needs of a comparative evaluation between U.S. automakers and those of off-shore.

Key Evaluation Checklist

The Key Evaluation Checklist (Scriven, 2013) is a tool that can be used as a standalone tool for evaluations, a guide to evaluation design, a cross reference tool used in addition to other evaluation models being deployed, or as a tool to evaluate an evaluation. The Key Evaluation Checklist (Scriven, 2013), for example, can be used evaluating programs, policies, writing reports about completed evaluations, and as a tool to assess a subjects evaluability.

Goal Free Approach

Michael Scriven's (1976b) goal free approach is a theory-based approach focused on not spoiling an evaluation's results by informing the evaluator of the goals of the program or product. Scriven (1976b) believed that goal free evaluation was the best method for determining side effects or unintended outcomes. For product evaluation, unintended outcomes could be many things: poor cost to benefit relationships, hazards, structural failures, unintended uses, or loss of market equity. However, Scriven (1976b) noted these unintended outcomes need to not only be judged against the goals, but also the needs of the consumer to be sure there is no unintended value that can be realized by the customer, even if the products intended goals are not hit.

The standards and specifications for goal free evaluations typically come from the following sources: the American Standards for Testing Materials (ASTM), the Consumer Products Safety Bureau, Underwriters Laboratories in conjunction with the Consumer

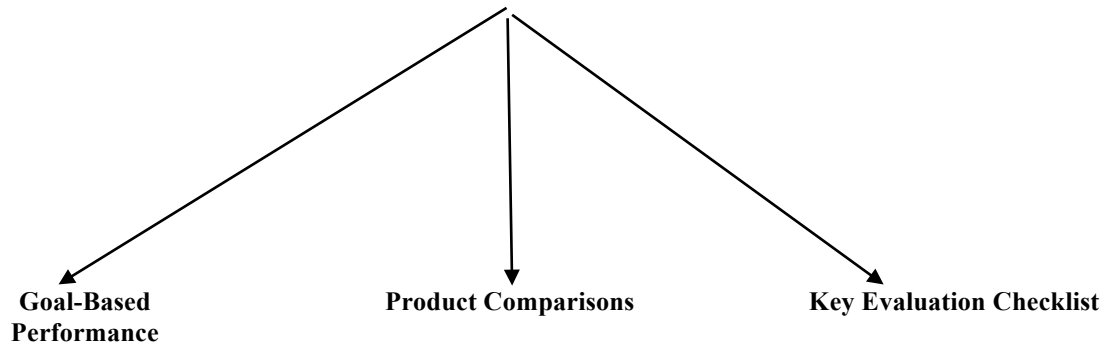
Products Safety Bureau, and proprietary standards established internally by the product evaluation center. The information and data obtained from the evaluation is used to inform consumer product safety personnel and help them make well informed decisions related to new products, existing products on the market, or products that require prior approval before commercialization. This information may also be used to develop new standards for testing, or metrics for pass/fail conditions of products or product types being tested. Further, this information may be used to inform consumers in both their purchase decisions, in the case of products currently on the market, and when informing consumers of what to do or where to go to get assistance on a faulty and unsafe products. Figure 2 summarizes Scriven's fundamental product evaluation approaches (Scriven, 2013).

Connoisseur Product Evaluation

A final approach is *connoisseur product evaluation*. According to Stufflebeam and Shinkfield (2007), "The methodology of the criticism and connoisseurship includes critics' systematic use of their perceptual sensitivities, past experiences, refined insights, and abilities to communicate their assessments" (p. 184). Is this the approach we are seeing used with online reviews and crowdsourced product reviews? It certainly seems to be. As Robbins (2006) noted,

Connoisseurship involves expert norm-referenced judgements being made by a person who is recognized as having the knowledge and experience necessary to do so. Examiner, observer, rater and assessor are all terms that may be used in different settings to describe this person.

Scriven's Fundamental Product Evaluation Approaches



Advantages

- Clear understanding of needs
- Easily identified
- Easily compared or benchmarked

Disadvantages

- Can goals be correct for all stakeholders?
- Focus on goals causes other subject impacts to be missed.
- Are standards for goals appropriate?

Advantages

- Provides a clear benchmark perspective
- Is easily accomplished
- Is understandable by stakeholders.

Disadvantages

- Are we comparing to a correct standard?
- If not THE standard why compare?
- What features should we compare? What are leaving out? Why?

Advantages

- Is a tool to address all aspects of an evaluation.
- Provides a framework for design.
- Can be easily used as a mixed method

Disadvantages

- Does not determine goals or standards for an evaluation.
- Cannot determine scope of criteria to be measured.
- Does not provide contextual methodology.

Figure 2. Scriven's fundamental product evaluation approaches.

This approach is used in evaluation or assessments of teachers, musicians, actors, firefighters, police, EMTs, medical personnel, food, wine, beer, and consumer products in general. Connoisseur product evaluation can be used in combination with other evaluation models as well. This approach is meant to be conducted by not only an expert

in the field, but a person close in the community of the object being evaluated or assessed, ensuring the context of the evaluation is not a barrier to success.

McAuley and Lekovec (2013) posited that connoisseur tastes change over time and experience, thereby possibly introducing an issue of rater-to-rater reliability within this approach. Their study was able to determine that there is a difference in agreement between product reviewers with less experience and those with greater experience. Further, expert ratings were easier to predict than less experienced reviewers, and inexperienced reviewers tended to high reviews to what were considered “low-average products” (p. 9).

Connoisseurs are without a doubt the larger population writing product reviews on the web today. They are providing reviews based on their experiences with products and services. However, the research conducted by McAuley and Lekovec (2013) clearly identified an inter-rater reliability issue between experienced and inexperienced product reviewers. They (2013) concluded:

Users’ tastes and preferences change and evolve over time. Shifting trends in the community, the arrival of new products, and even changes in users’ social networks may influence their rating behavior. At the same time, users’ tastes may change simply through the act of consuming additional products, as they gain knowledge and experience. Existing models consider temporal effects at the level of products and communities, but neglect the personal development of users: users who rate products at the same time may have less in common than users who rate products at different times, but whom are at the same stage in their personal evolution. (p. 10)

Scriven's Critical Review of Consumer Reports

As a consumer product evaluation icon in America, Consumer Reports (formerly Consumers Union) presents an important case study. To explain this case study, the researcher draws from Scriven's (1994) article, "Product Evaluation—The State of the Art," and his critical review of Consumers Union. At times this piece seems dated, yet it clearly calls out specific evaluation method and standard conflicts that are still prevalent today in work and publishing by Consumers Union, and typify the quandary of evaluation differences among practitioners. The article also clearly further points to the fact that there is not a generally accepted evaluation principle globally recognized. Being the consumer product evaluation flagship for our country Consumers Union is seemingly the standard of product evaluation. Inherent to such a large organization as this is the social responsibility to remain objective and immune to outside influences, both political and financial. According to Scriven (1994), a key shortcoming to Consumers Union has been its non-transparency to methodology. The following excerpt from Scriven's (1994) article, "Product Evaluation—The State of the Art," illustrates this:

- i. In some areas, the background knowledge and literature review is poor, contrary to a basic precept of any kind of research. A recent example is the coverage of kitchen knives in the August 1993, issue, which contained no durability testing and several errors that any attentive reader of American Blade or the mail-order knife catalogs would pick up...
- ii. CU often fails to tell consumers where to get items they test, which renders the results almost useless for many people, especially those who are homebound, inner city-bound, or live in remote areas.

- iii. Low-cost options too often get marginal comments, or no comments, instead of testing and rating. For example, their August 1993 eyeglass report did no testing of drugstore-bought eyeglasses, which at \$5-12 or so are a complete solution for many people who are paying an average of \$160.00 for prescription –nor of the \$50.00 prescription glasses from the warehouse stores, for example, Costco (there was a throwaway remark about them, but no testing)...

Other problems:

- iv. There are weaknesses in CU’s analysis of purchasing procedures: an example which costs consumers millions of dollars is their advice on car buying, which has stuck at the suggestion “get the dealers invoice price, and allow a few hundred dollars for profit,” which is non-optimal, although of course it is better than paying sticker price. (In any attempt at a balanced view of CU –not the intent here one would of course have to mention their many investigations which have saved consumers millions, such as the work on health and life insurance).
- v. They wrongly support the numerical weight and sum method of synthesizing performance on various dimensions (“weight the dimensions for importance, standardize the performance scores and multiply them by the weights, add up the totals for each about candidate and the one with the most points is best’). This is another expensive error about procedures for product evaluation, whether done by consumers or consumers union, and it is the basis for the many rankings published by CU on the basis for the “overall score”. Those

rankings will be invalid, to an unknown but possibly large degree, as demonstrated in the article on numerical weight and sum in the Evaluation Thesaurus, 3rd edition.

- vi. There's failure to review the presuppositions of their technical advisors: for example, to look critically at the cost of lifelong restorative dentistry by comparison with one-time fitting of plates, a shift in treatment that has made hundreds of millions for dentists, and cost the consumers the same...
- vii. There's no sign that CU does serious comparative reviews of Consumer Reports against the competition, especially Consumers Digest in the U.S. and various overseas publications; these competitors appear clearly superior in some respects and could be adopted by CU (e.g., in their use of specialist evaluators). (Scriven, 1994, pp. 48-50)

In this writing, Scriven (1994) was able to call out both philosophical and methodological issues he observed in Consumers Union's approach to post purchase consumer product evaluation. He explicitly objected to their lack of solid subject background information, as well as their lack of transparency regarding the past and current evaluation context for deeper understanding. He also objected to the limited amount of comparison study. In general, basic market research practice requires benchmarking a product against all competitors for the purpose of a SWOT (i.e., strengths, weaknesses, opportunities, and threats) analysis. It was in Consumers Union's best interest to do a SWOT analysis with its competitors and publish it, but they did not. Moreover, Scriven (1994) was very concerned about Consumers Union's application of

the numerical weight and sum method and their sources of bias and error, most of which are still the same today.

The Joint Committee on Standards for Education Evaluation

The 1960s were marked with a need for not only social reform in America, but social programming evaluation. As described on encyclopedia.com,

During the 1960s President Lyndon B. Johnson's administration declared an ostensible "War on Poverty" with its great society programs: Head Start, the Job Corps, food stamps, Medicaid, funded education, job training, direct food assistance, and direct medical assistance. Although the poverty rate declined in the 1960s, more than 4 million new recipients signed up for welfare.

(Encyclopedia.com, 2015)

Education in particular was still at a point where it was still grappling with ways to systematically measure and determine if direct learning objectives were being met. During this period, the United States government was dramatically expanding social programs. A great deal of evaluation was being conducted to determine if that investment was yielding intended outcomes. By the late 1960s, demand for feedback about social programs exceeded the supply of appropriate skills (Shadish, Cook, & Leviton, 1991). Methods and standards varied based on who was conducting the evaluations. It became clear to the those closest to the industry that standards needed to be developed and adopted by the profession that were complete, accurate, and useable.

The year 1975 marked the beginning of discussions among three professional evaluation organizations, which organized to form the Joint Committee on Standards for Education Evaluation (JCSEE). In 1981, this organization was formalized when it applied

for and was granted 501(c)(3) corporate status, simultaneously publishing its first set of evaluation standards—the standards for evaluation of educational programs, projects, and materials. The JCSEE is now supported by 12 to 17 North American professional organizations, and a membership of nearly three million (Stufflebeam & Coryn, 2014).

The JCSEE standards have a two-pronged approach to evaluation. These standards provide a framework for an example of good evaluation. The standards were developed with stakeholder input, vetted, refined, and re-written. The standards are very unique in many ways, yet they can replicate specifications found on a part print or laboratory manual. As Stufflebeam (2004) noted,

The Joint Committee's initial mission was to bring diverse stakeholder groups together to get a purchase on the meaning of evaluation in the context of the failures of evaluation in the early years of the U.S. War on Poverty programs of the 1960s and 1970s.

The organization of the JCSEE standards is very precise and consists of five domains: (1) Utility, (2) Feasibility, (3) Propriety, (4) Accuracy, and (5) Accountability. Evaluation involves persistent fact-finding and problem solving. Often, evaluations are conducted using evaluation models in many different ways, as intended, hybrid, or even mixed and matched. The standards provide a compass to guide the evaluator through the process in a sound and proven methodological way. In the course of any application, the standards can be applied in varying ways, combinations, and purposes.

Minimally, the JCSEE standards are a checklist to determine the overall fitness of an evaluation. These standards can be used for program evaluation, personnel evaluations, evaluation design, evaluation reviewers, evaluation sponsors, instructors, and

teachers. These standards can be applied in alternative evaluation settings as well, determining if a program is evaluable, demonstrating what evaluation can and cannot do, to help create agreements, contracts, and memoranda of understanding, identifying and interacting with stakeholders, evaluation design, managing evaluation implementation, data collection, analysis, communication, evaluation completion and follow through, accountability, and the appropriate uses of resources. Again, as Stufflebeam (2004) noted:

The Standards also helped evaluators and clients see the great limitations of standardized tests and experimental design studies. Perhaps most important, the standards effort brought together a diverse group of teachers, counselors, statisticians, psychometricians, evaluators, policy makers, administrators, etc., to pursue a common purpose of strengthening evaluation theory, practice, and utilization.

To conclude, the JCSEE standards are *the* guide to proficient, accurate, and systematic evaluation. The reason these standards are so used and adhered to is that having been developed by a field of practitioners, they are easily applied to the design of any evaluation. Furthermore, the standards are comprehensive in nature. According to Stufflebeam (2014):

The Standards are widely referenced in evaluation textbooks and other evaluation literature. They are used extensively in evaluation courses. There is evidence of widespread use in practice, though more use is needed. The Standards have been applied in a wide range of metaevaluations (Stufflebeam, 2001); the Louisiana and Hawaii state education departments adopted the standards as evaluation

policies; and the U.S. Marine Corps adopted the Joint Committee (1988) *Personnel Evaluation Standards* as the criteria for assessing and strengthening the Corps' system for evaluating officers and enlisted personnel. Arguably, the Joint Committee Standards have contributed to evaluators' expanded perspectives on what constitutes sound evaluation—from the early practice of looking myopically at an evaluation's accuracy to the prevalent, broader view that evaluations must also meet conditions of utility, feasibility, and propriety.

Other Standards Organizations

There are a few additional standards organizations that are worth mentioning, some of which are designed for the purpose of standardized evaluation. These are explored briefly to ensure greater understanding of the standards domain.

ISO 9001 Standards

ISO 9000 essentially had its profound birth in 1987 and was deployed as three different standards, which were based on the scope of activities of an organization. ISO 9001 was based on design, development, production, installation, etc., and was targeted toward companies that developed new products. ISO 9002 is a model for production, design, and development of systems in a new product development environment. ISO 9003 is a model for quality assurance of final product with no regard to how the product was manufactured. The ISO 9001 standards are developed to optimize and standardize company systemization. “These standards govern systems, not products” (Williams, 2004).

ANSI Standards

The American National Standards Institute was created in 1918 for the purpose of helping and facilitating standards co-development for business and government (<http://webstore.ansi.org/documents/What-Is-ANSI.pdf>). ANSI is not a standards developing organization, but rather an organization that partners with industry and government to oversee the development and use of thousands of standards globally with over 125,000 companies both domestically and internationally. The JSCEE Program Evaluation Standards are a member of ANSI.

ANSI also provides a service for product accreditation. A supplier can take a product and its specification to ANSI for attestation of conformance. ANSI will conduct the tests themselves if possible, and ANSI will determine if the product meets identified specifications and provide its accreditation. Again, ANSI is not developing standards for the accreditations of a company's product performance, but they will ensure and provide accreditation to the fact that it does or does not (<http://webstore.ansi.org/documents/What-Is-ANSI.pdf>).

Consumers are constantly challenging the envelope of product performance. The global market place seems to drive new specifications and product requirements. Global certifications in a similar way have been asked to stay current with the changing needs of consumers worldwide. In the past product certification companies have been hard to work with and very slow to react, coupled with the fact that they may not have completely understood the needs of producers and designers, it has been a struggle to arrive at global harmony (Barron, 2007):

Manufacturers today are faced with often complex regulations that affect their products, and they have a wide range of certification marks to consider. Choosing the right mark for a product involves analyzing the applicable regulations and customer need of each market sought and applying the most efficient, cost-effective means of obtaining the desired marks without, needlessly, repeating steps in the process.

Both ISO 9001 and ANSI standards are not involved in any way in the subject of this study which primarily is post purchase consumer product evaluation. ANSI does come close with its ability to provide accreditation for products, but its objective is to use their organization's reputation to provide attestation to a product performance specifications based on criteria and standards the producer established and developed.

ASTM

Founded in 1898 by Charles Dudley, its primary focus and objective was to develop standards for industrial materials. These standards comprise a variety of different steels, metals, cement, plastics, wood, etc. The goals of the standards are to properly perform tests and evaluations on raw materials used primarily in industrial applications. Further, in many cases these standards will provide a base standard for evaluation results. These standards are used by engineers, builders, product designers all over the world as they provide a key reference to established standards for performing tests and their related results for raw material. These standards have nothing to do with product evaluation standards in any manner.

The United States Government Accountability Office Standards

The United States government is held to its own standards by the United States tax payers. As one might assume, with hundreds of government programs all funded by U.S. tax payers, a clear and robust set of accountability standards must be used to determine governments success or failure of resource use. “Government auditing provides objective analysis necessary to help create a better future” (GAO, 2011). According to the GAO (2011), the standards set a stage for government auditors to be an example to any auditor by; exemplifying independence, transparency, accountability, and quality through the audit process. These standards seek to accomplish the following:

- a. Provide a framework for conducting high quality audits with competence, integrity, objectivity, and independence.
- b. To be used by auditors of government entities and entities that receive government awards and audit organizations performing GAGAS (General Accepted Government Auditing Standards).
- c. The GAGAS are standards for audits and act as a guidance for dealing with ethics, independence, professional judgement, competence, quality control, and reporting. (GAO, 2011).

AEA Guiding Principles

The American Evaluation Association has created its own set of guiding principles for evaluators. Based on years of experience and cooperation among practitioners, it was necessary to outline a list of guiding principles that all evaluation practitioners should minimally subscribe to. These principles, some of which date back as far as 1982 from

earlier efforts to standardize evaluation techniques, were developed and vetted by a task force and voted in January 1994 (<http://www.eval.org/p/cm/ld/fid=105>).

A. Systematic Inquiry: Evaluators conduct systematic, data-based inquiries.

B. Competence: Evaluators provide competent performance to stakeholders.

C. Integrity/Honesty: Evaluators display honesty and integrity in their own behavior, and attempt to ensure the honesty and integrity of the entire evaluation process.

D. Respect for People: Evaluators respect the security, dignity and self-worth of respondents, program participants, clients, and other evaluation stakeholders.

E. Responsibilities for General and Public Welfare: Evaluators articulate and take into account the diversity of general and public interests and values that may be related to the evaluation. (<http://www.eval.org/p/cm/ld/fid=51>)

Summary

For the purpose of this study, it was necessary to consider how product evaluation got its roots in order to understand the overall importance of product evaluation. Based on what was revealed through this literature review, consumer safety and fraud drove early product comparisons, testing, and the development of new consumer product standards (Strach & Russell, 2003). In particular, consumer safety and fraud spurred the eventual creation of the National Commission on Product Safety, which then led to the Consumer Product Safety Council, as well as several private consumer-focused protection agencies such as Consumers Union and Good Housekeeping. This was both monumental and yet contentious in the world of product testing and evaluation. In simple terms, it meant people performing product evaluations had the support of government and

consumers to develop oversight for producers (Waller et al., 2008). It also meant accountability for producers (Williams, 1965).

By harnessing massive numbers of opinions about a product, companies have a way to get a bigger picture of what other consumers think about their products, and consumers get quicker access to other consumers' experiences and opinions about products. Much of this activity today is taking place online. Producers either use their own websites or other sources of exchange to optimize transfer of market information. This process of information exchange through the use of technology has become incredibly popular with consumers. It is fast, employs large numbers, and cheap to obtain. Product reviews in many consumers' minds is the new method of product evaluation. A survey conducted by BrightLocal Research (2014), for example, found that 88% of survey takers ($N = 2,104$) used online reviews when choosing a local business. This is up 2% over 2013. Similarly, 39% read reviews on a regular basis, up 7% from 2013 (<https://www.brightlocal.com/2014/07/01/local-consumer-review-survey-2014/>). This data continues to point toward a trend of continued and increasing use of consumer reviews for decision-making.

CHAPTER III

METHOD

This chapter describes the methods that were used to conduct this study. This study examined crowdsourced product evaluation at one crowdsourcing company, MTurk. MTurk was chosen for this study because it represented the largest market share of the self-serve crowdsourcing industry's growing revenues. In the study, MTurk workers' product evaluation standards were measured against the Joint Committee on Standards for Education Evaluation (JCSEE) program evaluation standards (Yarbrough et al., 2011). The specific purpose, research questions, research design, sample, instrumentation, data collection procedures, and data analysis used in the study are described in the sections below.

Purpose of the Study

Evaluation standards define evaluation quality and guide evaluators and consumers in the pursuit of quality evaluation (Yarbrough et al., 2011). The purpose of this non-experimental, survey research was to identify if crowdsourcing workers use any evaluation standards. This study was also designed to determine how closely crowdsourcing product evaluations adhere to the JCSEE program evaluation standards. Specifically, this study sought to determine: (a) what, if any, evaluation standards are being used by crowdsourcing organizations and their reviewers, and (b) to what extent these standards adhere to accepted evaluation standards used by professional evaluators in the evaluation discipline, precisely, the Utility, Feasibility, Propriety, and Accuracy domains of the JCSEE program evaluation standards.

Research Questions

Using data collected from a web-based survey of workers of the self-serve crowdsource company MTurk, the following research questions were explored:

1. Are evaluation standards used on crowdsource product evaluations? If not, why?
2. To what extent do crowdsource workers adhere to the Utility, Feasibility, Propriety, and Accuracy domains of the JCSEE program evaluation standards (Yarbrough et al., 2011)?

Research Design

A descriptive, cross-sectional survey design was utilized for this study. As stated above, the purpose of this study was to explore if crowdsource workers, in particular, Amazon MTurk product reviewers, use any evaluation standards, and if so, how closely these standards adhere to the JCSEE program evaluation standards. According to Hall (2008), “A cross-sectional survey collects data to make inferences about a population of interest (universe) at one point in time” (p. 172). Overall, cross-sectional survey research consists of four primary components: (a) conceptualization, (b) sample design, (c) questionnaire or data collection instrument, and (d) operations planning. These components as applied to the current study are discussed in the paragraphs below.

Conceptualization is a significant aspect of cross-sectional research. Not only does the conceptualization phase of cross-sectional research influence decisions made in later phases of the study, but a thorough conceptualization process also helps to reduce errors and mistakes (Hall, 2008). Conceptualization includes the following:

1. Defining the study population
2. Formatting hypotheses, if any, to be tested

3. Defining the outcome (dependent) variables of interest and important classification or dependent variables
4. Specifying levels of precision, such as standard errors, confidence intervals (“margin of error”), or statistical power
5. Deciding whether the survey will be repeated
6. Establishing cost limits
7. Specifying whether the nature of the data to be collected—cost or other considerations—requires a certain data collection mode (Hall, 2008, p. 172).

As mentioned, the population for this cross-sectional study consisted of MTurk product reviewers. Because the research questions of the study are non-experimental and descriptive in nature, it is inappropriate to formulate hypotheses, and issues concerning levels of precision and statistical power are irrelevant. While this study is non-experimental and will not be repeated, rich data is expected to be collected nevertheless because as Tripodi and Potocky-Tripodi (2007) noted, “cross-sectional survey is the design choice for yielding quantitative descriptions” (p. 49). The researcher paid financial costs, and limited incentives to participants were in the form of paid work (see Data Collection section) and Internet survey-hosting fees.

Gaining access to the crowdsource workers required use of an Internet-based mode of data collection. Internet survey methods yield major advantages over mail survey methods including (a) quicker speed of response, (b) lower costs for distribution of questionnaires, and (c) precision of data compilations (Matsuo et al., 2004). This speaks to the sample aspect of the research design. *Sample design* refers to the design of the parameters (also called the *sampling frame*) used to ensure adequate coverage of a

study's population (Hall, 2008). Details of the sample design for this study are discussed in the Non-Probability Sample section below, however, in general, aspects of sample design include:

1. Selecting a sampling frame
2. Defining the strata, if any, to be employed
3. Deciding whether the sample is to be a single-stage, clustered, or multi-stage design, and
4. Determining the sample size (Hall, 2008, p. 173).

Another aspect of research design is questionnaire design. *Questionnaire design* refers to considerations given to the development of the survey instrument that will be used to obtain data for the study (Hall, 2008). It is critical that items on the questionnaire allow for the collection of data that would answer the study's research questions. Accordingly, the items on the questionnaire in this study consist of statements from the Utility, Feasibility, Propriety, and Accuracy domains of the JCSEE program evaluation standards. This allowed the researcher to the frequency participants in this study apply each of the specific JCSEE standards during product reviews. Further, if never or seldom were chosen by participants, open-ended questions would follow. Questionnaire design is discussed further in the Instrumentation section below.

Finally, *operations planning* refers to detailing the logistics of a cross-sectional study, such as making sure institutional procedures are clearly identified and followed, and scheduling data collection (Hall, 2008). In this study, operations planning included; meeting HSIRB requirements (Appendices A and B), obtaining survey-hosting services,

uploading the survey, contracting with MTurk, allowing enough time for data collection, and securing the proper data analysis software.

In sum, use of the cross-sectional research design described above allowed the researcher to quickly and efficiently capture information from a large sample of product reviewers that increased the accuracy and legitimacy of the data extracted (Kraut et al., 2004; Reips, 2002). However, as in all research, there are limitations inherent in cross-sectional survey research designs. Problems related to social desirability, or a phenomenon wherein participants respond to questions in a way they believe others will approve, have long been noted in the use of survey research (Johnson & Fendrich, 2002). The researcher made efforts to minimize problems with social desirability by making the survey anonymous and ensuring that no identifying information would be associated with responses. Beyond social desirability, the generalizability of cross-sectional research can be limited by sources of error in sampling, measurement, and non-response (Groves, 1989). These limitations were minimized by the controls and filtering placed on the subjects being surveyed. Specifically, use of MTurk.com allowed the researcher to filter unwanted subjects by establishing certain criteria potential subjects had to meet in order to participate in the study. Additionally, both sample and non-response errors were further minimized, as MTurk ensured an exact number of participants are recruited and sampled.

Non-probability Sample

A non-probability sampling method was used to gather information pertaining to the study's focal research questions. Battaglia (2008) noted that there are several types of non-probability sampling. For the purposes of this study, total population purposive

sampling was utilized (<http://dissertation.laerd.com/purposive-sampling.php#types>). Use of purposive sampling involves identification of a specific population, and then use of the researcher's judgment to identify characteristics necessary to be representative of the population and the sample size. The total number of the worker population that fits the inclusion criteria of this study is unknown. According to Amazon,

Mechanical Turk has more than 500,000 Workers from 190 countries, but we unfortunately do not retain or provide demographic information for our Workers. Unfortunately, this means that we are unable to provide information on how many Workers are located within the United States. (MTurk, 2015)

Even when the parameters of the total population of a study are unknown, "The selection of a purposive sample is often accomplished by applying expert knowledge of the population to select in a non-random manner a sample of elements that represents a cross-section of the population" (Battaglia, 2008, p. 524). Based on the goals of this study and the research questions being asked, the following criteria were utilized. To be included in the study, reviewers had to have more than one year of experience and be classified as a trained and qualified reviewer for product evaluations living in the United States, with at least a 90% approval rating. Moreover, participants must be from the United States to assure adequate English language skills. Exclusionary criteria included reviewers located outside the United States and those that had less than one year of product evaluation experience.

There are no guidelines for determining sample size in a descriptive study. Overall, the researcher intended to recruit a total of 500 reviewers. Because a non-probability sampling method was used, identification of strata and consideration of a

single-stage, clustered, or multi-stage design as suggested in the sampling procedure identified by Hall (2008) was not applicable.

Instrumentation

Instrumentation for this dissertation included a 27-item survey developed by the researcher for the purpose of this study and an 11-item demographic questionnaire (Appendix C). The items on the survey represented each of the four domains of the JCSEE program evaluation standards (Yarbrough et al., 2011) that were explored in this study (i.e., Utility, Feasibility, Propriety, and Accuracy), and assessed how often reviewers considered each program evaluation standard when conducting product evaluations or reviews within the last 12 months. They were rated on the following 4-point Likert-type scale: 1 = *Never*, 2 = *Seldom*, 3 = *Often*, 4 = *Always*. When selecting *Never*, *Seldom*, or *Always* on select items, participants were prompted to respond to an open-ended question that allowed participants to explain their answers. Higher scores indicated greater consideration of the standards when conducting product reviews. There were no cut-off scores indicating adequate or inadequate consideration of the standards.

Items of the demographic questionnaire were designed to explore how many years participants have worked in the crowdsource business, their role in crowdsourcing, the segment of product review for which they most frequently get contracts, if they have formal training in product review, the type of training they have, their highest level of education, their primary area of expertise, and if they have experience working with evaluation checklists. The survey ended with an open-ended response question that provided an opportunity for participants to make additional comments. High quality

responses were expected based on crowdsourcing worker incentives to succeed and advance within their organization as an evaluator or reviewer.

Data Collection Procedures

As indicated, this research was conducted using an online administration of the survey and demographic questionnaire with MTurk workers. The primary reason for using MTurk was it represented a very large population of product reviewers. MTurk was also both accessible and cost effective for research-based crowdsourcing. To use other product review crowdsourcing portals such as crowdsourcing.org was very expensive and for purposes of research, not very accessible. The sub-sections below describe how the survey was set up, recruitment of reviewers, and the informed consent process.

Setting up the Crowdsourcing Project

As indicated, this dissertation research was conducted through the single, privately held self-service crowdsourcing company, MTurk.com. Initiating services with MTurk required filling out forms found on the company's website (Appendices D through F). These forms captured the requester's (i.e., the researcher's) contact and billing information, as well as a description of what was being requested. After completing the initial forms, MTurk then asked for a hosting website (in this study, Qualtrics) where the survey was to be conducted, and allowed the requester to upload any special announcements or instructions prior to the reviewer being transferred to the hosting website for completion of the survey instrument.

Each time a requester sets up a project on MTurk, they must agree to MTurk Terms and Conditions. After gathering all background and billing data, MTurk then provides an invoice for services. The researcher paid this invoice prior to the project

launch using a credit card. Once paid, MTurk informed the researcher of when reviewers would be directed to the study's host website. Each survey participant was assigned a unique survey URL. Each survey was available for completion by each participant for one week.

Recruitment of Reviewers

Recruitment of reviewers occurred on MTurk through a job board that identified work opportunities (Appendix G). No pre-surveys or survey notices were necessary. However, in place of these activities was a requisition for the purpose of contracting the use of workers to complete surveys and execute work orders. If a reviewer was interested in completing the survey for this dissertation, he or she had to agree to contract with MTurk by clicking on the available opportunity. The reviewers then had to complete the required work (i.e., complete the survey in its entirety) for a fee paid through MTurk by the researcher as indicated above.

Informed Consent Process

Participation in this study was part of a voluntary crowdsource work opportunity. Nevertheless, all study participants were asked to indicate their consent in the study (Appendix H). MTurk workers had to be over the age of 18, so they were free to provide their own consent. Because this study was an online Qualtrics survey, consent was obtained using a consent screen. As Schmidt (1997) suggested, there was a separate consent screen before respondents gain access to the survey. On the consent screen, potential participants read the informed consent letter and affirmed their consent to participate in the study by clicking a button that stated, "I agree to participate." As part of the informed consent process, participants were also informed that they may abort the

survey at any time or refuse to answer all or certain questions; however, payment for participation in the study required completion of all survey questions.

Data Analysis

Analysis of all survey data was conducted according to generally accepted scientific standards. Closed-response items from the surveys were analyzed in SPSS using descriptive statistics including measures of central tendency such as means, medians, and modes, as well as frequencies and percentages. Open response items were analyzed using thematic methods by the researcher. All findings are presented using tables and figures where appropriate. Findings and conclusions for the research questions are based on examination and interpretation of the information arising from surveys.

Secondary Analysis: Thematic Synthesis

As mentioned, select items embedded and elicited open-ended responses from participants. These responses were thematically analyzed, coded, and themed accordingly. Braun and Clarke (2006) argued that thematic analysis provides both an accessible and flexible approach to qualitative data. Further, the literature has shown that thematic analysis is a growing and recognized method within the scientific community, and has become heavily relied upon as an addition to other methods of analysis (Boyatzis, 1998).

The method used for thematic analysis for this dissertation required a six-stage process that is widely recognized within the psychological and scientific communities (Braun & Clarke, 2006). Stage 1 consisted of data familiarization. Stage 2 consisted of generating initial codes. Stage 3 consisted of developing themes. Stage 4 consisted of

reviewing themes. Stage 5 consisted of defining and naming themes. Finally, Stage 6 consisted of producing the report.

All coding and theme development was considered inductive and was conducted by hand. Because themes were strongly linked to data gathered in the survey itself, it was appropriate to use an inductive approach to coding. Patton (1990) posited that when strong linkages exist between themes and data, it is not necessary fit it to a pre-existing theme or coding rationale, inductive coding works best. Coding consisted of reviewing each open-ended data response line by line, and then assigning a recognizable code word or phrase to each line item. Braun and Clarke (2006) argued the art and science of coding is not and does not need to be exact or precise; yet it does need to be consistent. Results of the line-by-line coding and themes can be seen in Appendices I and J.

Table 3 summarizes the total research methodology.

Table 3

Summary of Research Methodology

Method	Non-experimental, cross-sectional survey research
Participants	Trained and qualified product reviewers at MTurk with more than one year of experience and a 90% approval rating from the United States.
Sampling Procedures	Probability sampling of entire MTurk population meeting inclusionary criteria.
Instrumentation	Web-based survey assessing how often reviewers considered each program evaluation standard when conducting product evaluations or reviews within the last 12 months.
Data Analysis	Descriptive statistics including means, medians, modes, frequencies, and percentages.
Secondary Analysis	Six-stage thematic analysis and synthesis

CHAPTER IV

RESULTS

Chapter IV presents the results obtained from the study described in Chapter III. As stated in Chapter III, this study utilized a descriptive, cross-sectional survey design. Specifically, data was collected using a non-probability sampling method that included an online administration of a 27-item survey and 11-item demographic questionnaire developed by the researcher for the purpose of this study. Participants were workers at MTurk, the crowdsourcing division of Amazon.com. The section below reports the participant information obtained from the demographic questionnaire. The subsequent sections present the results of the data analyses conducted to answer the study's two research questions.

Participant Demographics

As stated, participant demographic information was collected using an 11-item questionnaire. The items of this demographic questionnaire were designed to explore: (a) how many years participants have worked in the crowdsourcing business, (b) their role in crowdsourcing, (c) the segment of product review for which they most frequently get contracts, (d) if they have formal training in product review, (e) the type of training they have, (f) their highest level of education, (g) their primary area of expertise, and (h) if they have experience working with evaluation checklists. A total of 653 workers initiated the survey; however, 1 person did not consent to participate in the study; 7 people made no indication of consent; 1 person gave consent and started the survey, but did not complete it; and 63 people gave consent to participate in the survey, but did not answer any questions. Further, 124 persons gave consent and answered questions, but did not

indicate if they reviewed products part or full time, and were filtered out of the final data set to assure the respondents were product reviewers either part or full time. Therefore, the final sample frame for this study was 454. Tables 2 through 4 present the frequencies and percentages from the demographic survey of these remaining participants. Missing data was considered MCAR (missing completely at random), and therefore not addressed further by the author.

Participant Crowdsourcing Evaluation Experience

Table 4 presents descriptive statistics for the demographic variables. As shown in Table 4, the majority of participants ($n = 385$, 84.80%) reported working 1 to 5 years in crowdsourcing evaluation. The second largest group of participants ($n = 60$, 13.22%) reported working 6 to 10 years in crowdsourcing evaluation, followed by participants who worked 11 to 15 years ($n = 4$, .88%), 16 to 20 years ($n = 1$, .22%), and 21 or more years ($n = 1$, .22%). The largest proportion of participants ($n = 428$, 94.27%), indicated they had reviewed products part time, while the remaining ($n = 26$, 5.73%) reviewed products full time. The majority of participants ($n = 425$, 93.61%) identified their role in crowdsourcing as reviewer. Among those who answered other ($n = 5$, 1.17%) three participants said they reviewed products they purchased on their own through vendors such as Amazon.com, while one participant indicated writing articles and other things. Finally, when asked what product segment do you get contracts for most often, the largest group of participants indicated they received contracts for medical products ($n = 164$, 36.12%), followed by consumer-packaged goods ($n = 154$, 33.92%). Among participants who answered other ($n = 18$, 3.96%), three participants indicated receiving contracts for

health and beauty, two participants for the service industry, and two participants for websites.

Table 4

Participant Crowdsourcing Evaluation Experience

Variable	Category	<i>n</i>	%
How many years have you worked in crowdsourced evaluation?	1-5 years	385	84.80
	6-10 years	60	13.22
	11-15 years	4	.88
	16-20 years	1	.22
	21+ years	1	.22
	Missing	3	.66
	Total	454	100.00
Do you perform product reviews?	Part Time	428	94.27
	Full Time	26	5.73
	Total	454	100.00
What in your role in crowdsourcing? Select all that apply.	Coordinator	20	4.41
	Manager	21	4.63
	Director	6	1.32
	Reviewer	425	93.61
	Other	6	1.32
What product segment do you get contracts for most often?	Consumer-packaged goods	154	33.92
	Medical	164	36.12
	Electronics	12	2.64
	Automotive	6	1.32
	Food & beverage	99	21.81
	Other	18	3.96
	Missing	1	.22
Total	454	100.00	

Participant Range of Pay Per Review

Lastly, the final section on the demographic questionnaire asked participants to indicate the approximate range of pay they receive in U.S. dollars for each review. As shown, the largest group of participants reported receiving \$1.00 to \$1.50 per review. Among participants who responded other ($n = 68$, 15.89%), 5 reported receiving \$3.00, 5 reported receiving \$5.00, and 3 reported receiving \$10.00. The remaining ($n = 55$, 80.82%) varied in response.

Table 5

Participant Range of Pay Per Review

Variable	Category	<i>f</i>	%
Approximately what is the range of pay in USD per review?	.25-.50	77	16.96
	.51-.75	74	16.30
	.76-1.00	94	20.70
	1.00-1.50	172	37.89
	Other	68	14.98

Pay ranges per review varied widely for crowdsourced reviewers. The largest segment of paid product reviewers supporting this survey were paid \$1.00-1.50 ($n = 172$, 37.89%). The next largest segment of paid product reviewers were \$.76-1.00 ($n = 94$, 20.70%). Finally, the third ranked pay category for product reviewers in this survey was \$.25-.50 ($n = 77$, 16.96 %). The category of “other” ($n = 68$, 14.98%) had many different, non-recurring monetary values that peaked at \$100.00 per review.

Crowdsource Workers' Use of the Utility, Feasibility, Propriety, and Accuracy Domains of the JCSEE Program Evaluation Standards

Research question 2 is: To what extent do crowdsource workers adhere to the Utility, Feasibility, Propriety, and Accuracy domains of the JCSEE program evaluation standards (Yarbrough et al., 2011). To answer this question, survey items were developed to assess how reviewers consider these domains when conducting product evaluations. Each item was rated on the following 4-point Likert-type scale: 1 = *Never*, 2 = *Seldom*, 3 = *Often*, 4 = *Always*. Table 6 provides a summary of the frequencies and percentages in the Utility, Feasibility, Propriety, and Accuracy domains. Following Table 6, Figures 3 through 7 present the frequencies and percentages for the items in each of the four domains, as well as a presentation of qualitative data from open-ended responses in each domain.

Table 6

Frequencies and Percentages of Crowdsource Workers' Use of the Utility, Domain of the JCSEE Program Evaluation Standards

Domain	Never	Seldom	Often	Always	Missing	Total
Utility						
I completed reviews of products without prior experience using them.	18 (3.96%)	43 (9.47%)	80 (17.62%)	313 (68.94%)	0 (0.00%)	454 (100.00%)
I completed reviews with an understanding of whom I was completing them for.	30 (6.61%)	41 (9.03%)	229 (50.44%)	153 (33.70%)	1 (.22%)	454 (100.00%)
I had a clear understanding of the product review's purpose.	3 (.66%)	10 (2.20%)	224 (49.34%)	215 (47.36%)	2 (.44%)	454 (100.00%)
I had complete understanding of the requestor's values underpinning the review.	10 (2.20%)	67 (14.76%)	266 (58.59%)	111 (24.45%)	0 (0.00%)	454 (100.00%)

Table 6—Continued

Domain	Never	Seldom	Often	Always	Missing	Total
Utility						
I was informed of the needs of users of the reviews.	14 (3.08%)	43 (9.47%)	298 (65.64%)	99 (21.81%)	0 (0.00%)	454 (100.00%)
My reviews encouraged use.	5 (1.10%)	25 (5.51%)	363 (79.96%)	61 (13.44%)	0 (0.00%)	454 (100.00%)
My review submissions were on time.	3 (.66%)	18 (3.96%)	140 (30.84%)	291 (64.10%)	2 (.44%)	454 (100.00%)
My reviews were misused.	1 (.22%)	12 (2.64%)	153 (33.70%)	288 (63.44%)	0 (0.00%)	454 (100.00%)

Utility

Subjects answered items in each of the four domains using the following 4-point Likert type scale: 1 = *Never*, 2 = *Seldom*, 3 = *Often*, 4 = *Always*. As shown in Table 7, “I completed reviews of products without prior experience using them” resulted in the following responses: Always ($n = 313$, 68.94%), Often ($n = 80$, 17.62%), Seldom ($n = 43$, 9.47%), and Never ($n = 18$, 3.96%). “I completed my reviews with an understanding of whom I was completing them for” produced the following results: Always ($n = 153$, 33.70%), Often ($n = 229$, 50.44%), Seldom ($n = 41$; 9.03%), and Never ($n = 30$, 6.61%). The item “I had a clear understanding of the product review’s purpose” yielded the following results: Always ($n = 215$, 47.36%), Often ($n = 224$, 49.34%), Seldom ($n = 10$, 2.20%), and Never ($n = 3$, .66%). The item “I had a complete understanding of the requestor’s values underpinning the review” produced the following results: Always ($n = 111$, 24.45%), Often ($n = 266$, 58.59%), Seldom ($n = 67$, 14.76%), and Never ($n = 10$, 2.20%). The item “I was informed of the needs of users of the reviews” had the following results: Always ($n = 99$, 21.81%), Often ($n = 298$, 65.64%), Seldom ($n = 43$, 9.47%), and Never ($n = 14$, 3.08%). The item “My reviews encouraged use” had the following results:

Always ($n = 61$, 13.44%), Often ($n = 363$, 79.96%), Seldom ($n = 25$, 5.51%), and Never ($n = 5$, 1.10%). The item “My review submissions were on time” returned the following results: Always ($n = 291$, 64.10%), Often ($n = 140$, 30.84%), Seldom ($n = 18$, 3.96%), and Never ($n = 3$, .66%). Finally, the item “My items were misused,” returned the following results: Always ($n = 288$, 63.44%), Often ($n = 153$, 33.70%), Seldom ($n = 12$, 2.64%), and Never ($n = 1$, .22%). Figure 3 summarizes the results presented in Table 6.

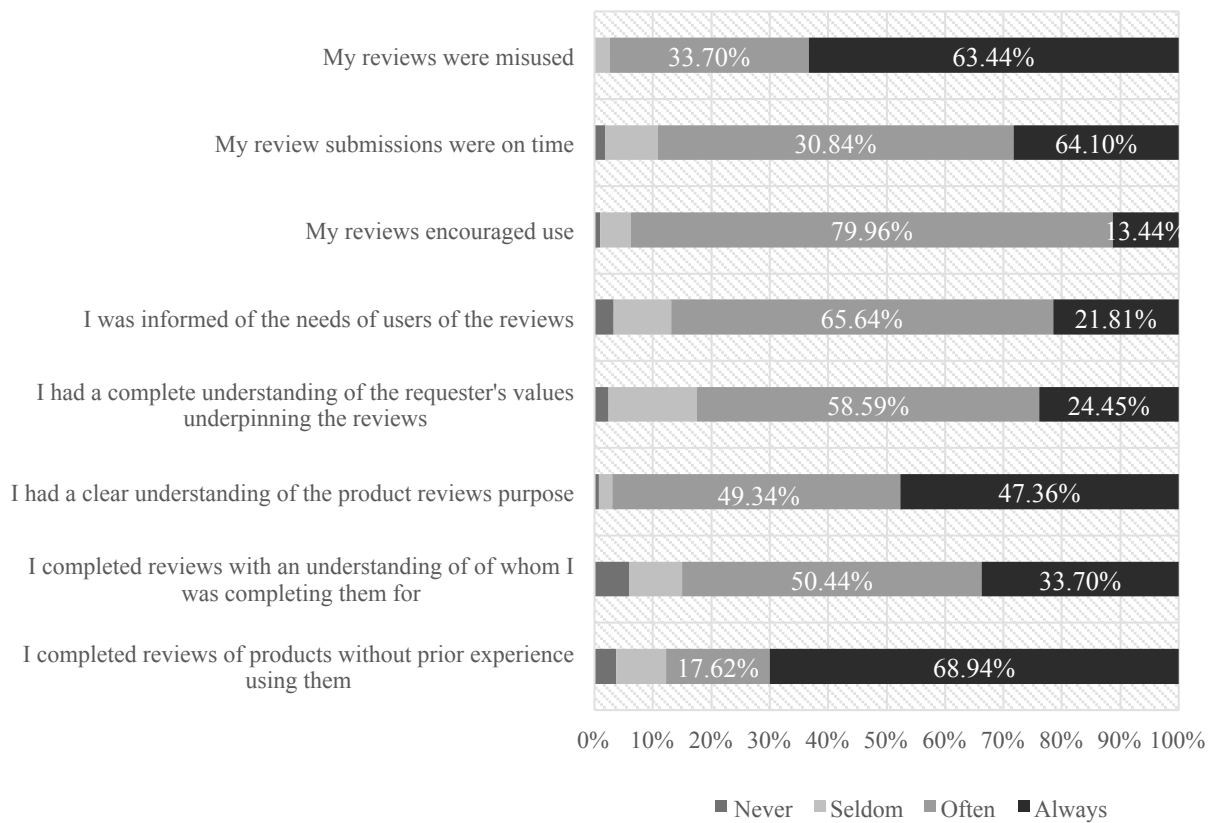


Figure 3. Percentages of crowdsource workers’ use of the Utility domain of the JCSEE program evaluation standards.

Follow-up questions were used in the Utility domain to gather qualitative data to inform this body of research even further. When subjects completing the survey answered questions linked to chosen options, a text box opened up in the survey asking them to

answer additional questions in more detail.

A thematic analysis was conducted on the secondary and follow-up responses to primary items. Inductive analysis was performed across all lines of open-ended responses and located recognizable patterns. Trends were recognized accordingly based on repetitive keywords and phrases, and were then coded for proper identification. The thematic analysis process used was a six-stage process as outlined in Braun and Clarke (2006), which presented clear and useful steps to thematic analysis and synthesis (pp. 16-23).

Inductive analysis was chosen for this study because there was a very strong link between open-ended responses and data received from subjects that participated in the survey. As mentioned earlier, inductive analysis is strongest when there are significant linkages between text and phrases and actual survey data. More detail as to the actual six-stage process taken to arrive at the theme coding and theme naming can be seen in Appendix I. Mapping of the thematic synthesis can be seen in Appendix J.

The following are qualitative responses relative to the Utility domain. If item “I had a complete understanding of the requestor’s values underpinning the review” obtained results Seldom or Never, a pop-up box in the survey opened asking the subject to respond to the additional question, “Can you briefly explain your strategy for your review if you did not know the requestors values?” Text segments, phrases, and keywords were thematically coded. The following themes were used to identify subject patterns: Values Based on Past Experiences (this theme was derived from texts and phrases that centered around personal experience); Objective Values (text and phrases that were central to an objective non-personal approach; and Random/Unrelated (random

words and orientations of sentence fragments).

There were 70 responses to this follow-up question. The majority ($n = 46$) of responses came from the theme Values Based on Past Experience. The theme Objective Values had 24 responses, and the theme Random/Unrelated, had seven responses that had no real meaning or purpose.

The gravity of these open-ended responses is large. The responses indicate a strong reliance on personal experience as the primary standard for their product review judgments. Therefore, participants appear to be projecting an experiential opinion of the product. Ott et al. (2012) argued that the more experience and knowledge a person has with a product, the more likely their opinion of it will change over time. Therefore reliability of experienced based product reviews is likely to be less than that of a more systematic and scientifically based product evaluation.

The following are some of the actual open-ended responses used to develop the above themes.

“I reviewed according to my experience.”

“I review something and give an honest opinion and me experiences.”

“I based it on what I know.”

“I made the review based on my experience with the product.”

In the Utility domain, the item “I was informed of the needs of users of the reviews” also triggered a follow-up question: “Can you briefly describe your review process if you were not informed of the needs of the reviews you were performing?” if respondents answered Seldom or Never. There were 58 responses total to this follow up question. Responses were again thematically coded into three groups. The following

themes emerged: Experienced-Based Needs (which again had a central orientation about the subjects' use of personal experiences), Objective Needs (this theme, as seen earlier, also reflected subjects' responses that were more based on an objective process of evaluation other than personal experience), and Random/Unrelated (random words and orientations of sentence fragments).

The results for "Can you briefly describe your review process if you were not informed of the needs of the reviews you were performing?" were as follows: Experienced-Based Needs had 33 responses, Objective Needs had 18 responses, Random/Unrelated had seven responses with various text and phrases that were random and could not link to a theme or data in a consistent manner or pattern.

Again, it is clear the majority of the open-ended responses were based on the subjects' personal experiences. Below I have listed some of those responses.

"I usually think of why I was looking to purchase the product and then give my take on if those needs were met or not."

"Review to the best of my ability and be honest."

"I only give reviews based on my experience."

"I would look up other reviews."

"It was either a product I had already tried out beforehand, or the product was included for me to test out."

"I detail my experience with the product to help others understand all of the aspects of the product that I know."

"I'd just do my best and give an honest opinion".

These examples of open-ended responses for the item "Can you briefly describe

your review process if you were not informed of the needs of the reviews you were performing?” are very transparent and identify use of experience-based opinions for product reviews among subjects that participated in the survey used for this dissertation.

Finally, the third follow-up question in the Utility domain was triggered from the item “My reviews were misused.” The triggering responses were Often and Always, which initiated the following question: “Please explain how your reviews have always been misused.” This item had two thematic groups: Were Not Misused (which was derived clearly from responses that indicated their reviews were not misused) and Unrelated Altogether (which was one random phrase). Were Not Misused had four responses, and Unrelated Altogether had one response.

Feasibility

Table 7 presents results from the Feasibility domain, and Figure 4 summarizes these results. Survey results showed that item “I used project management techniques to complete reviews” received the following subject responses: Always ($n = 20$, 4.41%), Often ($n = 167$, 36.78%), Seldom ($n = 130$, 28.63%), and Never ($n = 136$, 29.96%). Further, respondents endorsed item “I used practical procedures” as follows: Always ($n = 179$, 39.43%), Often ($n = 241$, 53.08%), Seldom ($n = 30$, 6.61%), and Never ($n = 4$; .88%). Item “My reviews recognized and balanced cultural and political interests of individuals and consumers” received the following results: Always ($n = 61$, 13.44%), Often ($n = 253$, 55.73%), Seldom ($n = 106$, 23.35%), and Never ($n = 33$, 7.27%). Finally, item “I used resources effectively and efficiently” provided the following results: Always ($n = 224$, 49.34%), Often ($n = 215$, 47.36%), Seldom ($n = 13$, 2.86%), and Never ($n = 1$, .22%).

Table 7

Frequencies and Percentages of Crowdsorce Workers' Use of the Feasibility Domain of the JCSEE Program Evaluation Standards

Domain	Never	Seldom	Often	Always	Missing	Total
Feasibility						
I used project management techniques to complete reviews.	136 (29.96%)	130 (28.63%)	167 (36.78%)	20 (4.41%)	1 (.22%)	454 (100.00%)
I used practical procedures.	4 (.88%)	30 (6.61%)	241 (53.08%)	179 (39.43%)	0 (0.00%)	454 (100.00%)
My reviews recognized and balanced cultural and political interests of individuals and consumers.	33 (7.27%)	106 (23.35%)	253 (55.73%)	61 (13.44%)	1 (.22%)	454 (100.00%)
I used resources effectively and efficiently.	1 (.22%)	13 (2.86%)	215 (47.36%)	224 (49.34%)	1 (.22%)	454 (100.00%)

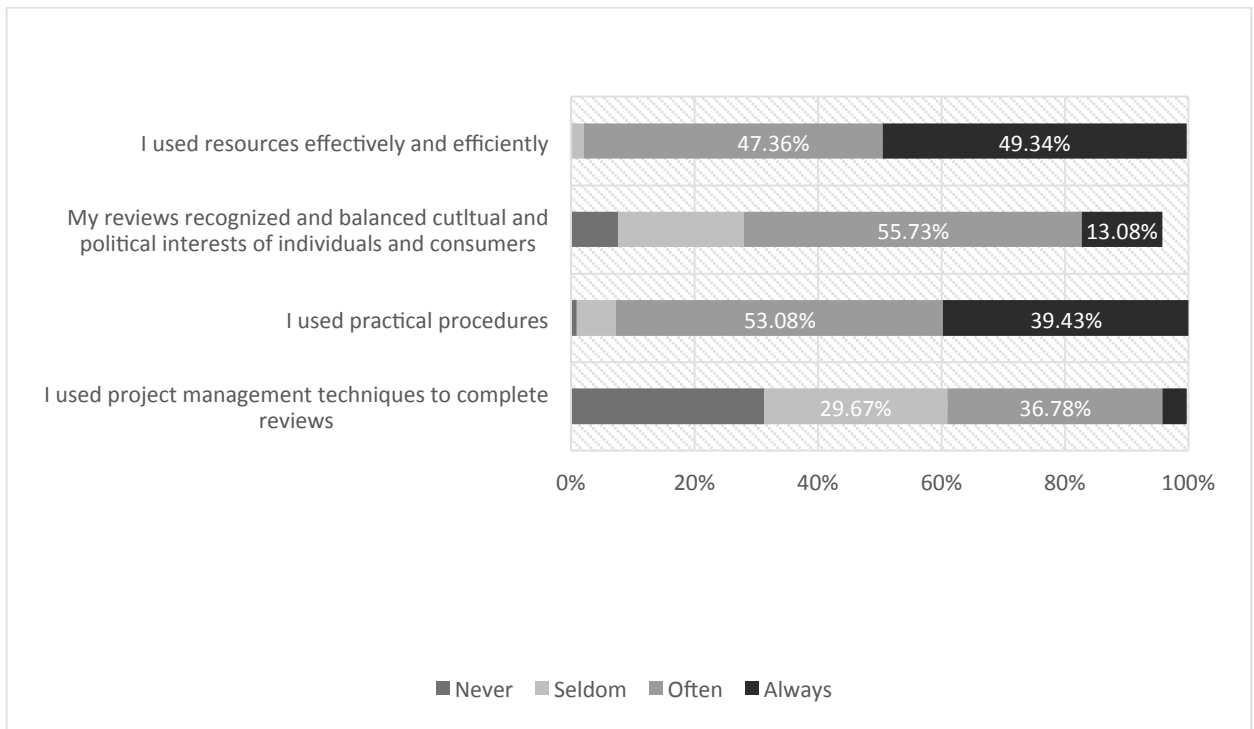


Figure 4. Percentages of crowdsorce workers' use of the Feasibility domain of the JCSEE Program Evaluation Standards.

Table 8 represents measurements of items from the Propriety domain of the survey instrument used in this study. Figure 5 summarizes these results. The item “My reviews were responsive to potential customers,” returned the following results: Always ($n = 190$, 41.85%), Often ($n = 243$, 53.52%), Seldom ($n = 19$, 4.19%), and Never ($n = 2$, .44%). Item “I was involved in negotiations with requesters about the reviews” provided the following results from the survey: Always ($n = 18$, 3.96%), Often ($n = 65$, 14.32%), Seldom ($n = 160$, 35.24%), and Never ($n = 211$, 46.48%). Item “My reviews considered consumer rights and safety” also provided the following feedback: Always ($n = 160$, 35.24%), Often ($n = 213$, 46.92%), Seldom ($n = 70$, 15.42%), and Never ($n = 9$; 1.98%). Item “My reviews are clear and fair,” returned the following results from the survey: Always ($n = 325$, 71.59%), Often ($n = 119$, 26.21%), Seldom ($n = 6$, 1.32%), and Never ($n = 2$, .44%). Item “My reviews disclosed all relevant information,” returned the following results: Always ($n = 246$, 54.19%), Often ($n = 194$, 42.73%), Seldom ($n = 10$, 2.20%), and Never ($n = 2$, .44%). Item “My reviews disclosed conflicts of interest,” returned the following results: Always ($n = 137$, 32.01%), Often ($n = 116$, 27.10%), Seldom ($n = 103$, 24.07%), and Never ($n = 75$, 16.59%). Finally, item “My reviews expended out of pocket resources,” provided the following results: Always ($n = 168$, 37.00%), Often ($n = 160$, 35.24%), Seldom ($n = 107$, 23.57%), and Never ($n = 18$, 3.96%).

Table 8

Frequencies and Percentages of Crowdsorce Workers' Use of the Propriety Domain of the JCSEE Program Evaluation Standards

Domain	Never	Seldom	Often	Always	Missing	Total
Propriety						
My reviews were responsive to potential consumers.	2 (.44%)	19 (4.19%)	243 (53.52%)	190 (41.85%)	0 (0.00%)	454 (100.00%)
I was involved in negotiations with requestors about the reviews.	211 (46.48%)	160 (35.24%)	65 (14.32%)	18 (3.96%)	0 (0.00%)	454 (100.00%)
My reviews considered consumer rights and safety.	9 (1.98%)	70 (15.42%)	213 (46.92%)	160 (35.24%)	2 (.44%)	454 (100.00%)
My reviews were clear and fair.	2 (.44%)	6 (1.32%)	119 (26.21%)	325 (71.59%)	2 (.44%)	454 (100.00%)
My reviews disclosed all relevant information.	2 (.44%)	10 (2.20%)	194 (42.73%)	246 (54.19%)	2 (.44%)	454 (100.00%)
My reviews disclosed conflicts of interest.	75 (16.59%)	103 (24.07%)	116 (27.10%)	137 (32.01%)	1 (.22%)	454 (100.00%)
My reviews expended out of pocket resources.	18 (3.96%)	107 (23.57%)	160 (35.24%)	168 (37.00%)	1 (.22%)	454 (100.00%)

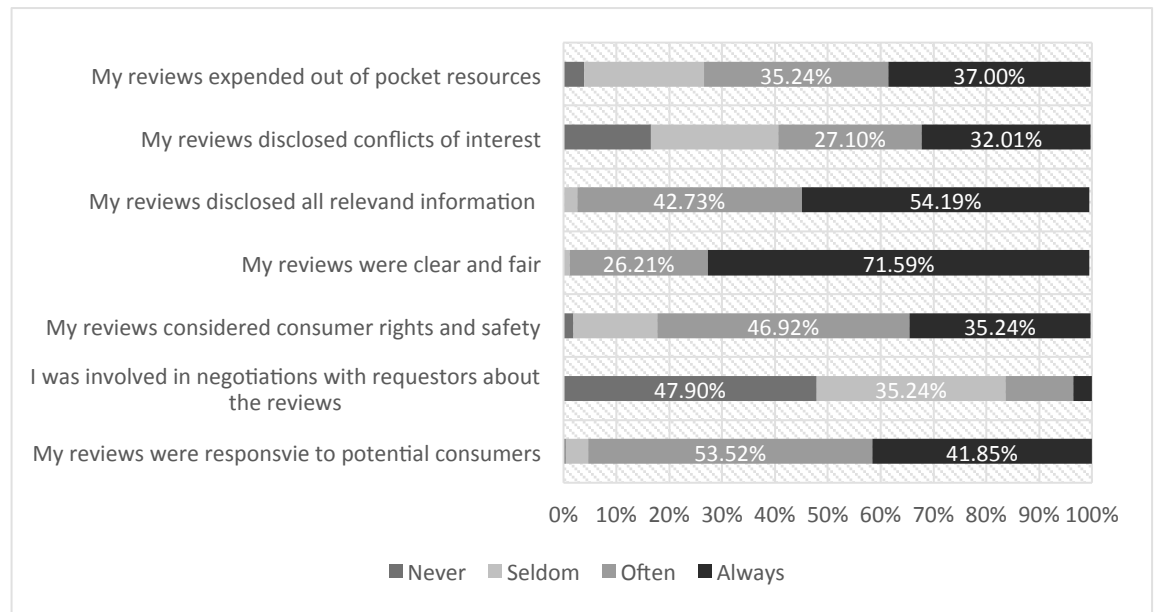


Figure 5. Percentages of crowdsorce workers' use of the Propriety domain of the JCSEE program evaluation standards.

Table 9 represents measurements from the Accuracy domain of the survey instrument used in this study. Figure 6 summarizes these results. Item “My review conclusions are or were justified” returned the following results: Always ($n = 299$, 65.86%), Often ($n = 142$, 31.28%), Seldom ($n = 9$, 1.98%), and Never ($n = 4$, .88%). Item “My review conclusions were valid,” returned the following survey results: Always ($n = 317$, 69.82%), Often ($n = 129$, 28.41%), Seldom ($n = 5$, 1.10%), and Never ($n = 3$, .66%). Item “My reviews were reliable,” returned the following survey results: Always ($n = 326$, 71.81%), Often ($n = 121$, 26.65%), Seldom ($n = 4$, .88%), and Never ($n = 3$, .66%). Item “I thoroughly documented reviews with great detail,” returned the following survey results: Always ($n = 131$, 28.85%), Often ($n = 268$, 59.03%), Seldom ($n = 47$, 10.35%), and Never ($n = 8$, 1.76%). Item “I used a system of information management” revealed these results: Always ($n = 45$, 9.91%), Often ($n = 202$, 44.49%), Seldom ($n = 118$, 25.99%), and Never ($n = 89$, 19.60%). The item “I used more technical review designs, when appropriate, for a specific review” revealed the following: Always ($n = 25$, 5.51%), Often ($n = 234$, 51.54%), Seldom ($n = 131$, 28.85%), and Never ($n = 64$, 14.10%). Item “I concisely documented review reasoning, findings, and conclusions” returned the following survey results: Always ($n = 129$, 28.41%), Often ($n = 260$, 57.27%), Seldom ($n = 53$, 11.67%), and Never ($n = 11$, 2.42%). Finally, the item “My review techniques guard against misinterpretations, biases, distortions, and errors” survey results returned the following results: Always ($n = 145$, 31.94%), Often ($n = 254$, 55.95%), Seldom ($n = 40$, 8.81%), and Never ($n = 15$, 3.30%).

Table 9

Frequencies and Percentages of Crowdsourcing Workers' Use of Accuracy Domain of the JCSEE Program Evaluation Standards

Domain	Never	Seldom	Often	Always	Missing	Total
Accuracy						
My review conclusions are or were justified.	4 (.88%)	9 (1.98%)	142 (31.28%)	299 (65.86%)	0 (0.00%)	454 (100.00%)
My review conclusions were valid.	3 (.66%)	5 (1.10%)	129 (28.41%)	317 (69.82%)	0 (0.00%)	454 (100.00%)
My reviews were reliable.	3 (.66%)	4 (.88%)	121 (26.65%)	326 (71.81%)	0 (0.00%)	454 (100.00%)
I thoroughly documented reviews with great detail.	8 (1.76%)	47 (10.35%)	268 (59.03%)	131 (28.85%)	0 (0.00%)	454 (100.00%)
I used a system of information management.	89 (19.60%)	118 (25.99%)	202 (44.49%)	45 (9.91%)	0 (0.00%)	454 (100.00%)
I used more technical review designs, when appropriate for the specific review.	64 (14.10%)	131 (28.85%)	234 (51.54%)	25 (5.51%)	0 (0.00%)	454 (100.00%)
I concisely documented review reasoning, findings, and conclusions.	11 (2.42%)	53 (11.67%)	260 (57.27%)	129 (28.41%)	1 (.22%)	454 (100.00%)
My review techniques guard against misinterpretations, biases, distortions, and errors.	15 (3.30%)	40 (8.81%)	254 (55.95%)	145 (31.94%)	0 (0.00%)	454 (100.00%)



Figure 6. Percentages of crowdsourcing workers’ use of the Accuracy domain of the JCSEE program evaluation standards.

Follow-up questions were also used to gather qualitative data in the Accuracy domain to inform this body of research even further. The following are the qualitative responses relative to the Accuracy domain. If item “I used a system of information management” were answered Seldom or Never, a follow-up question was triggered. In this case, the follow-up question was, “Briefly describe your system of information management.” There were three main themes that came from this follow-up question. The first theme was Systematic Process, which was named based on patterns that reflected the use of or creation of systematic process of information management. The second theme revealed was Unsystematic Processes. These were patterns of responses that reflected behavior that was unorganized and lacking systemization. The third theme was Random and Unrelated, which were text and phrases that did not fit into the previous patterns, and did not form a pattern of their own.

There were 42 responses to this follow-up question. Results from each theme are as follows: Systematic Process had 11 responses, Unsystematic Process had 23 responses, and Random and Unrelated had eight responses.

The results of the follow-up question “Briefly describe your system of information management” indicate that the majority of subjects were not using a systematic system of information management. This finding shows an emerging pattern indicating that a portion of the subjects in this study were not using systematic processes to perform product reviews, and therefore not adhering to the JCSEE standards. Below are examples of unsystematic responses:

“I go with my gut instinct and logic.”

“Experience.”

“I used my opinion then research.”

“Weighted my opinions based on need.”

Throughout the course of this section, evidence of non-systematic and non-comprehensive product evaluation has been prevalent. The JCSEE standards are very clear and transparent. They embody and project a foundation of evaluation standards that rely on complete, comprehensive, and scientific evaluation standards. Open-ended responses from participants in this survey reveal a core failure to align with and adhere to the JCSEE standards. The JCSEE standards go farther than experiential-based opinions.

Cross-tabulations

Cross-tabulations were conducted on multiple items to determine if a deeper understanding of this population sample could be obtained. The primary interest was to understand how reviewers thought about reliability and their approach to reviews, training, and validity. Unfortunately, no new information was revealed in these cross-

tabulations. However, that said, the consistency of item responses and follow-up questions is very positive and encourages the belief that the sample frame used from MTurk is both authentic and candid.

Table 10 shows a cross-tabulation of subject survey results for items, “Were your reviews solely based on past experience?” and “My reviews were reliable.” Responses for the item “My reviews were reliable” included Never, Seldom, Often, and Always.

Table 10

Cross-tabulation of Were Your Reviews Solely Based on Past Experience and My Reviews Were Reliable

My reviews were reliable		Were your reviews solely based on past experience?		Total
		Yes	No	
Never	Count	0	1	1
	% Within My reviews were reliable	0.0%	100.0%	100%
	% Within Were your reviews solely based on past experience?	0.0%	5.0%	.8%
Seldom	Count	0	2	2
	% Within My reviews were reliable	0.0%	100.0%	100.0%
	% Within Were your reviews solely based on past experience?	0.0%	10.0%	1.5%
Often	Count	9	8	17
	% Within My reviews were reliable	52.9%	47.1%	100.0%
	% Within Were your reviews solely based on past experience?	8.2%	40.0%	13.1%
Always	Count	101	9	110
	% Within My reviews were reliable	91.8%	8.2%	100.0%
	% Within Were your reviews solely based on past experience?	91.8%	45.0%	84.6%
Total	Count	110	20	128
	% Within My reviews were reliable	84.6%	15.4%	100.0%
	% Within Were your reviews solely based on past experience?	100.0%	100.0%	100.0%

Responses for the item “Were your reviews solely based on past experience?” were Yes and No. The results derived from responses within the item “Were your reviews solely based on past experience?” are as follow: Never: Yes ($n = 0$, 0.0%), No ($n = 1$, 0.8%); Seldom: Yes ($n = 0$, 0.0%), No ($n = 2$, 1.5%); Often: Yes ($n = 9$, 6.9%), No ($n = 8$, 6.2%); and Always: Yes ($n = 101$, 77.7%), No ($n = 9$, 6.9%).

Results from Table 10 only reinforce the idea that subjects show a strong endorsement to their product reviews being reliable based on their own personal experience. As stated earlier, prior research indicates that experiential-based opinions may not be reliable for product evaluations, and do not adhere to the JCSEE standards. Nevertheless, it appears based on these results that the only standards being used by this group of survey participants are personal and based on current experience and knowledge of the product. Therefore, the interpretation of these outcomes is that neither research question 1 nor research question 2 is being met.

Table 11 shows a cross-tabulation of subject survey results for items “My reviews were reliable” and “Do you have formal training for conducting product reviews?” Responses for the item “My reviews were reliable” included Never, Seldom, Often, and Always. Responses for the item “Do you have formal training for conducting product reviews?” were Yes and No. The results derived from responses within the item “Do you have formal training for conducting product reviews?” are as follows: Never: Yes ($n = 0$, 0.0%), No ($n = 3$, .7%); Seldom: Yes ($n = 1$, 0.2%), No ($n = 3$, 0.7%); Often: Yes ($n = 24$, 5.3%), No ($n = 97$, 21.4%); and Always: Yes ($n = 51$, 11.2%), No ($n = 275$, 83.3%).

The results of this analysis show, yet again, subjects’ confidence in the reliability of their reviews. As shown, 83.3% ($n = 275$) of the subjects do not have formal training

for conducting product reviews. It is difficult to think subjects could perform product reviews that adhere to the JCSEE standards without any formal training in evaluation.

Table 11

Cross-tabulation of My Reviews Were Reliable and Do You Have Formal Training for Conducting Product Reviews

My reviews were reliable		Do you have formal training for conducting product reviews?		Total
		Yes	No	
Never	Count	0	3	3
	% Within my reviews were reliable	0.0%	100%	100%
	% Within do you have formal training conducting product reviews?	0.0%	.8%	.7%
Seldom	Count	1	3	3
	% Within my reviews were reliable	25.0%	75.0%	100.0%
	% Within do you have formal training conducting product reviews?	1.3%	.8%	.9%
Often	Count	24	97	121
	% Within my reviews were reliable	19.8%	80.2%	100.0%
	% Within do you have formal training conducting product reviews?	31.6%	25.7%	26.7%
Always	Count	51	275	326
	% Within my reviews were reliable	15.6%	84.4%	100.0%
	% Within do you have formal training conducting product reviews?	67.1%	72.8%	71.8%
Total	Count	76	378	454
	% Within my reviews were reliable	16.7%	83.3%	100.0%
	% Within do you have formal training conducting product reviews?	100.0%	100.0%	100.0%

Table 12 shows a cross-tabulation of survey results for the items “Were your reviews solely based on past experience?” and “I used practical procedures.” Responses

for the item “Were your reviews solely based on past experience?” included Never, Seldom, Often, and Always. Responses for the item “I used practical procedures” were Yes and No. The results derived from responses within the item “Were your reviews solely based on past experience?” are as follows: Never: Yes ($n = 2$, 1.5%), No ($n = 1$, 0.8%); Seldom: Yes ($n = 6$, 4.6%), No ($n = 3$, 2.3%); Often: Yes ($n = 39$, 30.0%), No ($n = 10$, 7.7%); and Always: Yes ($n = 63$, 48.5%), No ($n = 6$, 15.4%).

Table 12

Cross-tabulation of Were Your Reviews Solely Based on Past Experience and I Used Practical Procedures

I used practical procedures		Were your reviews solely based on past experience?		Total
		Yes	No	
Never	Count	2	1	3
	% Within Were your reviews solely based on past experience?	66.7%	33.3%	100%
	% Within I used practical procedures?	1.5%	0.8%	2.3%
Seldom	Count	6	3	5
	% Within Were your reviews solely based on past experience?	66.7%	33.3%	100.0%
	% Within I used practical procedures?	4.6%	2.3%	6.9%
Often	Count	39	10	49
	% Within Were your reviews solely based on past experience?	79.6%	20.4%	100.0%
	% Within I used practical procedures?	35.5%	50.0%	37.7%
Always	Count	63	6	69
	% Within Were your reviews solely based on past experience?	91.3%	8.7%	100.0%
	% Within I used practical procedures?	57.3%	30.0%	53.1%
Total	Count	110	20	130
	% Within Were your reviews solely based on past experience?	84.6%	15.4%	100.0%
	% Within I used practical procedures?	100.0%	100.0%	100.0%

Table 13 shows a cross-tabulation of subject survey results for the items “Do you have formal training for conducting product reviews?” and “What is your role in crowdsourcing?” Responses for the item “Do you have formal training for conducting product reviews?” were Yes and No. Responses for the item “What is your role in crowdsourcing?” were Coordinator, Manager, Director, Reviewer, and Other. The results derived from responses within the item “Do you have formal training for conducting product reviews?” include Coordinator: Yes ($n = 9$, 45.0%), No ($n = 11$, 55.0%); Manager: Yes ($n = 10$, 47.6%), No ($n = 11$, 52.4%); Director: Yes ($n = 4$, 66.7%), No ($n = 2$, 33.3%); Reviewer: Yes ($n = 61$, 14.4%), No ($n=364$, 85.6%); and Other: Yes ($n = 1$, 16.7%), No ($n = 5$, 83.3%).

Table 13 results clearly show the majority of survey participants do not have formal training ($n = 394$, 86.8%). Further, the largest segment of participants was reviewers ($n = 425$, 93.6%). The fact that the majority of MTurk reviewers do not have formal training may explain why MTurk reviewers do not use more technical procedures when product evaluations become more technical and that they are less likely to use more sophisticated means of data management systems, learned from open item responses. All of this contradicts what was shown in the survey results. Although survey workers seem to indicate some use of standards according to overall quantitative survey results, yet do not have formal training in product evaluation in spite of identifying as product reviewers, almost precludes use of more technical procedures and systems of data management.

Table 13

Cross-tabulation of Do You Have Formal Training for Conducting Product Reviews and What is Your Role in Crowdsourcing

What is your role in crowdsourcing?		Do you have formal training for conducting product reviews?		Total
		Yes	No	
Coordinator	Count	9	11	20
	% Within What is your role in crowdsourcing? Coordinator	45.0%	55.0%	100.0%
	% Within Do you have formal training for conducting product reviews?	5.2%	6.3%	11.5%
Manager	Count	10	11	21
	% Within What is your role in crowdsourcing? Manager	47.6%	52.4%	100.0%
	% Within Do you have formal training for conducting product reviews?	5.8%	6.3%	12.1%
Director	Count	9	8	17
	% Within What is your role in crowdsourcing? Director	66.7%	33.3%	100.0%
	% Within Do you have formal training for conducting product reviews?	5.2%	4.6%	9.8%
Reviewer	Count	101	9	110
	% Within What is your role in crowdsourcing? Reviewer	14.4%	85.6%	100.0%
	% Within Do you have formal training for conducting product reviews?	58.0%	5.2%	63.2%
Other	Count	1	5	6
	% Within What is your role in crowdsourcing? Other	16.7%	83.3%	100.0%
	% Within Do you have formal training for conducting product reviews?	.5%	2.9%	3.4%

Finally, Table 14 shows a cross-tabulation of subject survey results for items “Were your reviews solely based on past experience?” and “I used more technical reviews designs, when appropriate for the specific review” Responses for the item “Were your reviews solely based on past experience?” were Yes and No. Responses for the item

“I used more technical reviews designs, when appropriate for the specific review” were Never, Seldom, Often, and Always. The results derived from responses within the item “Were your reviews solely based on past experience?” were Never: Yes ($n = 38$, 29.2%), No ($n = 6$, 4.6%); Seldom: Yes ($n = 44$, 33.8%), No ($n = 7$, 5.4%); Often: Yes ($n = 25$, 19.2%), No ($n = 6$, 4.6%); and Always: Yes ($n = 3$, 2.3%), No ($n = 1$, .80%). These results appear to be what would be expected given the results we saw in the other tables. Respondents have concluded that there is little change in technical designs driven by the scope and difficulty of the review, as they were based on the reviewers’ own experience and use of the item being reviewed.

A majority ($n = 182$, 83.8%) of subjects indicated they Never or Seldom use more technical evaluation designs when more technical evaluations called for it. This may be further indication that subjects were relying on personal experience to endorse their opinion in a product review. Below are actual responses taken from the survey open-ended response item “Briefly describe the product review designs which are used when product reviews become more technical.”

“They must be clear and innovative.”

“I compare the item with a similar competitor's item.”

“I stick to the same style of review but if it is for a more technical audience I'll be less concerned about ‘talking down’ to my audience and limiting the technical jargon.”

“Attributes, graphics, charts.”

“Video reviews”

“I make sure I add all details.”

“Owner’s manual, mostly Google the information needed.”

“Photographs, videos, research on proper use of the items as needed.”

“Being hands on in general and seeing how it functions and what makes it tick.”

“Used spec sheets, etc. when necessary.”

“I disclose the specifications of the products.”

“Detailed analysis of product uses and applications.”

Table 14

Cross-tabulation of Were Your Reviews Solely Based on Past Experience and I Used More Technical Reviews Designs, When Appropriate, for the Specific Review

I used more technical reviews designs, when appropriate for the specific review		Were your reviews solely based on past experience?		Total
		Yes	No	
Never	Count	38	6	58
	% Within I used more technical reviews designs, when appropriate for the specific review.	86.4%	13.6%	100.0%
	% Within Were your reviews solely based on past experience? % of Total	34.5%	30.0%	33.8%
Seldom	Count	44	7	124
	% Within I used more technical reviews designs, when appropriate for the specific review.	86.3%	13.7%	100.0%
	% Within Were your reviews solely based on past experience? % of Total	40.0%	35.0%	39.2%
Often	Count	25	6	31
	% Within <i>Were your reviews solely based on past experience?</i>	80.6%	19.4%	100.0%
	% Within <i>Were your reviews solely based on past experience?</i> % of Total	22.7%	30.0%	23.8%
Always	Count	3	1	4
	% Within I used more technical reviews designs, when appropriate for the specific review.	75.0%	25.0%	100.0%
	% Within Were your reviews solely based on past experience? % of Total	2.3%	0.8%	3.1%

These examples of open-ended responses from subjects illustrate their idea of more technical approaches to more technical product reviews. These are not examples of scientific or systematic evaluation, and do not adhere to the JCSEE standards.

CHAPTER V

CONCLUSION

This dissertation set out to determine what, if any, evaluation standards are used by crowdsourcing organizations and their reviewers, and to what extent these standards adhere to accepted evaluation standards used by professional evaluators in the evaluation discipline, precisely, the Utility, Feasibility, Propriety, and Accuracy domains of the JCSEE Program Evaluation Standards (Yarbrough et al., 2011). The literature shows significant changes have taken place recently with respect to the way consumers acquire evaluative information about products. The literature also shows these new methods of acquisition may not be as reliable and forthright as consumers expect. Yet, there is a paucity of empirical evidence to support this assumption. Given that the role of crowdsourcing organizations in evaluation has only recently developed, little is known about the standards used by crowdsourcing reviewers, particularly those found on open, self-serve sites such as MTurk. The results of this study provide insight into this newly developed area of product evaluation.

Context of the Study

Consumers have relied on rigorous and proven product evaluation standards to drive summative and formative product evaluations. In the past, Consumer's Research and Good Housekeeping were household names consumers relied on for consumer product evaluations. Historical evidence and the literature has shown these consumer product testing facilities invested in testing equipment, fixtures, and apparatuses to appropriately determine if the products being tested were performing as expected and meeting the understood consumer standards for acceptable consumer products.

Consumer's Research and Good Housekeeping both became formidable consumer products testers and reviewers in 19th century. Chuppa (2005) posited that Good Housekeeping presented itself as an authority in product testing and evaluation by testing and guaranteeing all products advertised in its magazine. Chase and Schlink (1927) identified Consumer's Research as the company consumers could depend on for unbiased and objective product testing. Kelley (2010) offered that most individuals have accessed a Consumer's Reports magazine to review product tests or opinions on consumer products. To that end, Consumer's Reports (2016) conducts 7,000 product evaluations annually and has a readership of 7,300,000 per month. Clearly, these numbers provoke a case for, justifiably, Consumer's Reports as being *the* consumers' source for consumer product evaluation.

Beyond Good Housekeeping and Consumer's Reports, Underwriters Laboratory is another popular product testing organization. According to the Underwriters Laboratory's (UL) website, in 1998, UL had over 14 billion UL trademarks on tested products. Further, UL has developed over 600 standards for product safety. UL is a non-profit organization designed to test products to determine their level of safety and conformity to safety standards in the household. Again, UL is an organization Americans look to for unbiased and objective product testing.

Good Housekeeping, Consumer's Reports, and Underwriters Laboratory are companies Americans have turned to in the past for product evaluation, safety, and recommendations. These companies are highly recognizable organizations, with extensive credibility in the eyes of consumers. It is important to note, however, there are some criticisms of these organizations. Scriven (1994), for example, expressed his

dissatisfaction about Consumer's Reports, citing lack of transparency, methodology, and qualified comparisons as reasons Consumer's Reports falls short of expectations for the highest quality product evaluation.

Technology has given consumers new tools to access product evaluations or product reviews. UNCTAD (2002) determined that online product reviews were important or extremely important to consumers. Jiang and Wang (2008) posited that ever since Amazon.com began publishing book reviews, an estimated 43% of online retailers offer the same service today. Floyd et al. (2012) cited a study conducted by Cisco Internet Business Solutions Group (2013), which used a sample of consumers ($N = 5,000$) to indicate the top three sources of information to make purchase decisions. Results revealed consumers reported online reviews on websites were used 52 percent of the time, while friends and family were used 49 percent, and advice from store employees 12 percent. Further, Nielson (2013) determined that there was a 120 percent incremental increase in smart phone (mobile) application use from 2011 to 2012. Clearly, this means this is a growing and useful practice for consumers. Further, as technology develops, access and scope of online reviews will continue to broaden.

Another, perhaps more influential, technology piece in this puzzle is the use of crowdsourcing for online product reviews. As Howe (2006) stated, crowd sourcing is taking a large group of consumers and eliciting their help or opinion to accomplish some goal or objective. The cost is minimal and the speed is unprecedented over former market feedback mechanisms. Open source crowdsourcing sites like Amazon.com owned MTurk allow the general public to use the sites' workers at a low cost and without much expertise. Hienerth and Rier (2007) argued that the relevance of crowd-based evaluation

is high. Biewald et al. (2010) noted research from crowdsourced applications may be superior to traditional, structured methods. Paolacci and Chandler (2014) argued researchers may use MTurk for any study suitable to be conducted online. Buhrmester and colleagues (2011) concluded MTurk workers were sensitive to pay rates and duration of time to complete an HIT, but their research did not conclude quality was dependent on pay rate. Chen et al. (2003) identified a problem of objectivity in reviews conducted online with producer-supported sites, and also concluded that online reviews tended to have small sample sizes. Further, Segal (2011) reported a case where MTurk workers were asked to write fake reviews for Southland Dentistry and post them on Southland's Yelp page.

The threat of unreliability when using an open crowdsourcing platform is unmistakable. Chen et al. (2003) determined that the number of postings in online reviews can improve accuracy, but bias is created by monetary incentives. Ott, Cardie, and Hancock (2012) argued that deception was higher in instances where posting requirements were lower. Conversely, deception was lower where posting requirements were higher. Further, it appears that when reviews and postings are either hidden or reviewed for authenticity, the incidence for deception decreased dramatically. Open sites like Yelp and Urban Spoon do invoke algorithms to detect deceptive reviews. This is based on many factors, a few of which are the number of reviews completed by a reviewer and reviews endorsed by other users. However, Chandler, Paolacci, and Mueller (2013) argued that deception within MTurk workers may increase substantially if they knowingly benefit from lying.

Therefore, a compelling dilemma presented itself in a most peculiar manner. In the context of a social desire for learning and information gathering among many consumers, a critical gap exists that prohibits the ability to consciously and risk adversely move forward. Consumers are very connected to media. Consumers appear to want reliable product evaluation within the context of accessibility and ease. Connectivity allows both information and transformation. However, it is critical to put into context levels of both certainty and acuity of online reviews. To start to understand this perceived gap, there was a need to understand what crowdsourcing workers conducting product reviews were using as standards for product evaluations.

To help determine what standards crowdsourcing workers were using, the researcher needed for a recognized and accepted set of evaluation standards that were comprehensive and applicable to product evaluation. The Joint Committee for Standards in Education Evaluation (JCSEE) Program Evaluation Standards (Yarbrough et al., 2011) were used to develop a survey to determine:

1. What, if any, evaluation standards are being used by crowdsourcing organizations and their reviewers.
2. To what extent do these standards adhere to accepted evaluation standards used by professional evaluators in the evaluation discipline, precisely, the Utility, Feasibility, Propriety, and Accuracy domains of the JCSEE Program Evaluation Standards (Yarbrough et al., 2011).

Why the JCSEE standards? The JCSEE standards provided known and accepted standards for evaluation practice. The JCSEE standards were developed by a diverse group of teachers, counselors, statisticians, psychometrics, evaluators, policy makers, and

administrators. The JSCEE standards are accepted as an ANSI standard, and are the most recognized, accepted, and credentialed evaluation standards to date by many evaluation practitioners. JSCEE standards were used as a model to develop a survey designed and deployed in this study to answer specific research questions.

Stufflebeam (2014) noted the JSCEE standards are used in textbooks. Durabi (2002), for example, noted the use of JCSEE standards during evaluation courses wherein students were required to use the JCSEE standards as guidelines and principles while completing class projects, and prescribed using and integrating the JCSEE standards in all five steps of evaluation logic model development. The JCSEE standards have also been used in evaluation practice by industry theorists, and were adopted by the U.S. Marines for personnel evaluation. The JCSEE standards are comprehensive and invoke inquiry using five relevant evaluative domains (i.e., Utility, Feasibility, Propriety, Accuracy, and Accountability), making them one of the most comprehensive standards used in many different types of evaluation. Stufflebeam (2004) posited that the JCSEE standards were used extensively to help evaluators see limitations to standardized tests and experimental design studies within academia and the social sciences.

Review of Methodology

A descriptive, cross-sectional survey design was used for this study. The purpose of this study was to explore if crowdsource workers, in particular, Amazon MTurk product reviewers, use any evaluation standards, and if so, how closely these standards adhere to the JCSEE program evaluation standards (Yarbrough et al., 2011).

It was critical that items on the questionnaire allowed for the collection of data that would answer the study's research questions. Accordingly, the items on the

questionnaire in this study consisted of statements from the Utility, Feasibility, Propriety, and Accuracy domains of the JCSEE program evaluation standards (Yarbrough et al., 2011). This allowed the researcher to determine if crowdsource workers who participated in the study were able to appropriately apply each of the specific JCSEE standards during product reviews.

As stated earlier in this dissertation, MTurk was unable to provide information on how many workers are located within the United States (MTurk, 2015). Thus, a non-probability sampling method was used to gather information pertaining to the study's focal research questions. Therefore, the researcher's goal was to obtain a sample size of 500 MTurk workers with five years of product review experience, and a minimum rating of 95 percent.

Instrumentation for this dissertation included a 27-item survey developed by the researcher specifically for the purpose of this study and an 11-item demographic questionnaire (Appendix C). The items on the survey represented the four domains of the JCSEE program evaluation standards (Yarbrough et al., 2011) that were explored in this study (i.e., Utility, Feasibility, Propriety, and Accuracy), and assessed how often reviewers considered each program evaluation standard when conducting product evaluations or reviews within the last 12 months on the following 4-point Likert-type scale: 1 = *Never*, 2 = *Seldom*, 3 = *Often*, 4 = *Always*. When selecting *never* or *seldom*, participants were prompted to respond to an open-ended question that allowed participants to explain their answers.

Summary of Empirical Findings

As stated, the first research question in this study explored what, if any, evaluation standards are being used by crowdsource organizations and their reviewers. Based on the results of this study, the only visible and detectable evaluation standard being used by MTurk product reviewers is personal experience.

Although the survey results suggest close adherence to the JCSEE standards on the surface, deeper qualitative analysis conducted in Chapter IV of open item responses suggested that MTurk product review workers primarily use personal experience. Fundamentally, the prevalence of personal opinion found throughout open-ended item responses do not adhere to JCSEE standards. Further, there was not sufficient evidence to suggest any additional evaluation procedures were performed to support personal opinions.

Finally, Ott et al. (2012) concluded tastes and preferences may change over time with experience and knowledge. This change is considered personal to consumers and product users. Therefore, this information posits that a reviewer's opinion today may change with more experience and knowledge. This has implications for on-line product reviews. Minimally, this puts into question the reliability of experience-based product reviews. There is no way to determine the depth of a reviewer's experience or knowledge of a product being reviewed.

The second research question in this study explored to what extent the standards used by reviewers adhere to accepted evaluation standards used by professional evaluators in the evaluation discipline, precisely, the Utility, Feasibility, Propriety, and Accuracy domains of the JCSEE Program Evaluation Standards (Yarbrough et al., 2011).

Conclusively, deeper analysis of qualitative data conducted in Chapter IV determined that the MTurk product reviewers sampled for this study did not adhere to all of the JCSEE standards in each domain, and further concludes the standards, domains, were *not* being adhered to as an aggregate outcome. The following table summarizes these results.

Table 15

Summary of Adherence to JCSEE Standards

Standard	Adherence to Standard
Utility	
I completed reviews of products without prior experience using them.	Yes
I completed reviews with an understanding of whom I was completing them for.	No
I had a clear understanding of the product review's purpose.	No
I had complete understanding of the requestor's values underpinning the review.	No
I was informed of the needs of users of the reviews.	No
My reviews encouraged use.	Yes
My review submissions were on time.	Yes
My reviews were misused.	Yes
Feasibility	
I used project management techniques to complete reviews.	No
I used practical procedures.	No
My reviews recognized and balanced cultural and political interests of individuals and consumers.	No
I used resources effectively and efficiently.	Yes
Propriety	
My reviews were responsive to potential consumers.	Yes
I was involved in negotiations with requesters about the reviews.	No
My reviews considered consumer rights and safety.	Yes
My reviews were clear and fair.	Yes

Table 15—Continued

Standard	Adherence to Standard
Propriety	
My reviews disclosed all relevant information.	No
My reviews disclosed conflicts of interest.	No
My reviews expended out of pocket resources.	Yes
Accuracy	
My review conclusions are or were justified.	Yes/No
My review conclusions were valid.	No
My reviews were reliable.	No
I thoroughly documented reviews with great detail.	No
I used a system of information management.	No
I used more technical review designs, when appropriate for the specific review.	No
I concisely documented review reasoning, findings, and conclusions.	No
My review techniques guard against misinterpretations, biases, distortions, and errors.	No

Further analysis of qualitative data can be found in the Appendices I and J. The results section of this dissertation showed a systematic organizing and cataloging of themes for each of the open response items deployed in the survey used for this dissertation. The most prevalent method described in each of the open response sections is personal use and experience. The responses lack a discussion about methods to determine if personal experiences or outcomes can be duplicated or are even shared in some common group of users. Further, there is no obvious or hint of reference to evaluation standards being used.

Evaluation methods and procedures were rarely noted among participants, and when directly queried, respondents referred to personal experience examples as a

reference point. The most technical references seemed to come from the open item response to information management. In this response window, many references were made to using more technical hardware, databases, and equipment from work. Interestingly, in this same window, the reviewers spoke of accessing specifications and data on other sites and sources of data for temporary use. These examples can be used in product evaluation, and in some ways adhere to the standards as shown above.

Missing from the qualitative examples are identifiable processes indicating product evaluation standards are being used. Furthermore, there is no evidence of a methodology that guards against misinterpretation, bias, distortions, and errors; that is, a method in conjunction with standards that can be seen as durable and reliable. As discussed earlier in this chapter, personal experience based opinions can change over time and with the acquisition of knowledge. There are no profound examples found in any of the open response windows that counter the fact that the MTurk product reviewers used for this survey were using personal experience as a bases to draw opinions about products tested. Crowdsourced product reviews are, therefore, weak in reliability, and based on the open testimony given in this survey, can be biased.

Implications

Stufflebeam (2014) argued, “The Standards are widely referenced in evaluation textbooks and other evaluation literature. They are used extensively in evaluation courses. There is evidence of widespread use in practice, though more use is needed.” The JCSEE standards are considered an ANSI standard. ANSI is a global organization for the purpose of establishing standards in nearly all industries, and is considered the highest distinction of standards (<http://webstore.ansi.org/documents/What-Is-ANSI.pdf>). As

indicated, the organization of the JCSEE standards is very precise and consists of five domains: (1) Utility, (2) Feasibility, (3) Propriety, (4) Accuracy, and (5) Accountability. In this dissertation, with the minimal applicable relationship of accountability to product reviewers, the Accountability domain was not used. The JCSEE standards are clear and objective standards that are designed to provide a comprehensive approach to systematic evaluation. The research for this dissertation sought to determine if the standards used by MTurk crowdsource product reviewers adhere to four of these domains. Deep analysis of the qualitative data found in this survey was determined to be an implied standard of personal experience, and that they do not adhere to the JCSEE standards.

Based on the above results, there are several implications for this study. First, given the open response item analysis, there was a significant lack of systematic inquiry being performed by the MTurk product reviewers. Most, if not all, reviews were based on past experience, which in and of itself does not employ nor imply systematic inquiry. To summarize, the results of this study show a significant lack of systematic inquiry being performed by the MTurk product reviewers, and most, if not all, reviews were based on past experience. Further, there did not appear to be a focus on building a base of data.

A second implication of this study concerns the reliability of crowdsourced product evaluations. Jacobsen (2015) noted, "I find clear evidence that consumer ratings are swayed by the reviews of experts, increasing their ratings in response to positive expert reviews and decreasing them in response to negative expert review" (p. 20). In addition, there are inconsistent results coming from online reviews compared to formal product evaluations. As Chen, Fay, and Wang (2003) noted, "For models that have extremely high or extremely low quality, reported ratings differ substantially from the

ratings reported by *Consumer Reports*” (p. 17). Consumers expect reliable and credible reviews. In the survey used for this dissertation the question stem read: *My reviews were reliable*. If always was selected, the qualitative responses ($n = 257, 73\%$) indicated reviews were based on participants own real life experiences. For reasons described earlier in this chapter, it is clear that using personal experience as a standard does not adhere to the JCSEE standards, particularly, the Accuracy domain.

A final implication of this study concerns the classification of crowdsourced product evaluations. Stufflebeam and Shinkfield (2007) noted, “The methodology of the criticism and connoisseurship includes critics’ systematic use of their perceptual sensitivities, past experiences, refined insights, and abilities to communicate their assessments” (p. 184). Is this the approach we are seeing used with online reviews and crowdsourced product reviews? It certainly seems to be. As Robbins (2006) noted,

Connoisseurship involves expert norm-referenced judgements being made by a person who is recognized as having the knowledge and experience necessary to do so. Examiner, observer, rater and assessor are all terms that may be used in different settings to describe this person. (p. 3)

This approach is used in evaluation or assessments of teachers, musicians, actors, firefighters, police, EMTs, medical personnel, food, wine, beer, and consumer products in general. Connoisseur product evaluation can be used in combination with other evaluation models as well. This approach is meant to be conducted by not only an expert in the field, but a person close in the community of the object being evaluated or assessed, ensuring the context of the evaluation is not a barrier to success (Stufflebeam & Shinkfield, 2007, p. 184).

After reviewing the details of the information provided by the MTurk product reviewers, categorically, this would explain the methodology and perhaps even the standard being used in online reviews. The exception that dispels this possibility, however, is that an observed, noted, and respected expert must conduct connoisseur product evaluations. That is not the case presented in the given sample used for this dissertation. Therefore, a case to classify online reviews as connoisseur type product evaluation cannot exist if we are to consider and follow the guidelines that Stufflebeam and Shinkfield (2007, p. 184) have given us.

Contributions

The practical contributions of this research may be many. First, as it relates to evaluation practices, people working in evaluation roles may be able to use this information to gain more understanding of the uses and limitations of crowdsourced product evaluations and online product reviews. This dissertation was able to identify the shortcomings of crowdsourced online product reviews. Evaluation practitioners can use this information to plan and design evaluations and evaluative research, or to develop specific standards that can be targeted toward online product reviewers.

Second, the results and conclusions for the research conducted in this dissertation can inform consumer marketing researchers and professionals in a manner that provides insight into greater uses of evaluation standards, methods, and expectations. Consumers expect usable and reliable information from online product reviews. Research confirms the uses of online product evaluation are increasing and that consumers are not getting the information they expect from online product reviews. Further, the literature has

shown that open crowdsourcing sites are gaming the system and are quite capable of facilitating bogus reviews with no detection, leaving the consumer misinformed.

Finally, this research confirms the fact that online reviews are opinion-based. Further, it reconfirms that these opinions can be biased and even untruthful. Therefore, this research and other similar research can be used to explore the plausibility of federal penalties for completing dishonest or distorted online reviews.

Limitations of the Study

The limitations of this study are as follows. First, as in any study, there are limitations to this study's instrumentation. A content validity analysis was not conducted on the instrument to ensure the instrument was appropriately designed to achieve optimal measurement outcomes. Terms and definitions may not have been clear to participants, and were not available. It is unclear if subjects understood all terms and definitions. This could have caused inaccurate responses, measurements, and interpretations within the quantitative data. According to Paulhus (1991), "a *response bias* is a systematic tendency to respond to a range of questionnaire items on some basis other than the specific item content (i.e., what the items were designed to measure)" (p. 1). Furthermore, positive responses or socially desirable responses (SDR) are responses subjects give to make them seem look better (Paulhus, 1991). It is possible that a SDR bias in this dissertation may have persisted enough to skew survey results favoring positive outcomes to adherence of the JCSEE standards.

A second limitation is that a non-probabilistic sample was used for this research. MTurk was not able to provide statistics for the population of product reviewers working

on the MTurk site. A non-probabilistic sample does not provide necessary information to accurately generalize this information beyond this research.

A final limitation concerns the question whether the JCSEE standards are appropriate standards to be used in determining if MTurk product reviewers are using appropriate standards for product evaluation. The JCSEE standards are incredibly useful when conducting an evaluation. Are they the correct standards for this research of MTurk product reviewers? Without a successful content validity analysis, it is unclear if the JCSEE standards can be made to work properly.

Research Opportunities

There are many research areas that are attractive for the domain of crowdsourced online reviews. It would be beneficial to learn about methods used by product reviewers: Are these methods transferable into any form of evaluation? If these workers followed a specific protocol, could there be a methodology that can be accepted and standardized for online reviews? Further, it would be important to learn more about how consumers use product reviews, and what they actually do with that information. Is it taken at face value? Are online product reviewers interested in learning how to perform formal product evaluation? This would be important to know as it informs stakeholders if a possibility exists to improve online review accuracy and reliability. This extends to the next question: Would companies be interested in deeper systematic inquiry if it improved review accuracy? Again, the more accurate and reliable we can make online reviews the more consumers can reliably depend on them.

REFERENCES

- Abelson, R. P., & Levi, A. (1985). Decision making and decision theory. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (3rd ed., Vol. 1, pp. 231-309). Hillsdale, NJ: Erlbaum.
- Azzam, T., & Jacobson, M.R. (2013). Finding a comparison group: Is online crowdsourcing a viable option? *American Journal of Evaluation*, 34, 372-384.
- Bagozzi, R. P., & Dholaki, U. M. (2006). Open source user communities: A study of participation in Linux user groups. *Journal of Management Science*, 52, 1099-1115.
- Barron, M. R. (2007). Creating consumer confidence or confusion: The role of product certification marks in the market today. *Marquette Intellectual Property Law Review*, 11, 413-442. Retrieved from <http://scholarship.law.marquette.edu/cgi/viewcontent.cgi?article=1099&context=iplr>
- Battaglia, M. P. (2008). Nonprobability sampling. *Encyclopedia of survey research methods*. Thousand Oaks, CA: Sage Publications.
- Boyatzis, R. E. (1998). *Transforming qualitative information: Thematic analysis and code development*. Thousand Oaks, CA: Sage.
- Braun, V., & Clarke, V. (2006) Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3, 77-101.
- Buhmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high quality, data? *Perspectives on Psychological Science*, 6, 3-5.

- Chandler, D., & Kapelner, A. (2013). Breaking monotony with meaning: Motivation in crowdsourcing markets. *Journal of Economic Behavior & Organization*, 90, 123-133.
- Chase, S., & Schlink, F. J. (1927). *Your money's worth*. New York, NY: The Macmillan Company.
- Chen, Y., Fay, S., & Wang, Q. (2003). Marketing implications of online product reviews. Retrieved from <http://plaza.ufl.edu/faysa/review.pdf>
- Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, 43, 345-354.
- Chilton, L. B., Horton, J. J., Miller, R. C., & Azenkot, S. (2010). Task search in a human computation market. Retrieved from <https://groups.csail.mit.edu/uid/other-pubs/search-HCOMP2010.pdf>
- Dasgupta, A., & Ghosh, A. (2013). Crowdsourced judgement elicitation with endogenous proficiency. Retrieved from <http://arxiv.org/pdf/1303.0799v1.pdf>
- Forman, C., Ghose, A., & Wiesenfeld, B. (2008). Examining the relationship between reviews and sales: The role of reviewer identity disclosure in the electronic markets. *Information Systems Research*, 19, 291-313.
- GAO (U.S. Government Accountability Office). (2011). *Government auditing standards*. Retrieved from <https://www.archives.gov/oig/pdf/government-auditing-standards-2011-revision.pdf>
- Hall, J. (2008). Cross-sectional survey design. In P. J. Lavrakas (Ed.), *Encyclopedia of survey research methods* (pp. 172-173). Thousand Oaks, CA: Sage.

- Hienerth, C., & Rier, F. (2013). The wisdom of the crowd vs. expert evaluation: A conceptualization of evaluation validity. Paper presented at the 35th DRUID Celebration Conference, Barcelona, Spain. Retrieved from http://druid8.sit.aau.dk/acc_papers/im4odl75ix421p8l11ry1qg6gko8.pdf
- Hofstadter, R. (1965). *The paranoid style in American politics*. New York, NY: Alfred A. Knopf, Inc.
- Howe, J. (2006, June). The rise of crowdsourcing. *Wired*, 14(06), 1-4. Retrieved from <http://www.wired.com/2006/06/crowds/>
- Huberman, B. (2008, November). Crowdsourcing and attention. *Computer Magazine*, 135, 103-105.
- Jacobsen, G. D. (2015). Consumers, experts, and online product evaluations: Evidence from the brewing industry. Retrieved from <http://pages.uoregon.edu/gdjaco/Beer.pdf>
- Jeppesen, L. B., & Frederiksen, L. (2006). Why do users contribute to firm-hosted user communities? The case of computer-controlled music instruments. *The Journal of Organizational Science*, 17, 45-63.
- Jiang, B.-J., & Wang, B. (2008). Impact of consumer reviews and ratings on sales, prices, and profits: Theory and evidence. ICIS 2008 Proceedings. Paper 141. Retrieved from <http://aisel.aisnet.org/icis2008/141>
- Johnson, T., & Fendrich, M. (2002). A validation of the Crowne-Marlowe Social Desirability Scale. Paper presented at the American Association for Public Opinion Research, St Petersburg, FL.

- Joint Committee on Standards for Educational Evaluation. (1994). *The program evaluation standards: How to assess evaluations of educational programs*. Thousand Oaks, CA: Sage.
- Kraut, R., Olson, J., Banaji, M., Bruckman, A., Cohen, J., & Couper, M. (2004). Psychological research online: Report of board of scientific affairs' advisory group on the conduct of research on the Internet. *American Psychologist*, 59, 105-117.
- Lakhani, K. R., & Wolf, R. G. (2005). Why hackers do what they do: Understanding motivation and effort in free/open source software projects. In J. Feller, B. Fitzgerald, S. Hissam, & K. R. Lakhani (Eds.), *Perspectives on free and open sources of software*. Cambridge, MA: MIT Press. Retrieved from <http://ocw.mit.edu/courses/sloan-school-of-management/15-352-managing-innovation-emerging-trends-spring-2005/readings/lakhaniwolf.pdf>
- Le, J., Edmonds, A., Hester, V., Biewald, L. (2010). Ensuring quality in crowdsourced search relevance evaluation: The effects of training question distribution. In *SIGIR 2010 workshop on crowdsourcing for search evaluation* (pp. 21-26).
- Lerner, J., & Tirole, J. (2002). Some simple economies of open source. *The Journal of Industrial Economics*, 50, 197-234.
- Luca, M. (2011, September 16). Reviews, reputation, and revenue: The case of yelp.com. Harvard Business School, Negotiations, Organizations and Markets Unit Working Paper No. 12-016. Retrieved from <http://ssrn.com/abstract=1928601>

- Matsuo, H., McIntyre, K. P., Tomazic, T., & Katz, B. (2004). The online survey: Its contributions and potential problems. *ASA Proceedings of the Section on Survey Research Methods*, 2004, 3998-4000.
- Mayzlin, D. (2006). Promotional chat on the Internet, *Marketing Science*, 25, 157-165.
- McAuley, J., & Lekovec, J. (2013). From amateurs to connoisseurs: Modeling the evolution of user expertise through online reviews. Retrieved from <http://arxiv.org/pdf/1303.4402v1.pdf>
- Morris, R., & Picard, R. (2012). Crowdsourcing collective emotional intelligence. *Proceedings of Collective Intelligence*. Retrieved from <http://affect.media.mit.edu/pdfs/12.Morris-Picard-CI2012.pdf>
- Nielson. (2013). The mobile consumer: A global snapshot. Retrieved from <http://www.nielsen.com/content/dam/corporate/uk/en/documents/Mobile-Consumer-Report-2013.pdf>
- Paolacci, G., & Chandler, J. (2014). Inside the Turk: Understanding Mechanical Turk as a participant pool. *Current Directions in Psychological Science*, 23, 184-188.
- Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2nd ed.). Newbury Park, CA: Sage.
- Poetz, M. K., & Schreier, M. (2012). The value of crowdsourcing: Can users really compete with professionals in generating new product ideas. *The Journal of Product Innovations*, 29, 245-256.
- Raymond, E. S. (1999). *The cathedral and the bazaar: Musings on Linux and open source by an accidental revolutionary*. Sebastopol, CA: O'Reilly Media, Inc.

- Robbins, J. H. (2006). Connoisseurship: Assessments of performance and questions of reliability. Paper presented at the 32nd Annual IAEA Conference, Singapore, China. Retrieved from https://www.academia.edu/2650240/Connoisseursip_assessments_performance_IAEA_2006
- Rubin, D. B. (2001). Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services & Outcomes Research Methodology, 2*, 169-188.
- Schmidt, F. L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for the training of researchers. *Psychological Methods, 1*, 115-129.
- Scriven, M. (1976a). Payoffs from evaluation. In C. C. Abt (Ed.), *The evaluation of social programs* (pp. 217-224). Beverly Hills, CA: Sage.
- Scriven, M. (1976b). Evaluation bias and its control. In G. V. Glass (Ed.), *Evaluation studies review annual* (Vol. 1, pp. 101-108). Beverley Hills, CA: Sage.
- Scriven, M. (1981a). The good news about product evaluation. *Evaluation News, 2*, 278-282.
- Scriven, M. (1981b). Product evaluation. In N. L. Smith (Ed.), *New techniques for evaluation* (pp. 121-166). Beverly Hills, CA: Sage.
- Scriven, M. (1993). *Hard won lessons in program evaluation*. San Francisco, CA: Jossey-Bass.
- Scriven, M. (1994). Product evaluation: The state of the art. *Evaluation Practice, 15*, 45-62.

Scriven, M. (2007). The logic of evaluation. OSSA Conference Archive. Paper 138.

Retrieved from

<http://scholar.uwindsor.ca/cgi/viewcontent.cgi?article=1390&context=ossaarchive>

Segal, D. (2011). A Rave, a pan, or just a fake? *The New York Times* [Online]. Retrieved

from http://www.nytimes.com/2011/05/22/your-money/22haggler.html?_r=0

Shadish, W. R., Jr., Cook, T. D., & Leviton, L. C. (1991). *Foundations of program*

evaluation: Theories and practice. Newbury Park, CA: Sage.

Shapiro, D. N., Chandler, J., & Mueller, P. A. (2013). Using Mechanical Turk to study

clinical populations. *Clinical Psychological Science, 1*, 213-220.

Strach, L., & Russell, M. (2003). The Good Housekeeping Seal of Approval: From

innovative consumer protection to popular badge of quality. *Essays in Economic and Business History*, 151-166. Retrieved from

http://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=9&ved=0ahUKEwjuiI_Txb3JAhVMqR4KHQwWDNQQFghYMAg&url=http%3A%2F%2Fwww.ebhsoc.org%2Fjournal%2Findex.php%2Fjournal%2Farticle%2Fdownload%2F105%2F100&usq=AFQjCNEfsEYI2JDC7QOm7P73i3HwXVa8zw

Streitfeld, D. (2012, August 25). The best book reviews money can buy. *The New York*

Times [Online]. Retrieved from

http://www.nytimes.com/2012/08/26/business/book-reviewers-for-hire-meet-a-demand-for-online-raves.html?_r=0

Stuart, E. A., & Rubin, D. B. (2008). Best practices in quasi-experimental designs. In J.

W. Osborne (Ed.), *Matching methods for causal inference* (pp. 155-176).

Thousand Oaks, CA: Sage.

- Stufflebeam, D. L. (1968). Evaluation as enlightenment for decision-making. Retrieved from <http://files.eric.ed.gov/fulltext/ED048333.pdf>
- Stufflebeam, D. (2001). Evaluation models. *New Directions for Evaluation*, 89, 7-98.
- Stufflebeam, D. L. (2004). A note on the purposes, development, and applicability of the Joint Committee Evaluation Standards. *American Journal of Evaluation*, 25, 99-102.
- Stufflebeam, D. L., & Coryn, C. L. S. (2014). *Evaluation, theory, models, and applications* (2nd ed.). Hoboken, NJ: Wiley Publications.
- Stufflebeam, D. L., & Shinkfield, A. J. (2007). *Evaluation, theory, models, and applications*. Hoboken, NJ: Wiley Publications.
- Tripodi, T., & Potocky-Tripodi, M. (2007). *International social work research: Issues and prospects*. Oxford: Oxford University Press.
- United Nations Conference on Trade and Development. (2002). *E-commerce and development report*. Retrieved from http://unctad.org/en/Docs/ecdr2002p1_en.pdf
- van Osselaer, S. M. J., & Janiezewski, C. (2012). A goal-based model of product evaluation and choice. *Journal of Consumer Research*, 39, 260-292.
- Waller, S. W., Brady, J. G., Acosta, R. J., & Fair, J. (2011, January 12). Consumer protection in the United States: An overview. *European Journal of Consumer Law*. Retrieved from <http://ssrn.com/abstract=1000226>
- Williams, G. L. (1995). Buyer beware: The Consumers Research Archives at the Rutgers University Libraries. *Journal of the Rutgers University Libraries*, LVII(1/2), 25-39.

- Williams, J. A. (2004). The impact of motivating factors on implementation of ISO 9001:2000 registration process. *Management Research News*, 27(1/2), 74-84.
- Yarbrough, D. B., Shula, L. M., Hopson, R. K., Caruthers, F. A. (2011). *The program evaluation standards: A guide for evaluators and evaluation users* (3rd ed.). Thousand Oaks, CA: Sage.
- Zhao, Y., Yang, S., Narayan, V., & Zhao, Y. (2013). Modeling consumer learning from online product reviews. *Marketing Science*, 32, 153-169.
- Zhao, Y., & Zhu, Q. (2014). Evaluation on crowdsourcing research: Current status and future direction. *Information Systems Frontiers*, 16, 417-434.

Appendix A

Human Subjects Institutional Review Board Approval


WESTERN MICHIGAN UNIVERSITY



Human Subjects Institutional Review Board

Date: August 5, 2015

To: Alexander Manga, Principal Investigator
Daniela Schroeter, Co-Principal Investigator for dissertation

From: Amy Naugle, Ph.D., Chair 

Re: HSIRB Project Number 14-04-06

This letter will serve as confirmation that your research project titled "Crowdsourcing Product Reviews: What are they? Are they Using Systematic Approach as to Methods and Standard Development" has been **approved** under the **exempt** category of review by the Human Subjects Institutional Review Board. The conditions and duration of this approval are specified in the Policies of Western Michigan University. You may now begin to implement the research as described in the application.

Please note: This research may **only** be conducted exactly in the form it was approved. You must seek specific board approval for any changes in this project (e.g., ***you must request a post approval change to enroll subjects beyond the number stated in your application under "Number of subjects you want to complete the study."*** Failure to obtain approval for changes will result in a protocol deviation. In addition, if there are any unanticipated adverse reactions or unanticipated events associated with the conduct of this research, you should immediately suspend the project and contact the Chair of the HSIRB for consultation.

Reapproval of the project is required if it extends beyond the termination date stated below.

The Board wishes you success in the pursuit of your research goals.

Approval Termination: August 4, 2016

1903 W. Michigan Ave., Kalamazoo, MI 49008-5456
PHONE: (269) 387-8293 FAX: (269) 387-8276
CAMPUS SITE: 251 W. Walwood Hall

Appendix B
HSIRB Application

CROWDSOURCE PRODUCT REVIEWS AND EVALUATIONS; REVIEW OF STANDARDS AND PRACTICE

Western Michigan University

HSIRB Application Principal Investigator: Alexander Manga

Co-Principal Investigator for Dissertation: Daniela Schroeter

Contents

Purpose and Background	129
Subject Recruitment	129
Setting up a Crowdsourcing Project	130
Recruitment of Reviewers	130
Informed Consent Process	131
Research Procedure	131
Methods for Data Collection	131
Instrumentation	132
Location of Data Collection	133
Duration of Study	133
Methodology	133
Design	133
Analysis	136
Dissemination	136
Confidentiality of Data	137
Risks and Costs to and Protection for Subjects	137
Benefits of Research	137

Abstract

The purpose of this study is to learn about crowdsourcing evaluation standards and methods. More specifically, this research seeks to determine what standards crowdsourcing evaluation workers are using and if these standards adhere to The Joint Committee on Standards for Educational Evaluation (JCSEE) Program Evaluation Standards (PES). This research will be conducted using a web-based survey that will collect data from 500 crowdsourcing workers using a self-serve crowdsourcing company; MTurk.com. The response rate of this study is expected to be 100%, because crowdsourcing workers will complete the survey as a job function. This is how crowdsourcing reviewers earn money, they accept contracts to review products, complete the review process, and are paid a pre-approved rate per review. The process for this study is straightforward: A reviewer will review an online solicitation by MTurk, review the terms, make a decision to participate or not, and if participating, complete the online survey for a small fee. Because of the operational model of crowdsourcing workers, Crowdsourcing reviewers do not work a conventional work week; they can complete reviews at all hours of the day seven days a week, the turnaround time for the surveys will be one to two weeks.. The results of the web-based survey will reveal crowdsourcing product review and evaluation standards, and explicate strategies used in online product reviews and evaluations. It is important to understand how these standards and methods compare to evaluation standards developed by the JCSEE, because we know from the literature that more and more people are using online product reviews and evaluations (Chandler, Mueller & Shapiro, 2013). However, we know little to nothing about the implicit standards and methods these crowdsourcing workers use and what their attitudes toward formal evaluation standards and methods are.

Evaluation professionals want to know how practitioners are conducting product evaluation in the field, it is important to understand what standards are being used in crowdsourced reviews, and how crowdsource evaluators feel about using changing the evaluation standards they are currently using.

Purpose and Background

Evaluation standards define evaluation quality and guide evaluators and consumers in the pursuit of quality evaluation (Yarbrough et al., 2011). The purpose of this non-experimental, survey research is to identify if crowdsource workers use any evaluation standards. This study is designed to determine how closely crowdsource product evaluations adhere to the JCSEE program evaluation standards. Specifically, this study seeks to determine: (a) what, if any, evaluation standards are being used by crowdsource organizations and their reviewers, and (b) to what extent these standards adhere to accepted evaluation standards used by professional evaluators in the evaluation discipline, precisely, the Utility, Feasibility, Propriety, and Accuracy domains of the JCSEE program evaluation standards.

Subject Recruitment

Recruitment of reviewers occurs on MTurk through a job board that identifies work opportunities. No pre-surveys or survey notices are necessary. However, in place of these activities will be requisition and contracting the use of workers to complete surveys and execute work orders. If a reviewer is interested in completing the survey for this dissertation, he or she must agree to contract with MTurk by clicking on the available opportunity. The reviewer will then complete the required work (i.e., complete the survey in its entirety) for a fee paid through MTurk by the researcher as indicated above.

Setting up a Crowdsourcing Project

Crowdsourced survey research will be conducted through a single privately held self-service crowdsourcing company; MTurk.com. To initiate services with the crowdsourcing company it requires filling out the forms found on the crowdsourcing company's website, these forms capture requester (Investigator) contact data, billing data, and brief instructions as to what is being requested. As shown in Appendix A, an online form to contract MTurk.com.

The crowdsourcing company will then ask for a hosting website (in this study, Qualtrics) where the survey is to be conducted and will ask the requester to upload any special announcements (Appendix B – reviewer agreement), or instructions for the reviewer prior to having the reviewer transferred to the hosting website for completion of the survey instrument.

Each time a requester sets up a project in MTurk, they must agree to MTurk Terms and Conditions. The Requester Terms and Conditions can be seen in Appendix C. After completion of gathering all background and billing data, the crowdsourcing company will then provide an invoice for services. This invoice will be paid prior to project launch using a credit card. Once paid the crowdsourcing company will inform the requester when reviewers will be directed to the studies site.

Recruitment of Reviewers

Recruitment of reviewers will occur with MTurk through a “Job Board” that identifies the opportunities provided by requesters. Each reviewer interested in completing the survey will then agree to contract with the MTurk by clicking on and linking with the available opportunity. The reviewer will then complete the required work

(complete the survey in its entirety) for a fee paid by the crowdsourcing company contracted.

When the reviewer completes the work sign up procedure he is also asked to agree to Terms and Conditions. Upon agreement of the terms and conditions the reviewer is free to continue and is directed to the survey by the recruiter.

Informed Consent Process

Participation in this study is part of a voluntary crowdsourcing work opportunity. Nevertheless, all study participants will be asked to indicate their consent in the study. MTurk workers must be over the age of 18, so they are free to provide their own consent. Because this study is an online Qualtrics survey, consent will be obtained using a consent screen. As Schmidt (1997) suggested, there will be a separate consent screen before respondents gain access to the survey (Appendix D). On the consent screen, potential participants will read the informed consent letter and be able to affirm their consent to participate in the study by clicking a button that states, "I agree to participate." As part of the informed consent process, participants will also be informed that they may abort the survey at any time or refuse to answer all or certain questions; however, payment for participation in the study requires completion of all survey questions.

Only the Co-Principal Investigator, Principal Investigator, and Mary Ramlow -office coordinator will have access to survey data. Data will be stored for five (5) years.

Research Procedure

Methods for Data Collection

Data will be collected via online questionnaires. Crowdsourcing workers will be asked to complete a 30-60 minute questionnaire online. Each subject will agree to

contract with MTurk via web based online self-serve crowdsourcing portals. The company being used for this research is MTurk.com. Upon agreeing to the crowdsourcing company terms and conditions and also completing the consent form provided at the survey website link; the reviewer will then participate in the completion of a survey instrument satisfying its requirements and will then be paid by the crowdsourcing company the reviewer contracted with.

Instrumentation

Instrumentation for this dissertation includes a 27-item survey developed by the researcher for the purpose of this study and a 11-item demographic questionnaire (Appendix E). The items on the survey represent each of the four domains of the JCSEE program evaluation standards (Yarbrough et al., 2011) that will be explored in this study (i.e., Utility, Feasibility, Propriety, and Accuracy), and assess how often reviewers considered each program evaluation standard when conducting product evaluations or reviews within the last 12 months. They are rated on the following 4-point Likert-type scale: 1 = *Never*, 2 = *Seldom*, 3 = *Often*, 4 = *Always*. When selecting *never* or *seldom*, participants will also be prompted to respond to an open-ended question that allows participants to explain their answers. Higher scores will indicate greater consideration of the standards when conducting product reviews. There is no cut-off score indicating adequate or inadequate consideration of the standards.

Items of the demographic questionnaire were designed to explore how many years participants have worked in the crowdsourcing business, their role in crowdsourcing, the segment of product review for which they most frequently get contracts, if they have formal training in product review, the type of training they have, their highest level of

education, their primary area of expertise, and if they have experience working with evaluation checklists. The survey ends with an open-ended response question that provides an opportunity for participants to make additional comments. High quality responses are assumed based on crowdsourcing worker incentives to succeed and advance within their organization as an evaluator or reviewer.

Location of Data Collection

Crowdsourced survey research will be conducted through one self-service web based crowdsourcing company; MTurk.com. There is no brick and mortar facility. To initiate services with MTurk it requires filling out the forms found on the company's website, these forms capture requester (Investigator) contact data, billing data, and brief instructions as to what is being requested.

Only the Co-Principal Investigator, Principal Investigator will have access to survey and interview data. Data will be stored for five (5) years on a secure server at WMU.

Duration of Study

Data collection is expected to take place as soon as HSIRB approval is obtained. Both research survey and interview participants will commit between 30 and 60 minutes of their time to the study. All data is expected to be collected within a two week time period or less.

Methodology

Design

A descriptive, cross-sectional survey design will be utilized for this study. As stated above, the purpose of this study is to explore if crowdsourcing workers, in particular,

Amazon MTurk product reviewers, use any evaluation standards, and if so, how closely these standards adhere to the JCSEE program evaluation standards.

Because the research questions of the study are non-experimental and descriptive in nature, it is inappropriate to formulate hypotheses, and issues concerning levels of precision and statistical power are irrelevant. Financial costs will be paid by the researcher to participants in the form of paid work (see Data Collection section) and Internet survey-hosting fees. Gaining access to the crowdsource workers requires use of an Internet-based mode of data collection. Furthermore, Internet survey methods yield major advantages over mail survey methods including (a) quicker speed of response, (b) lower costs for distribution of questionnaires, and (c) precision of data compilations (Matsuo et al., 2004).

Sample design refers to the design of the parameters (also called the *sampling frame*) used to ensure adequate coverage of a study's population (Hall, 2008). Details of the sample design for this study are discussed in the Sample section below, however, in general, aspects of sample design include:

5. Selecting a sampling frame
6. Defining the strata, if any, to be employed
7. Deciding whether the sample is to be a single-stage, clustered, or multi-stage design, and
8. Determining the sample size (Hall, 2008, p. 173).

Questionnaire design refers to considerations given to the development of the survey instrument that will be used to obtain data for the study (Hall, 2008). Items on

the questionnaire in this study consist of statements from the Utility, Feasibility, Propriety, and Accuracy domains of the JCSEE program evaluation standards.

Finally, operations planning refers to detailing the logistics of a cross-sectional study, such as making sure institutional procedures are clearly identified and followed, and scheduling data collection (Hall, 2008). In this study, operations planning includes meeting HSIRB requirements, obtaining survey-hosting services, uploading the survey, contracting with MTurk, allowing enough time for data collection, and securing the proper data analysis software.

In sum, use of the cross-sectional research design described above will allow the researcher to quickly and efficiently capture information from a large sample of product reviewers that increases the accuracy and legitimacy of the data being extracted (Kraut et al., 2004; Reips, 2002). However, as in all research, there are limitations inherent in cross-sectional survey research designs. Problems related to social desirability, or a phenomenon wherein participants respond to questions in a way they believe others will approve, have long been noted in the use of survey research (Johnson & Fendrich, 2002). These limitations are minimized by the controls and filtering placed on the subjects being surveyed. Specifically, use of MTurk.com will allow the researcher to filter unwanted subjects by establishing certain criteria potential subjects must meet in order to participate in the study. Additionally, both sample and non-response errors will be further minimized as MTurk ensures that an exact number of participants are recruited and sampled.

Analysis

Analysis of all survey data will be conducted according to generally accepted scientific standards. Closed-response items from the surveys will be analyzed in SAS 9.3 using descriptive statistics including measures of central tendency such as means, medians, and modes, as well as frequencies and percentages. Open response items will be analyzed using deductive and inductive coding procedures in MaxQDA 10. All findings will be presented using tables and figures where appropriate. Findings and conclusions for the research questions will then be based on examination and interpretation of the information arising from surveys. The table summarizes the total research methodology.

Summary of Research Methodology

Method	Non-experimental, cross-sectional survey research
Participants	Trained and qualified product reviewers at MTurk with more than one year of experience and a 90% approval rating from the United States.
Sampling Procedures	Probability sampling of entire MTurk population meeting inclusionary criteria.
Instrumentation	Web-based survey assessing how often reviewers considered each program evaluation standard when conducting product evaluations or reviews within the last 12 months.
Data Analysis	Descriptive statistics including means, medians, modes, frequencies, and percentages.

Dissemination

Findings from the study will be disseminated in a dissertation and possible publications from the dissertation. Findings may also be disseminated via presentations at conferences, informal meetings, or as part of future grant proposals.

Confidentiality of Data

Interviewees will sign and return the consent form (see Appendix E) and will be stored at the Western Michigan University Evaluation Centers Administration office electronic storage. Only the Co-Principal Investigator and the Principal Investigator will have access to survey and interview data.

Risks and Costs to and Protection for Subjects

There are no known risks to human subjects who participate or learn about the study. All data is expected to be collected within a two week time period. Participants receive a small fee for completing the survey. Interviewees will sign and return the consent form (see Appendix E), which will be stored on a secured drive at WMU. Only the Co-Principal Investigator, Principal Investigator, and office coordinator will have access to the folder. Data will be stored for five (5) years. Data Collection and interviews are confidential. Identifying information remains with the crowdsourcing company. The survey is encrypted through Qualtrics.

Benefits of Research

The literature indicates that crowdsourcing is used for product evaluation. “The basic assumption is that the crowd can bring interesting, non-trivial, and non-overlapping information, insights, or skills, which when harnessed through appropriate aggregation and selection mechanisms, can add to a solutions quality” (Davis, 2012, p.94.)The quality of the evaluation standards used in crowdsourced evaluation has not been assessed. It is not understood if crowdsourcing organizations use cogent standards to conduct evaluations. Further, it has not been determined if crowdsourced evaluations adhere to generally accepted evaluation standards such as those created by The JCSEE. As a result,

the benefits of the research can be seen in knowledge generation about and potential improvement in the way product evaluations are done in the future.

More specifically, information from this research may provide insight to inform evaluators in a manner to inspire creation and integration of acceptable standards and evaluation processes that can bridge gaps that may exist between crowdsourced evaluations and acceptable evaluation standards. The information from this study may also inform and contribute to the evolution and adaptation of the Program Evaluation Standards in such a manner that they may be useful in contexts other than educational program evaluation.

Literature has not revealed the details of the standards and methods these crowdsource-workers use and what their attitudes toward formal evaluation standards and methods are. It will be important for people in the evaluation profession to know what product evaluation standards crowdsource workers are currently using and what their attitudes toward using more formal standards may be, so that the profession can advance.

In addition to determining what standards are being used and how they adhere to the program Evaluation Standards, it is also important to understand how crowdsource organizations and workers feel about accepting and working with these evaluation standards. This is essential, as it may provide new information as to how evaluation standards may be perceived and accepted. This may have opportunities for future workforce development initiatives.

Appendix C

Survey of Evaluation Standards Questionnaire

Survey of Evaluation Standards Questionnaire

For product reviews that you have conducted within the last 12 months, please indicate how frequently the following applied. Indicate your choice by selecting the number that most adequately represents your response on the following scale:

1 = Never, 2 = Seldom, 3 = Often, 4 = Always

		Never	Seldom	Often	Always
1	I completed reviews of products without prior experience using it	1	2	3	4
2	I completed reviews with an understanding of whom I was completing them for.	1	2	3	4
3	I had a clear understanding of the product reviews purpose	1	2	3	4
4	I had complete understanding of the requestor's values underpinning the review.	1	2	3	4
5	I was informed of the needs of users of the reviews.	1	2	3	4
6	My reviews encouraged use	1	2	3	4
7	My review submissions were on time.	1	2	3	4
8	My reviews were misused.	1	2	3	4

IF: Never or Seldom

Then: Why? Source: Open Ended Question

1	I used project management techniques to complete reviews.	1	2	3	4
2	I used practical procedures	1	2	3	4
3	My reviews recognized and balanced cultural and political interests of individuals and consumers.	1	2	3	4
4	I used resources effectively and efficiently.	1	2	3	4

IF: Never or Seldom

Then: Why? Source: Open Ended Question

1	My reviews were responsive to potential consumers.	1	2	3	4
2	I was involved in negotiations with requesters about the reviews.	1	2	3	4
3	My reviews considered consumer rights and safety.	1	2	3	4
4	My reviews were clear and fair.	1	2	3	4
5	My reviews disclosed all relevant information.	1	2	3	4
6	My reviews disclosed conflicts of interest.	1	2	3	4
7	My reviews expended out of pocket resources.	1	2	3	4

IF: Never or Seldom

Then: Why? Source: Open Ended Question

1	My review conclusions are were justified	1	2	3	4
2	My review conclusions were valid	1	2	3	4
3	My reviews were reliable	1	2	3	4
4	I thoroughly documented reviews with great detail.	1	2	3	4
5	I used a system of information management.	1	2	3	4
6	I used more technical review designs, when appropriate for the specific review.	1	2	3	4
7	I concisely documented review reasoning, findings and conclusions.	1	2	3	4
8	My review techniques guard against misinterpretations, biases, distortions, and errors.	1	2	3	4

IF: Never or Seldom

Then: Why? Source: Open Ended Question

Please tell me a little bit about yourself.

1. How many years have you worked in crowdsourced evaluation?

Drop down:

1-5

6-10

11-15

16-20

2. What is your role in crowdsourcing? Select ALL that apply.

- Coordinator
- Manager
- Director
- Reviewer
- Other, please explain

3. What product segment do you get contracts for most often?

- Consumer Packaged Goods
- Medical
- Electronics
- Automotive
- Food & Beverage
- Industrial
- Other, please explain

4. Do you have formal training for conducting product reviews?

- Yes
- No

5. If yes, what type of training do you have? Select all that apply.

- University program
- On the job training (e.g., professional development)
- Off the job training (e.g., a certificate program)
- Other, please specify:

6. What is the HIGHEST educational degree you have obtained?

- Primary education
- Secondary school (e.g., high school)
- Professional training (e.g., certificate, apprenticeship)
- Undergraduate degree (e.g., baccalaureate)
- Graduate degree (e.g., MA, MBA)
- Postgraduate degree (doctorate)
- Other, please specify

7. Please indicate your primary area(s) of expertise or discipline

- Product Evaluation
- Product Reviews
- Other (Explain)

8. Approximately what is the range of pay in USD per review?

- a. .25-.50
- b. .51-.75
- c. .76-1.00
- d. 1.00-1.50
- e. Other (Explain)

9. Do you perform product reviews

- a. Part Time
- b. Full Time

10. Is there any other information that you would like to share?

11. What's your gender?

- a. Male
- b. Female

Appendix D
Contract with MTurk

Contract with MTurk

[← go to MTurk.com](#)

[Alex Manga](#) | [Account Settings](#) | [Sign Out](#) | [Help](#)

[Home](#) [Create](#) [Manage](#) [Developer](#) [Help](#)

User Registration

Personal Information

Please select an existing address or enter a new one below.

Alex Manga
1616 E CHICAGO RD
STURGIS, MI 49091-1973
United States
Phone: 269-69-7807

Enter a new address

Company Name*

Address Line 1*

Address Line 2

Address Line 3

City*

State, Province or Region*

ZIP or Postal Code*

Country* [\(Why?\)](#)

Phone number*

See the [Mechanical Turk Privacy Notice](#) for how we treat your information.
* indicates required information.

Mechanical Turk Participation Agreement

Amazon Mechanical Turk Participation Agreement

Last updated: November 1, 2012

Welcome to the Amazon Mechanical Turk services platform.

BY REGISTERING FOR AND USING THE SITE, YOU CERTIFY THAT (1) YOU ARE AT LEAST 18 YEARS OLD; (2) YOU HAVE THE AUTHORITY TO ENTER INTO THIS AGREEMENT AND BIND YOURSELF OR THE COMPANY YOU REPRESENT; (3) YOU AUTHORIZE THE ELECTRONIC TRANSFER OF FUNDS TO YOUR BANK ACCOUNT IN ACCORDANCE WITH SECTION 4 OF THIS PARTICIPATION AGREEMENT; AND (4) YOU AGREE TO BE BOUND BY ALL TERMS AND CONDITIONS OF THIS AGREEMENT, INCLUDING THE TERMS AND CONDITIONS OF THE PAYMENT SERVICE DESCRIBED IN SECTION 4 AND ALL APPLICABLE POLICIES, PROCEDURES AND GUIDELINES. This Participation Agreement (the "Agreement") is between you

Please check the box below:

I agree to the following:

- * I accept the User Agreement and Privacy Policy above.
- * I am over 18 years of age.
- * I will comply with all applicable laws, including without limitation tax and filing requirements.

Appendix E
Reviewer Agreement

Reviewer Agreement

Available at: <https://www.mturk.com/mturk/welcome>

Amazon Mechanical Turk - Welcome

Page 1 of 1



[Your Account](#) | [HITS](#) | [Qualifications](#)
Introduction | [Dashboard](#) | [Status](#) | [Account Settings](#)

Already have a worker account? [Sign In](#)

Working on HITS

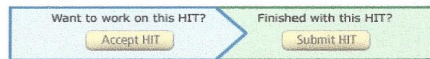
Step 1: Find work

Search or browse through the Human Intelligence Tasks (HITS) and click on the one that interests you.


Evaluate Search Results		View a HIT in this group
Requester: Powerset	HIT Expiration Date: April 1, 2008 Time Allotted: 10 minutes	Reward: \$0.02 HITS Available: 1114
Select the Best Category For a Product		View a HIT in this group
Requester: Channel Intelligence	HIT Expiration Date: April 24, 2008 Time Allotted: 10 minutes	Reward: \$0.01 HITS Available: 425
Expedited Podcast Transcription		View a HIT in this group
Requester: Casting Words	HIT Expiration Date: March 30, 2008 Time Allotted: 4 hrs 30 minutes	Reward: \$4.95 HITS Available: 2

Step 2: Work on your HIT

Accept the HIT and follow the instructions. When you're done, submit your work.



Step 3: Get paid for your work

 After the requester approves your work, money is deposited into your Amazon Payments account.

Examples

[Find HITS Now](#)

Here are just a few examples of HITS that workers have completed on Mechanical Turk.

- Select the correct spelling for these search terms
- Is this website suitable for a general audience?
- Find the item number for the product in this image
- Rate the search results for these keywords
- Are these two products the same?
- Choose the appropriate category for products
- Categorize the tone of this article
- Translate a paragraph from English to French

Currently 359,207 HITS available.

[Find HITS Now](#)

[FAQ](#) | [Contact Us](#) | [Careers at Mechanical Turk](#) | [Developers](#) | [Press](#) | [Policies](#) | [Blog](#)
©2005-2014 Amazon.com, Inc. or its Affiliates

An [amazon.com](#) company

Appendix F

Amazon Mechanical Turk Participation Agreement

Amazon Mechanical Turk Participation Agreement

Available at: <https://www.mturk.com/mturk/welcome>

Last updated: November 1, 2012

Welcome to the Amazon Mechanical Turk services platform.

BY REGISTERING FOR AND USING THE SITE, YOU CERTIFY THAT (1) YOU ARE AT LEAST 18 YEARS OLD; (2) YOU HAVE THE AUTHORITY TO ENTER INTO THIS AGREEMENT AND BIND YOURSELF OR THE COMPANY YOU REPRESENT; (3) YOU AUTHORIZE THE ELECTRONIC TRANSFER OF FUNDS TO YOUR BANK ACCOUNT IN ACCORDANCE WITH SECTION 4 OF THIS PARTICIPATION AGREEMENT; AND (4) YOU AGREE TO BE BOUND BY ALL TERMS AND CONDITIONS OF THIS AGREEMENT, INCLUDING THE TERMS AND CONDITIONS OF THE PAYMENT SERVICE DESCRIBED IN SECTION 4 AND ALL APPLICABLE POLICIES, PROCEDURES AND GUIDELINES. This Participation Agreement (the "**Agreement**") is between you and Amazon Mechanical Turk (as defined below) and governs your and Amazon Mechanical Turk's respective rights and obligations with respect to your offering for sale, selling, requesting, purchasing, and/or providing Services (defined below) on or through the Site (as defined below).

For purposes of this Agreement, (a) "**Amazon Mechanical Turk**", "**we**", "**us**" or "**our**" means Amazon Mechanical Turk, Inc. a Delaware Corporation, (b) "**Site**" means the Amazon Mechanical Turk web site located at mturk.amazon.com, requester.mturk.com, www.mturk.com and any successor website thereto, including all services provided by us to you through the service platform on the Site, (c) "**Services**" means any service that you sell, offer to sell, request, purchase, and/or provide on or through the Site, (d) "**Affiliate**" means any entity controlled by, in control of, or under common control with Amazon Mechanical Turk, (e) "**Requester**" means you, if you use the Site to request that a Provider perform Services, (f) "**Provider**" means you, if you use the Site to perform Services for a Requester, (g) "**Amazon Account**" means any customer account that you have established with a website owned or controlled by Amazon or its Affiliates, or operated by Amazon or its Affiliates on behalf of third parties, including without limitation those websites currently located at <http://www.amazon.com>, <http://www.amazon.co.uk>, <http://www.amazon.de>, <http://www.amazon.fr>, <http://www.amazon.ca>, <http://www.amazon.co.jp> and <http://www.joyo.com>, and any successor or replacement websites.

This Agreement consists of the terms and conditions set forth in this document together with all applicable policies, procedures and/or guidelines that appear on the Site from time to time (collectively, the "**Policies**" which are hereby incorporated by this reference into, and made part of, this Agreement). Amazon Mechanical Turk reserves the right to change any of the terms and conditions contained in this Agreement and/or any Policies governing the Site, at any time, in its sole discretion. Any changes will be effective upon posting of the Agreement or Policies on the Site and may be made without any other notice of any kind. You are at all times responsible for reading and understanding each version of this Agreement and the Policies. YOUR CONTINUED USE OF THE SITE FOLLOWING AMAZON MECHANICAL TURK'S POSTING OF ANY CHANGES WILL CONSTITUTE YOUR ACCEPTANCE OF SUCH CHANGES. IF YOU DO NOT AGREE TO ANY CHANGES TO THIS AGREEMENT (INCLUDING TO ANY OF THE POLICIES INCORPORATED HEREIN), DO NOT CONTINUE TO USE THE SITE.

1. **Registration.**

- a. **Registration.** When you register with the Site, you will be asked to provide us with, at a minimum, your name, a valid email address, your phone number, and your physical address. Providers may also be asked to provide certain tax information at registration or afterwards. You agree to provide us with true and accurate information, and to update that information to the extent it changes in any way. When registering or updating your information, you will not impersonate any person or use a name that you are not legally authorized to use.

- You may register with the Site either by (i) using your existing Amazon Account or (ii) creating a new Amazon Account. If you do not have an existing Amazon Account at the time you register with the Site, an Amazon Account on the Amazon.com website located at <http://www.amazon.com> (hereinafter, "**Amazon.com**") will be automatically and concurrently established in your name with the same e-mail address and password you provide to us. Amazon Accounts used in conjunction with the Site are governed by the [Conditions of Use](#) and [Privacy Notice](#) applicable to Amazon.com, as well as the [Amazon Mechanical Turk Privacy Notice](#). You may not use multiple Amazon Accounts to register with Mechanical Turk. Your Amazon Account username must not suggest affiliation with Amazon, Amazon Mechanical Turk, or any third party unless that third party specifically gave you permission to do so.

 - **b. Passwords and Account Use.** You are solely responsible for maintaining the secrecy and security of your password. You may not disclose your password to any third party (other than third parties authorized by you to use your account) and are solely responsible for any use of or action taken under your password on the Site. If your password is compromised, you must change your password. You may not permit any other person to perform Services as Provider using your Amazon Account. Additionally, if you are using the Site as a Provider, you may not use different Amazon Accounts to perform Services.
2. **Amazon Mechanical Turk's Role.** Amazon Mechanical Turk provides a venue for third-party Requesters and third-party Providers to enter into and complete transactions. Amazon Mechanical Turk and its Affiliates are not involved in the transactions between Requesters and Providers. As a result, we have no control over the quality, safety or legality of the Services, the ability of Providers to provide the Services to Requesters' satisfaction, or the ability of Requesters to pay for Services. We are not responsible for the actions of any Requester or Provider. We do not conduct any screening or other verification with respect to Requesters or Providers, nor do we provide any recommendations. As a Requester or a Provider, you use the Site at your own risk.
3. **Your Use of the Site**
- **a. Requesters in General.** Upon completion of Services to Requesters' reasonable satisfaction, Requesters must pay Providers for their Services. As a Requester, you agree that upon your approval of the Services performed by a Provider, payment will be remitted to the Provider automatically (as described in Section 4 below). After you have approved the applicable Services, you are not entitled to any refund of your payment for such Services. If a Requester is not reasonably satisfied with the Services, the Requester may reject the Services. As a Requester, you will be charged a fee for your use of Amazon Mechanical Turk in connection with each request for Services. Please review the applicable Amazon Mechanical Turk Fees contained in the Policies for all applicable fees associated with your use of the Site pursuant to this Agreement. All fees are in U.S. dollars unless stated otherwise. The Amazon Mechanical Turk Fees may vary in the future. You agree to pay the amounts set forth in the Amazon Mechanical Turk Fees from time to time on the terms set forth herein and therein, and to check the fees and terms each time you use the Site. You acknowledge that, while Providers are agreeing to perform Services for you as independent contractors and not employees, repeated and frequent performance of Services by the same Provider on your

behalf could result in reclassification of that employment status. If you have any questions about your obligations to comply with local laws and regulations pursuant to Section 6, you should seek independent legal advice. To the extent you receive any contact or personal information regarding any Provider who has performed Services for you, such information may only be used as necessary for you to comply with applicable laws and for no other purpose whatsoever. Further, you agree that you will only accept work product from Providers that has been submitted through the Site.

- **b. Providers in General.** You may only register once with Mechanical Turk as a Provider. Providers may perform Services for any Requester in accordance with the specifications submitted by the Requester. However, if the Services do not meet the Requester's reasonable satisfaction, the Requester may reject the Services and repost the specific request. As a Provider, the Requester for whom you provide Services is your client, and as such, you agree that the work product of any Services you perform is deemed a "work made for hire" for the benefit of the Requester, and all ownership rights, including worldwide intellectual property rights, will vest with the Requester immediately upon your performance of the Service. To the extent any such rights do not vest in Requester under applicable law, you hereby assign or exclusively grant (without the right to any compensation) all right, title and interest, including all intellectual property rights, to such work product to Requester. As a Provider you are performing Services for a Requester in your personal capacity as an independent contractor and not as an employee of the Requester. You specifically acknowledge and agree to the following: (i) you will not use robots, scripts or other automated methods to complete the Services; (ii) you will submit all work product through the Site only, and not directly to a Requester; (iii) you will provide Requesters for whom you perform Services with any information reasonably requested by them in connection your performance of such Services; (iv) you are responsible for, and have and will, comply with all applicable laws and registration requirements, including those applicable to independent contractors and maximum working hours regulations; (v) this Agreement does not create an association, joint venture, partnership or franchise, employer/employee relationship between Providers and Requesters, or Providers and Amazon Mechanical Turk; (vi) you will not represent yourself as an employee or agent of a Requester or Amazon Mechanical Turk; (vii) you will not be entitled to any of the benefits that a Requester or Amazon Mechanical Turk may make available to its employees, such as vacation pay, sick leave, insurance programs, including group health insurance or retirement benefits; (viii) you are not eligible to recover worker's compensation benefits in the event of injury; and (ix) if you are not a resident or citizen of the United States, all Services that you use the Site to perform for a Requester will be performed outside of the United States. If you have any questions about your obligations to comply with local laws and regulations pursuant to Section 5, you should seek independent legal advice.

- **c. Listing and Promotions Generally.** As a Requester or Provider, you may not sell, offer for sale, request, purchase, or provide any Service that violates applicable law or is prohibited by the Policies. Notwithstanding any provision of this Agreement, Amazon Mechanical Turk will have the right, in its sole discretion, to determine the content, appearance, design, functionality and all other aspects of the Site (including the right to re-design, modify, remove and alter the content, appearance, design, navigation, functionality, and other aspects of the Site and/or any page thereof and any element, aspect, portion or feature thereof, from time to time).

- d. **Information and Feedback.** You must supply accurate and complete information for all Services in accordance with our data requirements, as may be designated by us from time to time, including in the Policies. You recognize and agree that Amazon Mechanical Turk will implement mechanisms allowing us and others to track your requests for, or your performance of, Services and rate your performance as a Requester or Provider, and Amazon Mechanical Turk reserves the right to collect feedback regarding your performance and to post such feedback on the Site. You may not take any actions that may undermine the integrity of the feedback system. You agree that submission of any information, feedback, content, data or other materials (collectively, "**Materials**") is at your own risk, and that none of Amazon Mechanical Turk, its Affiliates, Requesters or Providers has any obligations (including without limitation obligations of confidentiality) with respect to such Materials. You represent and warrant that you have all rights necessary to submit the Materials. You hereby grant to Amazon Mechanical Turk and its Affiliates a royalty-free, non-exclusive, worldwide, perpetual, irrevocable right and license to use, reproduce, perform, display, distribute, adapt, modify, re-format, create derivative works of, and otherwise commercially or non-commercially exploit in any manner, any and all Materials, and to sublicense the foregoing rights, in connection with the operation and maintenance of the Site. For avoidance of doubt, if you are a Requester, if you create any tests, specifications, criteria lists or other programs for use on the Site to evaluate or otherwise select Providers they will be considered Materials for purposes of this Agreement and may be used and/or referenced by us or other Requesters and Providers pursuant to the license granted above.

 - e. **Customer Service.** Amazon Mechanical Turk will be responsible for and will have sole discretion regarding all customer service issues relating to use of the Site and its features.

 - f. **Disputes between Requesters and Providers.** Your use of the Site is at your own risk. Because Amazon Mechanical Turk is not involved in the actual transaction between Providers and Requesters, Amazon Mechanical Turk will not be involved in resolving any disputes between participants related to or arising out of the Services or any transaction.
4. **Payment Service.** Amazon Mechanical Turk or its Affiliates will process all payments made by Requesters to Providers (the "**Payment Service**"). Requester payments made through the Payment Service are received by Amazon Mechanical Turk or its Affiliates on behalf of Providers, and may be disbursed only in accordance with the terms outlined below.
- a. **Prepaid HITs.** Each Requester must prepay for work they intend on acquiring through the Services by purchasing Mechanical Turk Prepaid HITs ("**Prepaid HITs**") from Amazon Payments, Inc. Prepaid HITs are subject to the [Mechanical Turk Prepaid HITs Terms and Conditions](#). Prepaid HITs are maintained in a single Prepaid HIT account for you solely for use of the Services. The amount of Prepaid HITs purchased must be at least equal to the total amount that will be owed to Providers upon completion and acceptance of the Services and any amounts payable to Amazon Mechanical Turk in connection with Requester's use of the Site. If the Prepaid HITs are purchased with proceeds from a bank account, the Prepaid HITs may not be available for use for up to four (4) days before such funds are available for disbursement to a Provider's Payment Account (defined below). After Requester's acceptance

of the Services, the Payment Service will debit the amount owed to each Provider from the Requester's Payment Account, and credit each Provider's Payment Account that amount.

- **b. Disbursement of Funds to Providers.** When Providers register with the Site, a payment account ("**Payment Account**") will automatically be established in conjunction with their registration. Providers may disburse funds from their Payment Account by the following methods, at their option: (i) to an ACH-Enabled Bank Account in U.S. dollars; (ii) or by converting such funds to a credit that is held for the benefit of Provider in an [Amazon.com gift certificate account](#). For select countries, Amazon Mechanical Turk may enable Providers to request disbursements through physical checks in U.S. or local currency. Check disbursements may be subject to additional fees, registration and documentary requirements. See our [FAQs](#) for more information. Funds will only be disbursed in compliance with applicable laws and regulations, including without limitation the United States Patriot Act and the regulations of the Office of Foreign Assets Control. Providers may not share a bank account. We reserve the right to cancel your Payment Account for any reason.

- **c. Authorizations for ACH-Enabled Bank Account.** If you are a Provider, you hereby authorize Amazon Mechanical Turk and its Affiliates, and any third party service providers or agents acting on their behalf, to debit or credit your ACH-Enabled Bank Account (including by creating a paper draft or an electronic funds transfer) and/or your Payment Account, as applicable, (i) to transfer, disburse or process other payment transactions associated with the Services; and (ii) to settle payment for any fees that may be charged under this Agreement. In the event there is an error in the processing of any transaction described above, you authorize us to initiate debit or credit entries to your ACH-Enabled Bank Account or your Payment Account, as applicable, to correct such error, provided that any such correction is made in accordance with applicable laws and regulations, and to make any inquiries we consider necessary to validate the error, which may include ordering a credit report, performing credit checks, or verifying the information you provide against third party databases. If we are unable to debit any ACH-Enabled Bank Account you select for any reason, you authorize us to resubmit the debit, plus any applicable fees, to any other ACH-Enabled Bank Account you have on file with us (or, in the case of any fees that are owed under this Agreement, to deduct such amounts from the funds in your Payment Account). Your authorizations will remain in full force and effect until we receive written notification from you of any termination. Any termination will become effective as soon as we have had a reasonable amount of time to act on it, but in any event not later than thirty (30) days after written notice of termination is received by us in accordance with Section 12(e).

- **d. Restrictions and Limitations.** We reserve the right to terminate or suspend any Payment Account, or to delay the availability of any Prepaid HITs, transfer or disbursement of any amounts, in each case for any reason in our sole discretion, including, without limitation, if we believe that a Requester or Provider is in violation of this Agreement. We reserve the right to restrict the transfer to Providers of any amounts held in a Requester's Prepaid HIT account for such time as we reasonable deem necessary to protect us or others: (a) if we are subject to financial risk, (b) if Provider has violated any term of this Agreement or the Policies, (c) if any dispute exists involving Provider's Payment Account or involving the Services provided by Provider, or (d) in connection with fraudulent, abusive or unlawful activities as determined by us. Further, we reserve the right to restrict the transfer to Providers of any amounts held in a Requester's Prepaid HIT account for up to ten (10) Business Days (as defined below) following Requester's acceptance of the Services provided by Provider. Other than a credit to

a Provider's Payment Account for Services rendered by such Provider, amounts held in Payment Accounts cannot be transferred to other Requesters or Providers. If Amazon Mechanical Turk terminates this Agreement because you have violated the Policies then (i) any Services that have been completed by Providers but not yet accepted by you will be deemed accepted and the applicable payments will be remitted to the Providers and deducted from your Prepaid HITS balance and (ii) your remaining Prepaid HITS balance (if any) will become the property of Amazon Mechanical Turk.

- e. **Our Liability.** We (and our Affiliates) act only in the capacity of a payment processor in facilitating the transactions between Requesters and Providers, and are not otherwise involved in the actual transactions. We will only be responsible for initiating purchases of Prepaid HITs and the transfers or disbursements at the direction of Requesters and Providers. We will be entitled to rely on the instructions of Requesters and Providers without any further inquiry or liability whatsoever. We will not be liable if we are not able to complete a transaction for any reason, including, but not limited to,
 - If any system or equipment was not working properly and you knew or had been advised about the breakdown before you initiated the transaction;
 - If you do not have enough available funds in your Prepaid HIT account or in your Payment Account to complete the applicable transaction, or if the transfer would cause you to exceed any applicable transfer limit with respect to your ACH-Enabled Bank Account;
 - If circumstances beyond our control (such as, but not limited to, power outages, fire, flood, mechanical or systems failure) prevent the proper execution of the transaction, despite reasonable precautions we have taken;
 - If your transaction is intercepted by legal process or other encumbrances restricting transfer, or your participation in the Site has been terminated or suspended for security purposes;
 - If we are unable to confirm your identity or have reason to believe that the transfer requested is unauthorized; or
 - If you have not provided us with correct, current and complete payment information.
- f. **Statements and Account Balances.** We will send an e-mail confirmation to you after you purchase any Prepaid HITs, make payments to a Provider or other payment transaction occurs with respect to a Payment Account. In addition, you may access your transaction information (your "Activity History") online in the "Your Account," and "View Transaction History" (or equivalent) areas of the Site. You may access this feature only with a browser that is compatible with the Service, including any security features that are part of the Service. Interest will not be paid on Prepaid HITs or any amounts held in Payment Accounts. If no transfer, disbursement or other payment transaction occurs with respect to your Payment Account for at least two (2) years and six (6) months, consecutively, the balance in your Payment Account will be automatically converted into an Amazon.com gift certificate and sent electronically to your then-current e-mail address associated with your Payment Account.
- g. **Transaction Errors.** If you believe that any payment transaction initiated by us (or our agent) is erroneous, or if you need more information about any such transaction, you should contact us as soon as possible.

5. Compliance with Laws.

- a. **Taxes.** You agree that it is your responsibility to determine any and all taxes and duties, including without limitation, sales, use, transfer, value added, withholding and other taxes and/or duties assessed, incurred or required to be collected, paid or withheld for any reason in connection with any request for, or performance of Services, or your use of the Site, or otherwise in connection with any action, inaction or omission of you or any of affiliate of yours, or any of your or their respective employees, agents, contractors or representatives ("**Taxes**") and to collect, withhold, report, and remit correct taxes to the appropriate tax authority, and to otherwise be responsible for the collection and payment of any and all Taxes. YOU ALSO AGREE THAT AMAZON MECHANICAL TURK AND ITS AFFILIATES ARE NOT OBLIGATED TO DETERMINE WHETHER TAXES APPLY AND ARE NOT RESPONSIBLE TO COLLECT, REPORT, OR REMIT ANY TAXES ARISING FROM ANY TRANSACTION.

- b. **Registrations** You agree that is your responsibility to determine whether and to what extent any permits, registrations, authorization or filings (including without limitation with respect to the transfer of technology) are required by any governmental agency in any jurisdiction in which you have requested or are performing Services ("**Permits**"). YOU ALSO AGREE THAT AMAZON MECHANICAL TURK AND ITS AFFILIATES ARE NOT OBLIGATED TO DETERMINE WHETHER ANY SUCH PERMITS APPLY TO ANY TRANSACTION.

- c. **Compliance with Laws;** The Site may be used only for lawful purposes and in a lawful manner. You may not use the Site in any manner that violates any applicable law or governmental regulation. In addition to your obligations with respect to Taxes and Permits above, you agree to comply with all applicable laws, statutes, and regulations of any jurisdiction in which you request or perform Services.

- d. **Investigation.** Amazon Mechanical Turk has the right, but not the obligation, to monitor any activity, content and Materials associated with the Site. Amazon Mechanical Turk may investigate any reported violation of its Policies or complaints and take any action that it deems appropriate.

6. Disclosure of Information; Confidentiality; Privacy

- a. **Our Use of Data and Communications.** Our Privacy Notice and this Agreement describe our collection, use, and disclosure of information associated with the Site, including how we handle personal information. In addition to the disclosures described in our Privacy Notice, we may disclose to Requesters your name, address, data on HITs you have completed, and Provider Tax Information. "**Provider Tax Information**" means tax identification information of Providers, such as a Social Security Number or Employer Identification Number. You hereby consent to our use and disclosure of Provider Tax Information and other data as described in this Section 6 and our Privacy Notice.
- b. **Your Use of Data and Communications.** You may use information or other data acquired from your use of the Site solely to the extent necessary for you to use the Site and for no

other purpose, including but not limited to, for purposes of solicitation, advertising, marketing, unsolicited e-mail or spamming, harassment, invasion of privacy, or otherwise objectionable conduct.

- c. **Press Releases and Public Disclosures.** You may generally publicize your use of the Site, however you may not issue any press release with respect to Amazon Mechanical Turk or the Site, without Amazon Mechanical Turk's express prior written consent.
7. **No Warranties.** THE SITE, THE PAYMENT SERVICE AND THE SITE SERVICES ARE PROVIDED ON AN "AS IS" BASIS. TO THE FULLEST EXTENT PERMITTED BY APPLICABLE LAW, AMAZON MECHANICAL TURK MAKES NO REPRESENTATIONS OR WARRANTIES OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING WITHOUT LIMITATION:
- a. ANY IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, TITLE, OR NON-INFRINGEMENT;
 - b. THAT THE SITE, THE PAYMENT SERVICE OR THE SITE SERVICES WILL MEET YOUR REQUIREMENTS, WILL ALWAYS BE AVAILABLE, ACCESSIBLE, UNINTERRUPTED, TIMELY, SECURE, OPERATE WITHOUT ERROR, OR WILL CONTAIN ANY PARTICULAR FEATURES OR FUNCTIONALITY;
 - c. THAT THE INFORMATION, CONTENT, OR MATERIALS INCLUDED ON THE SITE WILL BE AS REPRESENTED BY REQUESTERS OR PROVIDERS, THAT THE SERVICES ARE LAWFUL, OR THAT REQUESTERS OR PROVIDERS WILL PERFORM AS PROMISED; OR
 - d. ANY IMPLIED WARRANTY ARISING FROM COURSE OF DEALING OR USAGE OF TRADE.
8. **General Release.** BECAUSE AMAZON MECHANICAL TURK IS NOT INVOLVED IN TRANSACTIONS BETWEEN REQUESTERS AND PROVIDERS OR OTHER PARTICIPANT DEALINGS, YOU HEREBY RELEASE AMAZON MECHANICAL TURK AND ITS AFFILIATES (AND THEIR RESPECTIVE EMPLOYEES, DIRECTORS, AGENTS AND REPRESENTATIVES) FROM ANY AND ALL CLAIMS, DEMANDS, AND DAMAGES (ACTUAL, CONSEQUENTIAL AND OTHERWISE) OF EVERY KIND AND NATURE, KNOWN AND UNKNOWN, SUSPECTED AND UNSUSPECTED, DISCLOSED AND UNDISCLOSED, ARISING OUT OF OR IN ANY WAY CONNECTED WITH ANY DISPUTE BETWEEN ONE OR MORE REQUESTERS, PROVIDERS, OR OTHER USERS OF THE SITE.

9. **Indemnity; Limitation of Liability.**
- a. **Indemnity and Defense.** You will indemnify and hold harmless Amazon Mechanical Turk and its Affiliates (and their respective employees, directors, agents and representatives) from and against any and all claims, costs, losses, damages, judgments, penalties, interest and expenses (including reasonable attorneys' fees) arising out of any claim, action, audit, investigation, inquiry or other proceeding instituted by a person or entity ("**Claim**") that arises out of or relates to: (i) any actual or alleged breach of your representations, warranties, or obligations set forth in this Agreement; (ii) your Services and any Materials, including any actual or alleged infringement of any intellectual property or proprietary rights by any of your Services or Materials; and/or (iii) your failure to comply with any applicable laws and regulations in connection with your use of the Site.

 - b. **Limitation of Liability.** TO THE FULLEST EXTENT PERMITTED BY APPLICABLE LAW, AMAZON MECHANICAL TURK AND ITS AFFILIATES WILL NOT BE LIABLE FOR ANY INDIRECT, INCIDENTAL, PUNITIVE OR CONSEQUENTIAL DAMAGES ARISING OUT OF OR IN CONNECTION WITH THIS AGREEMENT, THE SITE, THE PAYMENT SERVICE, THE SITE SERVICES, THE INABILITY TO USE THE SITE SERVICES, OR ANY SERVICES PURCHASED OR OBTAINED OR MESSAGES RECEIVED OR TRANSACTIONS ENTERED INTO THROUGH THE SITE. TO THE FULLEST EXTENT PERMITTED BY APPLICABLE LAW, IN NO EVENT WILL AMAZON MECHANICAL TURK'S OR ITS AFFILIATES' AGGREGATE LIABILITY ARISING OUT OF OR IN CONNECTION WITH THIS AGREEMENT OR THE TRANSACTIONS CONTEMPLATED HEREBY, WHETHER IN CONTRACT, TORT (INCLUDING NEGLIGENCE, PRODUCT LIABILITY OR OTHER THEORY), WARRANTY OR OTHERWISE, EXCEED THE AMOUNT OF FEES EARNED BY AMAZON MECHANICAL TURK IN CONNECTION WITH YOUR REQUEST FOR, OR YOUR PERFORMANCE OF, SERVICES DURING THE TWELVE (12) MONTH PERIOD IMMEDIATELY PRECEDING THE EVENT GIVING RISE TO THE CLAIM FOR LIABILITY.
10. **Applicable Law; Arbitration.** The Site is arranged, sponsored, and managed by Amazon Mechanical Turk in the state of Washington, USA. The laws of the state of Washington govern this Agreement and all of its terms and conditions, without giving effect to any principles of conflicts of laws. You agree that any action at law or in equity arising out of or relating to these terms and conditions shall be submitted to confidential arbitration in Seattle, Washington, except that, to the extent you have in any manner violated or threatened to violate Amazon Mechanical Turk's intellectual property rights, Amazon Mechanical Turk may seek injunctive or other appropriate relief in any state or federal court in the state of Washington, and you consent to exclusive jurisdiction and venue in such courts. Arbitration under this agreement shall be conducted under the rules then prevailing of the American Arbitration Association. The arbitrator's award shall be binding and may be entered as a judgment in any court of competent jurisdiction. To the fullest extent permitted by applicable law, no arbitration under this Agreement shall be joined to an arbitration involving any other party subject to this Agreement, whether through class arbitration proceedings or otherwise.
11. **Termination.** You may at any time elect to stop using the Site, provided that in discontinuing any Site activities, you must use Amazon Mechanical Turk's standard functionality and further must abide by all applicable Amazon Mechanical Turk Policies, Procedures and Guidelines. Amazon

Mechanical Turk, in its sole discretion, may terminate this Agreement, suspend access to the Site, or remove any Service listings immediately without notice for any reason.

12. General Provisions

- a. **Entire Agreement.** This Agreement and the general terms and conditions of the Site, including the Policies, constitute the entire agreement of the parties with respect to the subject matter hereof, and supersede and cancel all prior and contemporaneous agreements, claims, representations, and understandings of the parties in connection with the subject matter hereof.
- b. **Assignment.** You may not assign this Agreement without our prior written consent. We may assign this Agreement at any time, without notice. Subject to the foregoing, this Agreement will be binding on each party's successors and permitted assigns.
- c. **Severability.** If any provision of this Agreement shall be deemed unlawful, void, or for any reason unenforceable, then that provision shall be deemed severable from these terms and conditions and shall not effect the validity and enforceability of any remaining provisions.
- d. **No Waiver.** We will not be considered to have waived any of our rights or remedies, or portion thereof, unless the waiver is in writing and signed by us. Amazon Mechanical Turk's failure to enforce the strict performance of any provision of this Agreement will not constitute a waiver of Amazon Mechanical Turk's right to subsequently enforce such provision or any other provisions of this Agreement.

- e. **Notices.** All notices relating to this Agreement (including the Payment Service) will be sent by e-mail or will be posted on the Site. We will send notices to you at the e-mail address maintained in our records for you. You must send notices to us at our current e-mail address published on the Site. E-mail notices are deemed written notices for all purposes for which written notices may be required. E-mail notices are deemed received the business day after transmission if properly addressed to the intended recipient

Appendix G
Request for Participation

Request for Participation

To be located on the front page of the survey site and completed prior to entry into survey:

Dear Crowdsourcer Reviewer;

My name is Alexander Manga. I am a doctoral candidate in the Interdisciplinary Ph. D. in Evaluation program at Western Michigan University. I would like to request your participation in my dissertation research, ***“Crowdsourcer product reviews and evaluations: standards and practice.”*** This study will help us understand what evaluation methods and standards are used in crowdsourced product evaluations.

The crowdsourcer survey will be implemented via an online questionnaire and the manager interviews by telephone. Completion of the survey is anticipated to take between 30 and 60 minutes. Questions center on how closely crowdsourcer evaluation methods and standards adhere to The Joint Committee on Standards for Educational Evaluation (JCSEE) Program Evaluation Standards (PES).

There are no known risks to participating in the study. This research will contribute to the body of knowledge regarding crowdsourcing and crowdsourced product evaluations. In the future, this research may be used to further inform and improve the state of crowdsourced evaluations, inform workforce development initiatives, and provide other useful information for your companies.

Study respondents will be reimbursed \$.50 by their respective employer. The information collected during the study will be available only to the student investigator and advisors of the study. Participants can choose to stop participating in the study at any time for any reason clicking: **“Agree to Consent”**.

Findings from the study will be disseminated in a dissertation and possible publications from the dissertation. Findings may also be disseminated via presentations at conferences, informal meetings, or as part of future grant proposals

Appendix H
Consent Letter

Dear Crowdsourcer Reviewer;

My name is Alexander Manga. I am a doctoral candidate in the Interdisciplinary Evaluation PhD program at Western Michigan University. I would like to request your participation in my dissertation research titled, "Crowdsourced Product Evaluations: A Determination and Analysis of Standards Used." This study will help us understand what evaluation methods and standards are used in crowdsourced product evaluations.

The crowdsourcer survey will be implemented via an online questionnaire. Completion of the survey is anticipated to take between 30 and 60 minutes. Questions center on how closely crowdsourcer evaluation methods and standards adhere to the Joint Committee on Standards for Educational Evaluation (JCSEE) program evaluation standards.

There are no known risks to participating in the study. This research will contribute to the body of knowledge regarding crowdsourcing and crowdsourced product evaluations. In the future, this research may be used to further inform and improve the state of crowdsourced evaluations, inform workforce development initiatives, and provide other useful information for your companies.

The information collected during the study will be available only to the student investigator and advisors of the study. Participants can choose to stop participating in the study at any time for any reason. No names or identifying information will be associated with the data collected in this study.

You may contact the Chair, Human Subjects Institutional Review Board at (269) 387-8293 or the Vice President for Research (269) 387-8298 if questions or problems arise during the course of the study. You may also contact this study's principal investigator, Dr. Daniela Schroeter, by email at daniela.schroeter@wmich.edu.

This consent document has been approved for use for one year (To be completed on or before August 4th, 2016) by the Human Subjects Institutional Review Board (HSIRB) at Western Michigan University. Do not participate in this study if the date indicated has passed.

Your consent is needed to fully participate in this study. By clicking the "I agree to participate" circle you are acknowledging that you are aware of the nature and purpose of the study and wish to proceed onto the survey questions. By clicking the "I do not agree to participate" circle you are indicating you do not wish to participate in this study.

- I agree to participate
- I do not agree to participate

Appendix I
Thematic Analysis

Thematic Analysis

This dissertation sought (1) to determine what, if any, product evaluation standards were being used by crowdsourced product reviewers. Further, it sought to determine (2) if these standards adhere to the JCSEE Program Evaluation Standards.

The survey instrument used to determine both research questions above used items designed to closely represent four domains that comprise the JCSEE Program Evaluation Standards. Further, these survey items were embedded with open response items designed to trigger based on specific answer choices.

The measured outcome of the online survey instrument completed by subjects indicated that crowdsource product reviewers do closely adhere to the JCSEE Program Evaluation Standards. However, deeper analysis of the open response items indicated a lack of prevalent use of product evaluation standards. Further, open response items revealed little understanding of end-user needs, values, and systematic processes. In terms of a systematic approach to product evaluations, open-ended responses revealed the majority of respondents used very little systematic practices or standards to handle evaluation data. Further, possibly most important, open-ended response items also showed that the majority of subjects responded indicated no change in evaluation design or strategy if the evaluation increased in level of difficulty or sophistication.

Without a doubt, the deeper qualitative analysis used in this research has been very important, if not invaluable. As such, it was important to complete this analysis properly. Braun and Clarke (2006) posed a six-step process to accurately determine and use thematic analysis and synthesis. This process was systematic, pragmatic, and understandable in its application.

The Utility domain of the JCSEE Program Evaluation Standards was concerned with the product evaluator understanding the needs and values of the consumer. As such, survey items were developed to ask subjects if they understood the need and values of the customer. If Seldom or Never was selected, the open item response question would come up and ask, "If you didn't know the needs or values of the customer, how did you perform your reviews?" Open-ended responses to the needs of the customer had 58 responses, and open-ended responses to customer values had 70 responses. Phase or step one was to critically reason these phrases and rich text as to get familiar with them and to get a feel for what the subject meant by them. The second step, was to assign codes based on reasoned and systematic method to split them. They were very easy to code, as they seemed to split almost perfectly without a lot of debate. The third phase was to assign names. This was rather easy, as a dichotomous relationship presented itself. Either the subjects were coming from a personal experience perspective, or an objective perspective based on additional information or evaluation. One theme that emerged was subjects explaining that they simply completed reviews based on their experience with the product, or what they knew. The second theme was almost the opposite. Subjects seemed to have some objective approach. Given the relationship to the data we had already collected, it became apparent it was important to determine if a response was about personal experience or "some" objective method. Further, there were seven responses that were random and without a linkage. The development of theme names made sense to develop roots of the name in personal experience and objective strategy. Therefore, the theme names developed for the Utility domain based on understanding the customer's needs and values were simple and can be seen in the table below.

Below is an example of the coding used for texts produced for the ask, “If you didn’t know the needs or values of the customer, how did you perform your reviews?”

Thematic Coding: Utility Domain. Open Item: If I had complete understanding of the requestor’s needs underpinning the review. Never or Seldom is selected

Coding Used	Content Rich Text
Needs	I was asked to just create a positive, or negative review, with no other information given. I just went from there.
Needs	I stick strictly to the main purpose of the items and if it serves that purpose well.
Personal Needs	I just explained what was important to me.
Personal Needs	I usually think of why I was looking to purchase the product and then give my take on if those needs were met or not.
Personal Needs	review to the best of my ability and be honest
Personal Needs	i only give reviews based on my experience

Below is a table that illustrates in detail the themes derived from the thematic analysis and the frequency associated with each.

Open-ended Question Related to Values	Open-ended Question Related to Needs
“Can you briefly explain your strategy for your review if you did not know the requestors values?”	“Can you briefly describe your review process if you were not informed of the needs of the reviews you were performing?”
Themes (Responses)	Themes (Responses)
Values Based on Past Experiences (46)	Experienced-based Needs (33)
Objective Values (24)	Objective Needs (18)
Random/Unrelated (7)	Random/Unrelated (7)

The final open-ended response item residing in the Utility domain, which also employed thematic analysis, was located on the item “My reviews were misused.” There were a total of five open-ended responses to this item. The triggering responses were Often, and Always, which initiated the following question: “Please explain how your

reviews have always been misused.” The process of coding (Phase 2) these responses was even more basic than the previous two questions, as these responses were clearer. Two thematic groups were coded: Not Misused, which was derived from responses that included words from the subject such as “reviews were not misused,” and the code Other, which was one random phrase. Finally, phase three, naming, was conducted as you see below. This was very natural and easy to follow and understand for the reader.

Thematic Coding: Utility Domain. Open Item: If My reviews were misused. Always Is Selected; Please explain how your reviews have always been misused.

Thematic Code		Open Item Response
Were Not Misused	Not Misused	As far as I know they haven't been misused. It is entirely possible that they have been misused without my knowledge.
Were Not Misused	Not Misused	I can't think of any time they were misused, or I was not aware of it.
Were Not Misused	Not Misused	I feel like they were honest and true and were not misused.
Randon Unrelated	Other	I was always completely honest with my reviews.
Were not Misused	Were Not Misused	The platform which I submitted my reviews have a policy that protects the review from being used for other than the purpose stated.

Finally, the Accuracy domain produced one very important open ended response. This particular open-ended response item stemmed from the survey item “I used a system of information management.” This item was designed to measure the amount of systematic organization deployed by the subject. The open-ended response item was triggered when the survey response Seldom or Never was chosen. The follow-up question deployed was, “Briefly describe your system of information management.” Again, during

the process of phase two, the text and phrases that were given could quickly be split into two groups: systematic functions or methods and non-systematic functions or methods. As such, they were coded Systematic and Un-systematic. The third phase of naming was not difficult with the linkages and readers in mind. In total, the themes were named Systematic Process, Unsystematic Process, and Random/Unrelated. There were 42 responses to this follow-up question. Results from each theme are as follows: Systematic Process had 11 responses. Theme two, Unsystematic Process, had 23 responses. Finally, theme three, Random/Unrelated, had 8 responses.

Thematic Coding: Proprietary Domain. Open Item: If I used a system of information management and Always is selected, then briefly describe your system of information management

Theme	Thematic Coding	Open Item Response
	Ad Hoc	Creating and listing reasons behind views
Unsystematic Process	Ad Hoc	Recording notes on my computer, photographs taken with my phone
Unsystematic Process	Ad Hoc	hospital data
Unsystematic Process	Ad Hoc	Often I would have a technical chart that allowed me to compare similar products/services on various dimensions.
Unsystematic Process	Ad Hoc	data banks created by me
Unsystematic Process	Ad Hoc	I keep notes on what I am testing and reviewing and different things that typically happen with each item. like with toys which is what I review mostly I have my kids play with them then I have different kids play. Its work but hey free toys
Unsystematic Process	Ad Hoc	Google
Unsystematic Process	Ad Hoc	Gathering details and taking notes
Unsystematic Process	Ad Hoc	I write down my thoughts daily as I use the product, and organize them into categories.
Unsystematic Process	Ad Hoc	I kept a list of the pros and cons
Unsystematic Process	Ad Hoc	I made sure the I used the who what where why and how method. If there was pros and cons they were listed

with the cons first.

Unsystematic Process	Ad Hoc	Tracking testing results whether objective or subjective.
Unsystematic Process	Ad Hoc	Studying each and every source of reliable information.
Unsystematic Process	Ad Hoc	sales reports with filters
Unsystematic Process	Ad Hoc	my memory
Unsystematic Process	Ad Hoc	I go with my gut instinct and logic.
Unsystematic Process	Ad Hoc	Experience
Unsystematic Process	Ad Hoc	i used my opinion then research
Unsystematic Process	Ad Hoc	weighted my opinions based on need
Unsystematic Process	Ad Hoc	I use a socratic method, "from the most basic thing, go out". Start with a root concept or goal and expand outwards.
Unsystematic Process	Ad Hoc	My system of information management provides effectiveness in strategic management. I often worked with experts in the field I was managing.
Unsystematic Process	Ad Hoc	I use spreadsheets.
Unsystematic Process	Ad Hoc	I gather up my experiences, data and information and put it into the review so I can produce good reviews that are reliable and trustworthy.
Unsystematic Process	Other	None
Unsystematic Process	Other	HP
Unsystematic Process	Other	I decided pros and cons
Unsystematic Process	Other	No
Unsystematic Process	Other	I write based on my own experience depending on the situation
Unsystematic Process	Other	I am not sure what you are asking
Unsystematic Process	Other	It goes through a process defined by management
Systematic Process	Systems Based	computerize system to manage the data
Systematic Process	Systems Based	A system or organizational ideas that allows everyone to complete a project
Systematic Process	Systems Based	Database used in process of creation as well as publication of reviews.
Systematic Process	Systems Based	DIKAR model data, information, knowledge, action, result
Systematic Process	Systems Based	Human resource management system

Systematic Process	Systems Based	It is one that I have designed on my computer which allows me to keep everything in an organized manner
Systematic Process	Systems Based	I keep a large database with all of my reviews and requesters.
Systematic Process	Systems Based	Good old fashioned trees of information. Branch = main idea. Leaves = relevant info to main idea.
Systematic Process	Systems Based	My system of information management is technology (software) based.
Systematic Process	Systems Based	I use a template for new items I purchase and input pros and cons with a general experience at the end.
Systematic Process	Systems Based	I have a 10 point system I use in every review. It is not a review if it does not cover each point thoroughly.

Thematic Coding: Utility Domain. Open Item: If I had complete understanding of the requestor's values underpinning the review. Never or Seldom is selected

Theme	Thematic Code	Open End Responses
Random Unrelated	Other	Don't understand the relevance of "values" - I simply report my experience with the product
Random Unrelated	Other	I'm not sure why the 'requestors values' are overly relevant to things like product reviews.
Random Unrelated	Other	I think broadly.
Random Unrelated	Other	Making it seemed informed and intelligent.
Random Unrelated	Other	Worth vs. Cost
Random Unrelated	Other	I just guessed or assumed.
Random Unrelated	Other	I assumed he shared my values.
Values Based On Past Experiences	Personal Values	I reviewed according to my experience
	Personal Values	To make money
	Personal Values	I review something and give an honest opinion and me experiences
	Personal Values	I only post honest reviews, regardless of who the requestor might be.
	Personal Values	I based it on what I know
	Personal Values	I made the review based on my experience with the product.
	Personal Values	I just review based upon what I was looking for in a product and whether it met my expectations
	Personal Values	I would strickly go by what I have done for others in the past and by what I would want to know.
	Personal Values	I look for the practical application and go from there. The requestors values can make it easier to identify what to highlight, and the direction to take, but it is not needed to complete the review.
	Personal Values	If I didn't know what the requestor was looking for, I would just give an honest assessment of what I thought of the product. This is what I'd do anyway usually anyway.
	Personal Values	Based it on my own values
	Personal Values	I make the attempt to be even handed in my approach. I try to give both positive and negative qualities and try my best to be as honest as possible.
	Personal Values	My experience with the product
Personal Values	the overall experience, value and how it work for me personally	
Personal Values	I review products just as I see them . The requestors values have nothing to do with it.	
Personal Values	I use honesty no matter what	

Values Based On Past Experiences

Personal Values
Personal Values

I would just give an honest review that doesn't seem too biased.

Tell them my honest opinion

I would simply tell them what I thought about it, no matter what the values were they were looking for.

Personal Values

I would be honest as to how I felt about the product.

Personal Values

i understand that a review needs to be honest and explanatory

Personal Values

Personal Values

I just provided an honest and fair review.

Personal Values

Just putting truthful stuff about my thoughts and experiences.

Personal Values

I was just honest.

Personal Values

I just was honest about my experience so users will know the truth.

Personal Values

To do the best job I could.

Personal Values

I just tried to be honest and straightforward and create something that would help an average user looking at the product like me.

Values Based On Past Experiences

Personal Values

I just give my honest review f

Personal Values

I would tell them my opinion of the product I do the best I can with the information I'm given. If unsure I research my goals before reviewing. If still unsure I ask others for ideas.

Personal Values

I know my own values, so I applied them to the review process. I write the review if I have used the product and liked it, and can give a positive review that's honest.

Personal Values

The review would be based on my experience with the product. The details of the product (good and bad.) Shipping, customer service interaction, cost, functionality.

Personal Values

Personal Values

To just review as honestly as I can

I review the products objectively. It doesn't matter much what the requestor wants. If a product is good, it will get a good ranking. If a product is bad, it gets a bad ranking. I will nitpick, I will rewrite my review if my opinion later changes. I will cover shipping, packaging, customer service and any thing else relevant I can think of in a review.

Values Based On Past Experiences

Personal Values

Based upon personal experience with similar products, desire to obtain reviewed product, pricing, product information.

Personal Values

		I always attempt to review something with how well I think the product works or how well I found it from my own experiences. I try not to let what the requester might think or feel influence me so that I can give a fair opinion to others. Not everyone is going to approach something from the same set of values or beliefs so trying to provide a review that is fair without any bias is my main goal.
	Personal Values	I did it based on my personal beliefs
	Personal Values	giving out my honest review of the product
	Personal Values	went with my gut
	Personal Values	I was reviewing based on my own experience with the products.
Values Based On Past Experiences	Personal Values	i just review about my experience with the product
	Personal Values	I'm just honest in how I feel about the product.
	Personal Values	personal experience
	Personal Values	I use my own values and talk about the things I personally like or dislike
	Personal Values	I just do it honestly of what I think.
	Personal Values	I just review it by giving my honest opinion
	Personal Values	Go with gut instinct and give honest opinion.
	Personal Values	I'd do my best to give an honest assessment.
	Personal Values	I just review products, I don't concern myself with things that have nothing to do with trying the product and being honest about my opinions.
	Personal Values	I just give honest reviews to what I have evaluated regardless of the requesters values.
	Values	To express my views and possibly be a help to others.
	Values	review to the best of my ability
	Values	I attempted to address practical considerations that 'average users' might face in comparison with financial ones.
	Values	To give a fair opinion
	Values	I JUST GAVE MY HONEST OPINION AS I ALWAYS DO
	Values	I tried to give a clear, unbiased review. I didn't want to know anything about the company.
Objective Values	Values	Assumption, or common sense
	Values	To educate other consumers about my experience with the product or service
	Values	used their products

Values	I just reviewed the product
Values	Maybe you mean something different than what I do by "requestors values". The value of the requestor does not factor into my feelings or opinions of a product. The views of a requestor should not influence a reviewer.
Values	To be as objective as possible and point out the pros/cons
Values	I would look over the product specifications and get feedback from other reviews.
Values	I would review the object from a neutral standpoint. Listing the pro's and con's of the object from my perspective. I try to be neutral as I do not know exactly who will use my review.
Values	I would see if the product is practical and has relevance for me to use
Values	I wrote about what I felt other people who might be interested in the product would want to know.
Values	If I didn't know what the requestor wanted, I would simply do my best to give my honest and truthful opinion
Values	I tried to think objectively but often the requester would say "don't leave a negative review if you don't like the product". So I had to reach to find compliments I may not otherwise say.
Values	I would just ask for more information
Values	i based it off the quality and functionality
Values	I try and write a very unbiased review that typically has pros and cons.
Values	I used my common sense and what I thought the best review was
Values	I would go into detail about pros and cons of a product. I will explore how and why this product will work for you or why it will not work for you in detail.
Values	I look to explain the product to potential customers.
Values	read other reviews, and follow the general line of overall tone and thought.

Thematic Coding: Utility Domain. Open Item: If I had complete understanding of the requestor's needs underpinning the review. Never or Seldom is selected

Theme	Thematic Code	Open Ended Response
Objective Needs	Needs	I was asked to just create a positive, or negative review, with no other information given. I just went from there.
Objective Needs	Needs	I stick strictly to the main purpose of the items and if it serves that purpose well.
Objective Needs	Needs	I just go with a general usage review and cover anything that seems pertinent for the item in question.
Random Unrelated Random Unrelated	Other	I tried to think of the users as a whole and what their needs could be
Random Unrelated Random Unrelated	Other Other	Some just ask you to write about what you would want to know about the product before hand. There is no other specific information and this is usually what I see. It will vary depending on the product I would just focus on the main information I would want about something first
Random Unrelated Random Unrelated	Other	I wanted to inform people
Random Unrelated	Other	I was told to review a product, I did so. I did not know who paid for the research
Random Unrelated	Other	I just tried to complete a general review with any details I though would be relevant to the review.
Random Unrelated	Other	I could assume the needs based on the product but I am not typically informed formally.
Random Unrelated	Other	I try to go in as neutral as possible. I try not to bring in any personal bias to the review.
Random Unrelated	Other	How would you know what the needs of the users are? My review would provide the most helpful information possible. To some it would be helpful and to others it would not. There's no way to know the needs of the users.
Experience Based Needs	Personal Needs	The same as above. I will review objectively and detailed.
	Personal Needs	Normally if a need to review is know it is brief and discreet. My process is the same and based upon integrity, whether I have use for the product, whether I

		find it useful or beneficial.
Experience Based Needs		I write about what I am using the product for and how it performed in its use. I don't think about other people that much unless it something obvious that should be adressed that other people need to know.
	Personal Needs	
	Personal Needs	It was someting that i did without consulting the needs of the users. I think that is much more fair.
	Personal Needs	I would just write the review, it doesn't matter if thye say they need it or not.
	Personal Needs	common sense and what I would want from the product
Experience Based Needs		I would review the product based on general needs or uses that I thought might be the most common.
	Personal Needs	
	Personal Needs	I would look up other reviews
	Personal Needs	Again, what does that mean, and how would I know? I just post them - I don't analyze the people who might read them!
	Personal Needs	I think broadly.
Experience Based Needs		Needs are to ambiguous to determine fully.
	Personal Needs	
Experience Based Needs		I never even thought about this.
	Personal Needs	
Experience Based Needs		I don't know
	Personal Needs	
Experience Based Needs		Guessed
	Personal Needs	I was informed that it would be posted on Amazon for other users to see.
Experience Based Needs		I just explained what was important to me.
	Personal Needs	
Experience Based Needs		I usually think of why I was looking to purchase the product and then give my take on if those needs were met or not.
	Personal Needs	
Experience Based Needs		review to the best of my ability and be honest
	Personal Needs	
Experience Based Needs		i only give reviews based on my experience
	Personal Needs	
Experience Based Needs		If asked to post a review, I reflect my true feelings about the product or service.
	Personal Needs	
Experience Based Needs		It was either a product I had already tried out beforehand, or the product was included for me to test out.
	Personal Needs	
Experience Based Needs		I detail my experience with the product to help others understand all of the
	Personal Needs	

aspects of the product that I know.

Experience Based Needs	Personal Needs	I described what I thought I would want to read in a review.
Experience Based Needs	Personal Needs	I'd just do my best and give an honest opinion.
Experience Based Needs	Personal Needs	I try to give detail about what I like and dislike about the product in order to facilitate their decision to buy or not to buy. I usually write a review with the assumption it is supposed to be honest though I am pretty sure a lot of requesters have nefarious intentions to just inflate a product's stats.
Experience Based Needs	Personal Needs	I review based on what I would need
Experience Based Needs	Personal Needs	I based them on experience.
Experience Based Needs	Personal Needs	I reviewed how I felt. I didn't feel like I needed to censor my feelings on the product.
Experience Based Needs	Personal Needs	I just give an honest review of the product
Experience Based Needs	Personal Needs	I just read them and made sure they were truthful to my experience.
Experience Based Needs	Personal Needs	Tell them my honest opinions
Experience Based Needs	Personal Needs	i base my reviews soley on my experience with the porduct or service
Experience Based Needs	Personal Needs	I provided an honest and fair review.
Experience Based Needs	Personal Needs	I was just honest.
Experience Based Needs	Personal Needs	I still just give an honest review and let the requester take it for what it is worth.
Experience Based Needs	Personal Needs	I just reviewed it based on my opinions of the product.
Experience Based Needs	Personal Needs	I would tell them my honest opinion of the product
Experience Based Needs	Personal Needs	I know my own needs from a product, and what I would want to learn from a review. That seemed sufficient.
Experience Based Needs	Personal Needs	I used my gut feeling
Experience Based Needs	Personal Needs	I'd just tell my honest opinion
Experience Based Needs	Personal Needs	I reviewed it based on my needs and wants and preferences
Experience Based Needs	Personal Needs	I just wrote true reviews.
Experience Based Needs	Personal Needs	i just review about my experience with the product
Experience Based Needs	Personal Needs	Again I'm just honest about the product and let the user take it how it will.
Experience Based Needs	Personal Needs	I was told to give my honest opinions. It was a fairly thorough process, so I assume they take what they need from

		what I give.
Experience Based Needs	Personal Needs	personal experience
Experience Based Needs	Personal Needs	Go with gut instinct and give honest opinion.
Experience Based Needs	Personal Needs	I put myself in their place.

Thematic Coding: Utility Domain. Open Item: If My reviews were misused. Always is selected, please explain how your reviews have always been misused.

Thematic Code		Open Item Response
Were Not Misused	Not Misused	As far as I know they haven't been misused. It is entirely possible that they have been misused without my knowledge.
Were Not Misused	Not Misused	I can't think of any time they were misused, or I was not aware of it.
Were Not Misused	Not Misused	I feel like they were honest and true and were not misused.
Randon Unrelated	Other	I was always completely honest with my reviews.
Were not Misused	Were Not Misused	The platform which I submitted my reviews have a policy that protects the review from being used for other than the purpose stated.

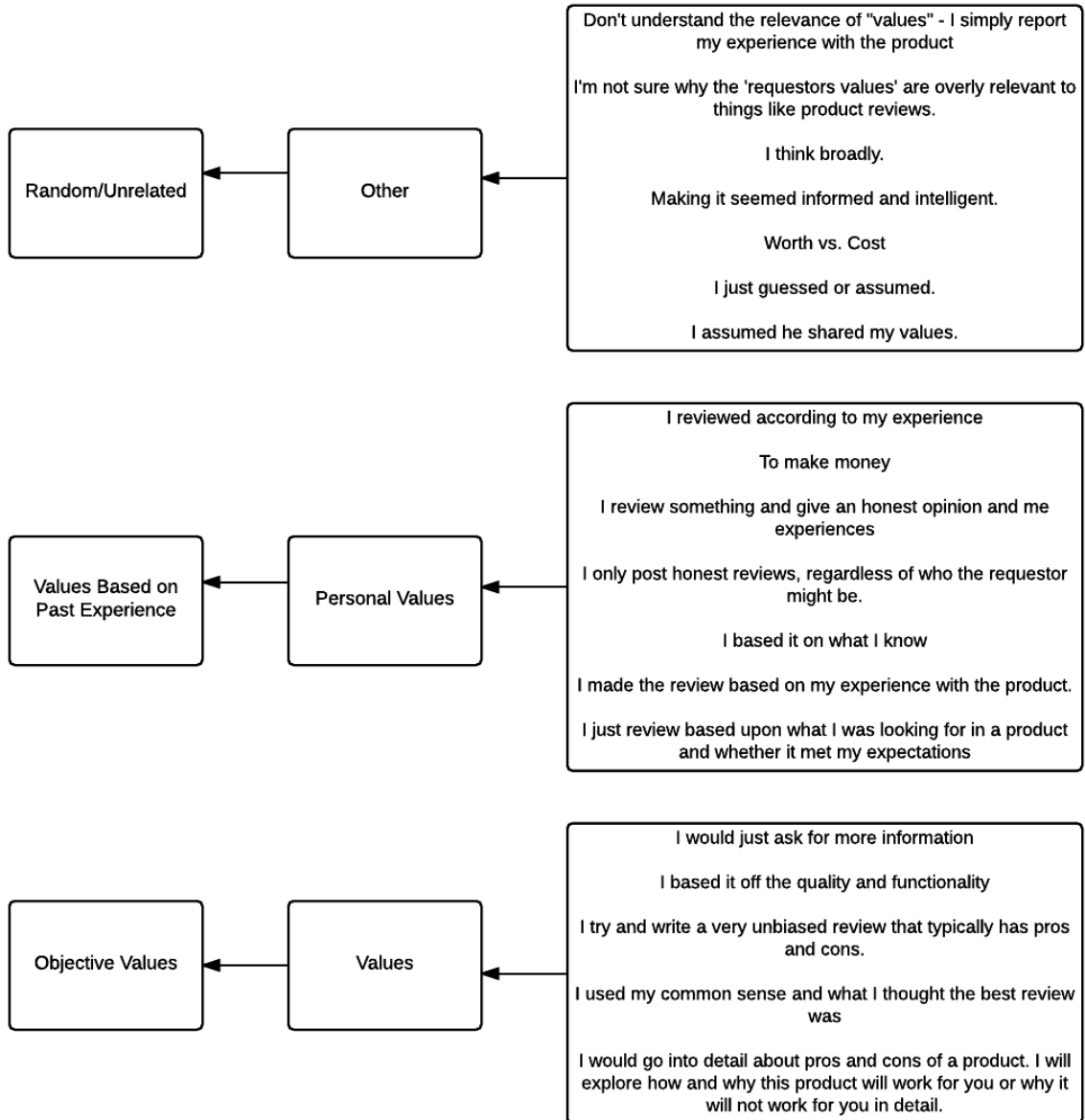
Thematic Coding: Proprietary Domain. Open Item: If I used a system of information management and Always is selected, then briefly describe your system of information management.

Theme	Thematic Coding	Open Item Response
		Creating and listing reasons behind views
Unsystematic Process	Ad Hoc	Recording notes on my computer, photographs taken with my phone
Unsystematic Process	Ad Hoc	hospital data
Unsystematic Process	Ad Hoc	Often I would have a technical chart that allowed me to compare similar products/services on various dimensions.
Unsystematic Process	Ad Hoc	data banks created by me
Unsystematic Process	Ad Hoc	I keep notes on what I am testing and reviewing and different things that typically happen with each item. like with toys which is what I review mostly
		I have my kids play with them then I have different kids play. Its work but hey free toys
Unsystematic Process	Ad Hoc	Google
Unsystematic Process	Ad Hoc	Gathering details and taking notes
Unsystematic Process	Ad Hoc	I write down my thoughts daily as I use the product, and organize them into categories.
Unsystematic Process	Ad Hoc	I kept a list of the pros and cons
Unsystematic Process	Ad Hoc	I made sure the I used the who what where why and how method. If there was pros and cons they were listed with the cons first.
Unsystematic Process	Ad Hoc	Tracking testing results whether objective or subjective.
Unsystematic Process	Ad Hoc	Studying each and every source of reliable information.
Unsystematic Process	Ad Hoc	sales reports with filters
Unsystematic Process	Ad Hoc	my memory
Unsystematic Process	Ad Hoc	I go with my gut instinct and logic.
Unsystematic Process	Ad Hoc	Experience
Unsystematic Process	Ad Hoc	i used my opinion then research
Unsystematic Process	Ad Hoc	weighted my opinions based on need
Unsystematic Process	Ad Hoc	I use a socratic method, "from the most basic thing, go out". Start with a root concept or goal and expand outwards.
Unsystematic Process	Ad Hoc	My system of information management provides effectiveness in stragteic managment. I often worked with experts in the field I was managing.
	Ad Hoc	

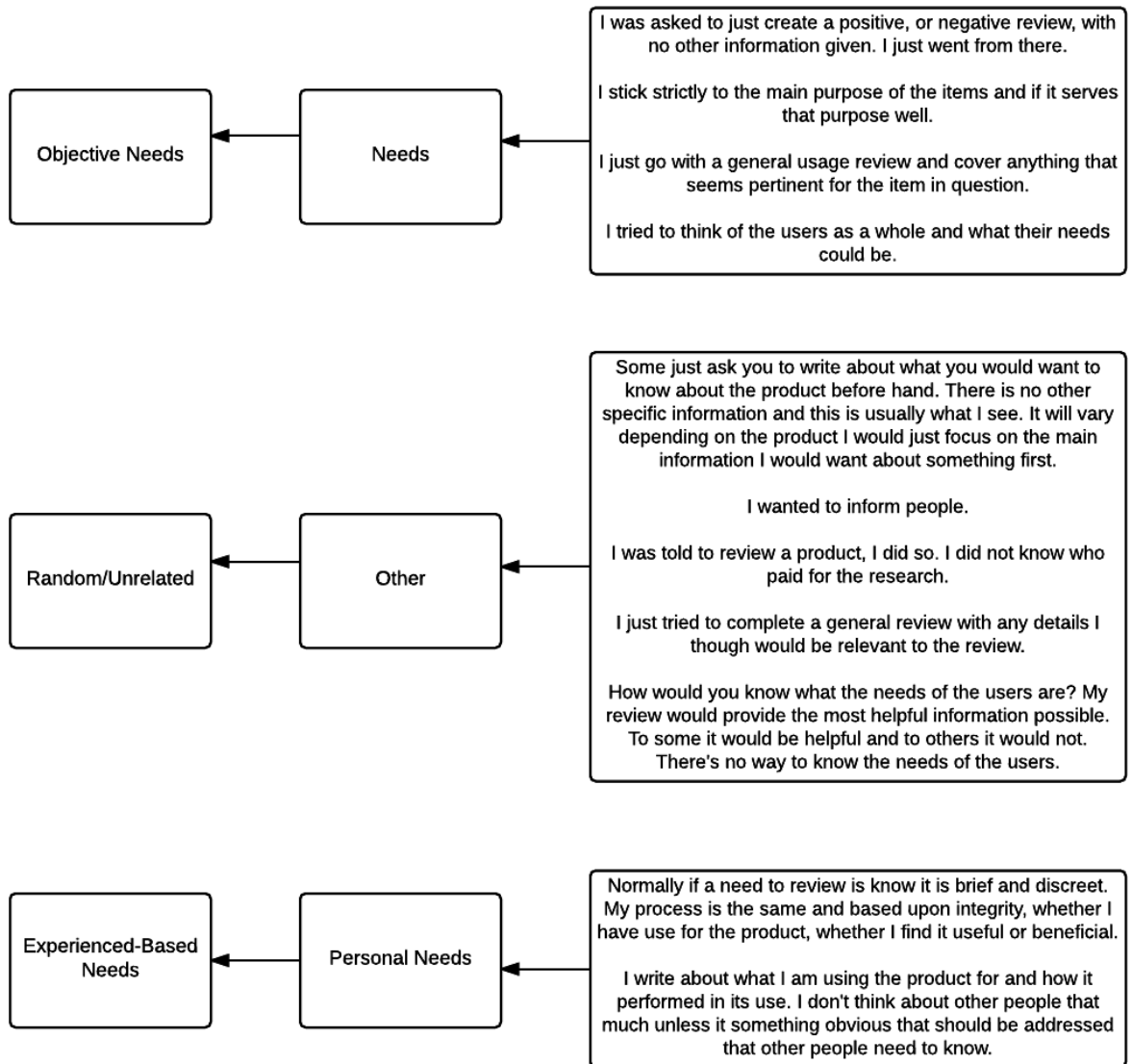
Unsystematic Process	Ad Hoc	I use spreadsheets.
Unsystematic Process		I gather up my experiences, data and information and put it into the review so I can produce good reviews that are reliable and trustworthy.
	Ad Hoc	
Unsystematic Process	Other	None
Unsystematic Process	Other	HP
Unsystematic Process	Other	I decided pros and cons
Unsystematic Process	Other	No
Unsystematic Process	Other	I write based on my own experience depending on the situation
Unsystematic Process	Other	I am not sure what you are asking
Unsystematic Process		It goes through a process defined by management
	Other	
Systematic Process	Systems Based	computerize system to manage the data
Systematic Process		A system or organizational ideas that allows everyone to complete a project
	Systems Based	
Systematic Process		Database used in process of creation as well as publication of reviews.
	Systems Based	
Systematic Process		DIKAR model data, information, knowledge, action, result
	Systems Based	
Systematic Process		Human resource management system
Systematic Process		It is one that I have designed on my computer which allows me to keep everything in an organized manner
	Systems Based	
Systematic Process		I keep a large database with all of my reviews and requesters.
	Systems Based	
Systematic Process		Good old fashioned trees of information. Branch = main idea. Leaves = relevant info to main idea.
	Systems Based	
Systematic Process		My system of information management is technology (software) based.
	Systems Based	
Systematic Process		I use a template for new items I purchase and input pros and cons with a general experience at the end.
	Systems Based	
Systematic Process		I have a 10 point system I use in every review. It is not a review if it does not cover each point thoroughly.
	Systems Based	

Appendix J
Thematic Mapping

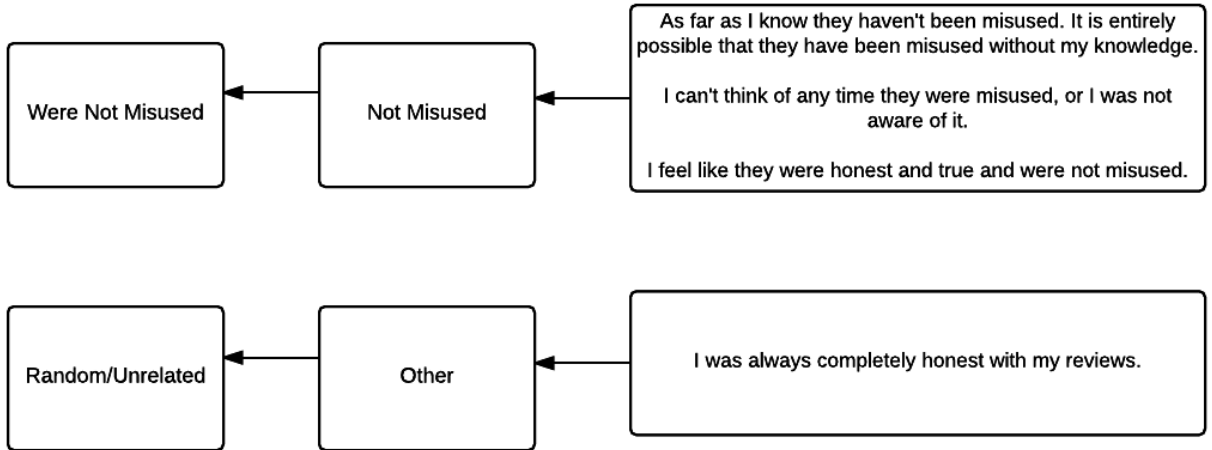
Thematic Coding: Utility Domain. Open Item: If I had complete understanding of the requestor's values underpinning the review. Never or Seldom is selected



Thematic Coding: Utility Domain. Open Item: If I had complete understanding of the requestor's needs underpinning the review. Never or Seldom is selected



Thematic Coding: Utility Domain. Open Item: If My reviews were misused Always is selected: Please explain how your reviews have always been misused.



Thematic Coding: Proprietary Domain. Open Item: If I used a system of information management and Always is selected, then Briefly describe your system of information management.

