MIS Quarterly

RESEARCH ESSAY

DATA COLLECTION IN THE DIGITAL AGE: INNOVATIVE ALTERNATIVES TO STUDENT SAMPLES[1]

Zachary R. Steelman
Spears School of Business, Oklahoma State University, Tulsa, OK 74106 U.S.A. {zach@steelman.com}

Bryan I. Hammer
Spears School of Business, Management Science and Information Systems, Oklahoma State University, Stillwater, OK 74078 U.S.A. {bryan.hammer@okstate.edu}

Moez Limayem
College of Business, University of South Florida, Tampa, FL 33620 U.S.A. {mlimayem@usf.edu}

Online crowdsourcing markets (OCM) are becoming more popular as a source for data collection. In this paper, we examine the consistency of survey results across student samples, consumer panels, and online crowdsourcing markets (specifically Amazon's Mechanical Turk) both within the United States and outside. We conduct two studies examining the technology acceptance model (TAM) and the expectation–disconfirmation theory (EDT) to explore potential differences in demographics, psychometrics, structural model estimates, and measurement invariances. Our findings indicate that (1) U.S.-based OCM samples provide demographics much more similar to our student and consumer panel samples than the non-U.S.-based OCM samples; (2) both U.S. and non-U.S. OCM samples provide initial psychometric properties (reliability, convergent, and divergent validity) that are similar to those of both student and consumer panels; (3) non-U.S. OCM samples generally provide differences in scale means compared to those of our students, consumer panels, and U.S. OCM samples; and (4) one of the non-U.S. OCM samples refuted the highly replicated and validated TAM model in the relationship of perceived usefulness to behavioral intentions. Although our post hoc analyses isolated some cultural and demographic effects with regard to the non-U.S. samples in Study 1, they did not address the model differences found in Study 2. Specifically, the inclusion of non-U.S. OCM respondents led to statistically significant differences in parameter estimates, and hence to different statistical conclusions. Due to these unexplained differences that exist within the non-U.S. OCM samples, we caution that the inclusion of non-U.S. OCM participants may lead to different conclusions than studies with only U.S. OCM participants. We are unable to conclude whether this is due to of cultural differences, differences in the demographic profiles of non-U.S. OCM participants, or some unexplored factors within the models. Therefore, until further research is conducted to explore these differences in detail, we urge researchers utilizing OCMs with the intention to generalize to U.S. populations focus on U.S.-based participants and exercise caution in using non-U.S. participants. We further recommend that researchers should clearly describe their OCM usage and design (e.g., demographics, participant filters, etc.) procedures. Overall, we find that U.S. OCM samples produced models that lead to similar statistical conclusions as both U.S. students and U.S. consumer panels at a considerably reduced cost.

Keywords: Data collection, crowdsourcing, sampling, online research, crowdsourcing market

[1]Rajiv Kohli was the accepting senior editor for this paper. Eric Walden served as the associate editor.

The appendices for this paper are located in the "Online Supplements" section of the *MIS Quarterly*'s website (http://www.misq.org).

Introduction

Increasing pressures on both faculty members and Ph.D. students to publish (Dean et al. 2011) have also led to a greater number of active researchers, thereby enhancing competition for survey respondents and making data collection even more challenging. Simultaneously, relevant data is becoming more difficult to gather from external sources as a result of strict corporate information policies, competition for corporate participation, and a lack of availability, especially for researchers in smaller universities. A common solution is to recruit students as a sampling frame,[2] but this population poses its own issues, including low participation motivation, limited demographic heterogeneity, and demanding university policies regarding student participation (Gordon et al. 1986; Peterson 2002; Thomas 2011). The quantity of student participants available to nonacademic researchers or those from smaller universities is a limitation as well (Mason and Suri 2012).

Thus, some researchers have begun to experiment with alternative sampling frames, using online crowdsourcing markets (OCM) such as Amazon's Mechanical Turk,[3] InnoCentive,[4] or Crowdflower[5] (Bohannon 2011; Marks 2011; Paolacci et al. 2010). Crowdsourcing initially was conceptualized to outsource tasks to an undefined network of laborers, using an open call (Howe 2006). Accordingly, *online crowdsourcing markets* entail web-based environments in which employers post outsourced tasks to an undefined, anonymous network of laborers who are compensated for their contribution. The large, demographically diverse (Ross et al. 2010), and motivated (Kaufmann et al. 2011) international online participant workforce thus offers some promise as a mainstream research tool (Bohannon 2011; Conway and Limayem 2011; Mogilner et al. 2012; Spiller 2011; Yu and Nickerson 2011) that is slowly emerging through conferences in the Information Systems discipline and currently accepted in leading journals of allied disciplines such as Marketing (see Appendix A). Despite such increasing uses of OCMs, however, we have limited insights into their validity and reliability for academic research, especially in IS.

Some researchers have begun to investigate the demographic properties of one such popular OCM, Amazon.com's Mechanical Turk,[6] as an appropriate and acceptable means to collect individual-level data (Buhrmester et al. 2011; Marge et al. 2010; Paolacci et al. 2010; Ross et al. 2010; Sprouse 2011). Applications of Mechanical Turk include user evaluation studies with crowdsourcing workers (Kittur et al. 2008; Sorokin et al. 2010), language transcriptions (Marge et al. 2010), experimental designs (Bursztein et al. 2010), and qualitative designs (Ward and Broniarczyk 2011). However, although such studies offer an important exploratory foundation, a rigorous empirical analysis that compares OCM samples with more traditional approaches, on aspects other than simple demographics, is crucial for evaluating the validity of these samples for academic research. Some authors using OCM samples have examined the reliability of their construct measurement (Buhrmester et al. 2011), but many fail to evaluate consistency with prior methods by comparing their results against previously validated results for a particular construct. Validating novel techniques and methods demands consideration of not only reliability (e.g., Cronbach's alpha, composite reliability) but also a comparison of results with previously validated techniques and methods (Honaker 1988; Meyerson and Tryon 2003). For example, as Internet-mediated studies emerged, it was necessary to reexamine the construct and measurement validation of the new sampling and testing techniques to ensure consistency with previous techniques (Buchanan and Smith 1999). Similarly, the use of OCMs requires a further empirical examination of the demographics, psychometric properties (reliability, convergent validity, and discriminant validity), and structural model estimations, as well as a comparison of any differences among alternative sampling frames to determine parallels in the results obtained (Meyerson and Tryon 2003).

This study aims to examine the similarities and differences between OCMs and more traditional sampling frames through the empirical investigation of each sample's demographic composition, psychometric properties, construct validity and reliability, theoretical models, and measurement invariance. We open with an introduction to OCMs and their potential issues as well as benefits for academic research. We then outline our methods and analyses of the demographic differences across five distinct samples (consumer panels, college students, worldwide OCM, U.S. OCM, and non-U.S. OCM), the psychometric properties associated with each construct from the technology acceptance model (TAM) (Davis 1989; Davis et al. 1989), the structural path models, and a series of group invariance tests. We additionally repli-

[2]Sampling frames are defined as an accessible listing of a population of participants used for sample selection (Trochim and Donnelly 2006).

[3]http://www.mturk.com/.

[4]http://www.innocentive.com/.

[5]http://crowdflower.com/.

[6]"Mechanical Turk" refers to an 18th century chess-playing machine that was secretly operated by a human.

cate our set of analyses utilizing the expectation–disconfirmation theory[7] (EDT; Oliver 1980) to further strengthen our findings. We conclude with a discussion of our results, their theoretical and practical implications, potential limitations, and future research. We also offer a set of procedural and sample attributes that researchers should report when relying on OCMs.

Online Crowdsourcing Markets

We define OCMs as web-based environments where employers post outsourced tasks for an undefined, anonymous network of laborers to perform and receive compensation for their contributions. On Amazon's Mechanical Turk, registered users (called Workers) participate in tasks issued by individual employers (Requesters) that solicit the work. These OCM tasks include simple traditional crowdsourcing processes such as identifying pictures, creating keywords or tags, and cleaning data; however, they may also be as complex as audio transcriptions, detailed product reviews, and experimental surveys. For any given task, workers receive a predetermined monetary amount for the successful completion of the task, and the payment moves through built-in OCM payment systems. However, requesters can deny payment for poor quality work or provide bonuses for exceptional work, which provides an additional incentive for workers (Bohannon 2011).

OCMs such as Amazon's Mechanical Turk exhibit several unique traits, compared with traditional online sampling, including (1) the complete anonymity of the sample, (2) the motivation to visit the recruitment location, (3) controls for participant selection and recruitment, and (4) built-in payment systems for incentive disbursement. If researchers use a sampling frame obtained from traditional online techniques, such as discussion forums, online communities, and chat rooms, they possess some information that is not typically available for OCMs. For example, they can determine with some accuracy the total number of registered members in an online forum who see and might respond to a posted message. However, with OCMs, no accessible list of members defines the sampling frame, because the participants are typically hidden behind anonymous identifiers. This can also lead to difficulties in estimating accurate response rates.

In addition, the motivation to frequent an online forum, chat room, or other online community tends to be homogenous across participants (Miller et al. 1996). For example, visitors to an automotive forum focused on a specific make and model have fairly consistent interests and motivations for participating. In contrast, OCMs support a much broader set of visitation motivations for workers, who might pursue payment but also tend to derive hedonic value from the experience (Kaufmann et al. 2011; Sun et al. 2011). OCMs also provide the filtering and participant selection control features, which increase the researcher's ability to select specific candidates, but are difficult to incorporate in online forums, chat rooms, and communities. Finally, the integrated marketplaces that make up OCMs also include automated payment features, which facilitate the distribution of incentives to workers without the need to hire third-party services.

Online crowdsourcing markets provide unique benefits (Mason and Suri 2012). First, these populations are diverse in terms of culture, occupation, and age (Mason and Suri 2012; Ross et al. 2010). Therefore, researchers are potentially able to access a greater variety of demographic attributes to investigate within their studies compared to those of a more homogenous student sample (Peterson 2002). If the researcher needs to target a specific demographic and, therefore, filter out some participants, it also is possible to design a participant qualification requirement, such that only those respondent profiles specified by the researcher may view, access, and participate in the study.

Second, the large and highly motivated population of workers is persistently available to participate in research studies. Although student participants may be similarly available, various studies indicate that the median sample sizes of student-based studies are smaller than those with nonstudent participant samples (Shen et al. 2011). Furthermore, it is often difficult to motivate students to participate in research studies; researchers rely on course credit or institutional funding to pay students for their participation. In turn, finding faculty willing to grant course credit or provide class time for students to participate is not easy, especially when course restrictions and university policies prohibit such methods. No such problems arise with OCMs, which also tend to cost much less than typical consumer panels, laboratory studies, and prior online sample recruitment techniques (Mason and Suri 2012). Finally, identifying appropriate student participants is both time consuming and resource intensive, whereas the constant flow of OCM workers remains relatively stable, especially compared with the inherent seasonality that affects traditional student sampling (Mason and Suri 2012). In this sense, OCMs remove the leg work to find a large sample—by their very nature, OCMs simply attract workers to participate.

However, the presence of large populations at low cost likely increases the need for a more detailed inspection and valida-

[7] Also known as expectation–confirmation theory.

tion of responses, as is similarly required by scenarios that grant researchers little control over the testing environment (Vadillo and Matute 2011). In novel sampling frames such as OCMs, method biases might have significant impacts on construct measurements (Burton-Jones 2009). Therefore, the need for increased scrutiny of the responses, such as to detect workers with lowered attention or automated robotic responses, must be addressed and planned for in advance (Mason and Suri 2012). Another assessment should exploit the OCMs' ability to track and control for multiple responses by the same person, while still providing anonymity, because they use unique identifiers to track participation. Therefore, researchers should filter out workers who might have participated in an earlier version of a survey or pilot study to reduce bias. Finally, only workers with high reputation ratings for their previous quality work should ideally be allowed to participate in an OCM study, which gives researchers a further means to address reliability and quality issues. Despite the ability of researchers to design quality controls for a recruitment method, the potential for biased results still exists.

When examining the validity and benefits of a new recruitment and sampling technique, we must make note of the potential impacts of a variety of biases such as nonresponse, coverage, and sampling bias (Groves 1989). Nonresponse bias occurs when individuals cannot or will not participate in the survey, creating deviations between the population estimate and sample estimate due to differences between respondents and nonrespondents. Coverage bias may occur when sets of individuals within the population are nonexistent within the sampling frame and therefore their responses are not captured. Sampling bias such as selection bias can occur when individuals within the sampling frame have a higher probability of participation and selection than others. The presence of any bias within the results derived from a sample can lead to inconclusive, inaccurate, or inconsistent results from theory and prior research. The evaluation of each of these biases is directly measurable only in probability samples, or those where the sampling frame and chance of selection are known (Couper 2000). The context of this research is specifically on ***nonprobability*** samples which include convenience samples such as college students, online consumer panels, open-call web surveys, and OCMs. The use of convenience samples such as these inherently increases the potential for a variety of biases as well as the lack of ability to directly evaluate them which should at a minimum be reported and noted by researchers.

One of the largest problems related to web-based surveys, such as OCMs, is the inability to accurately define the sampling frame and thus the potential response rates (Couper 2000). Within an OCM environment, the number of responses to be collected is predefined and prepurchased such that when the quota is reached, the recruitment link is removed from the market. Therefore, there is no indication of or ability to determine which types of individuals would have participated if they still had access to the survey. For example, in Mechanical Turk a Requester can set the number of Workers to 400 participants. Once this number is reached, no more participants can engage in the task. This problem is not one of OCMs alone but is similar to traditional online consumer panels that are purchased via a similar quota-based response collection. Accordingly, the composition of participants that could have potentially seen and responded to a survey is a function of the number of available slots as well as the timeline of the recruitment listing. A task can be posted for a set number of days, for example, and once the time has elapsed the request is closed, barring any other participants from engaging in the task. Therefore, to capture the potential nonresponse bias present within an OCM requires the comparison to alternative, more accurate sources,[8] which have the same or similar goals as the focal and auxiliary variables, which may be correlated with response propensity (Couper 2000; Groves 2006). However, the simple comparison of demographics alone does not indicate a lack of bias in respondent attitudes and selection (Couper 2000). Interestingly, recent research has suggested that changes in response rates may not alter some survey estimates (Curtin et al. 2000; Keeter et al. 2000). Therefore, to determine the extent of differences and potential biases between OCMs and traditional techniques, this study examines not only the demographic differences but the psychometric properties and mean levels of each measurement scale as well as the structural models of theoretical relationships.

In addition to nonresponse biases, there inherently exists a selection bias within web-based surveys of individuals who have access to the internet (Schmidt 1997). The argument on the potential for this bias to impact results has been debated for years despite the increasing proliferation of Internet access. However, researchers utilizing OCMs or other web-based recruitment techniques must keep this potential bias in mind when examining theoretical models that may be influenced by a variety of socio-economic attributes related to Internet access. A further concern of consumer panels as well as OCMs is the potential for biased responses due to repeated survey participation, or panel conditioning (Toepoel et al. 2008). Individuals who are continuously participating in surveys may become accustomed to specific designs and

[8]We utilize both students and online consumer panels as the accepted norms for comparisons to OCM.

respond differently than non-conditioned individuals. However, our focus is not on directly testing the effects of panel conditioning but on comparisons among OCMs, consumer panels, and student samples, which may each exhibit a level of panel conditioning due to their continued use for research participation.

Several researchers have investigated differences between web-based and traditional survey methods (Birnbaum 2004; Gosling et al. 2004; Lewis et al. 2009; Meyerson and Tryon 2003), noting many of the same potential threats to validity: (1) nature of the sample, (2) volunteer status, (3) nature of the testing environment, (4) technological issues, (5) multiple completions, and (6) mischievous responses (Buchanan and Smith 1999). Despite some guidelines and procedures for addressing certain issues with online studies (see Birnbaum 2000, 2004), many articles still fail to report or justify the validity of their samples and measures when using OCMs. For example, studies often fail to list the specific recruitment restrictions for participation, payment levels, or demographic information other than age and gender (e.g., Bagchi and Li 2011; Leonhardt et al. 2011; Paolacci et al. 2011; Sakamoto and Bao 2011; Yu and Nickerson 2011). The use of computer-based questionnaires has become broadly accepted, but despite the examination of web surveys and the sampling of anonymous online populations over the years it has yet to be fully established across research fields, especially in terms of the unfamiliarity and validity concerns of the sampling frame (Couper 2000; Paolacci et al. 2010; Schmidt 1997).

To increase awareness and address such concerns, we undertake a rigorous empirical examination of the validity of emerging OCMs as a viable alternative to traditional sampling frames to recruit study participants. We specifically note unique caveats of OCMs, which may invalidate, bias, or mask results if not clearly addressed and reported by researchers. Using Amazon's Mechanical Turk as an illustrative example, we conduct a series of data collections that reveal the effectiveness, validity, and reliability of this sampling frame.

Method

Because the focus of this study is not to propose new theory or measures, we selected TAM in its most basic form to conduct our initial study; thus, our primary focus remains squarely on the technique and sampling properties, not the theoretical model. TAM (Figure 1) has been subject to extensive assessments of its reliability, consistency, and validity across multiple contexts (King and He 2006), so it supports reliable comparisons with previous research expectations (Buchanan and Smith 1999). This consistency across samples gives us confidence that the differences in responses are due in part to the sample compositions and not the questionnaire or theoretical design. The number of measures it features is not extensive, such that we can perform a more focused analysis of the variables. As one of the most familiar and well-known information systems theories, TAM also allows readers to direct focus to the analysis of differences in each sampling frame and their resulting psychometric properties. In Appendix B we provide a complete listing of all items we used for this study.

In an attempt to duplicate and increase the strength of our findings in the initial TAM study we replicated the entire analyses procedure utilizing the EDT.[9] Due to space constraints and to provide a clear and concise description of each study we have placed all analyses for our second study in Appendix C. Our findings were highly consistent across both studies, in both differences and similarities, providing robust insights into the potential use of OCMs as a sampling frame and the resulting research implications and recommendations. We present below the results from Study 1 while noting any distinct differences between Study 1 and Study 2 within the discussion section.

Participants and Demographics

To collect data for this study, we used three sampling frames across our recruitment techniques: (1) a nationwide consumer panel in the United States, (2) college students from a major midwestern U.S. university, and (3) users of Amazon's Mechanical Turk. From these frames, we gathered three primary samples: a consumer panel sample of adults in the United States, a sample of university students, and a sample of worldwide OCM participants without restrictions. To detail the differences that may arise from an unrestricted worldwide listing, we further collected two additional OCM samples of users from (1) the United States and (2) any non-U.S. country. With these additional subsamples, we are able to examine more closely the potential for problems that may arise if researchers use OCMs without accounting for the country of origin of workers. In all five samples, we collected participant data using the same survey questionnaires.

Procedure

We conducted an empirical test of the research model using data collected from an online survey. After giving their consent, participants were instructed to watch a short video that

[9] We thank the review team for recommending this addition.

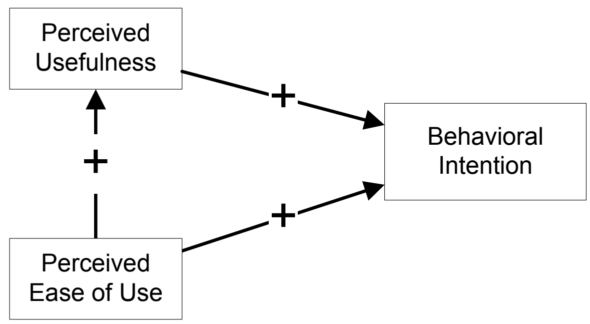


Figure 1. Original TAM Model

showed various Windows 7 features.[10] Windows 7 easily enables work-related tasks such as word processing, spreadsheet usage, and presentation preparation, all of which align with the conceptualizations in the original TAM scales (Davis 1989). All participants were required to have both audio and video capabilities to participate in the survey. After the video, participants received the scales, followed by demographic items.

The worldwide, U.S., and non-U.S. OCM task listings on Amazon's Mechanical Turk were identical, with the exception of the relevant physical location restrictions.[11] The OCM samples were compensated $0.20 for their participation and only those workers who completed the entire survey were paid. We requested a consumer panel with a quota of 250 usable responses based on a quoted price of $5.25 per response.[12] Student participants were entered into a drawing for ten $10.00 gift cards for completed responses.

While the calculation of actual response rates is not possible with many of our samples, the examination of complete responses may provide some insight into differences between techniques (Bethlehem and Biffignandi 2012). To begin this analysis we dropped responses that were incomplete or failed the attention and response quality questions employed in the study (see Appendix B). This procedure was conducted to eliminate responses provided by computerized programs or systematic responses or by participants with low participation and attention as recommended by prior research on online surveys (Birnbaum 2004). In Table 1 we display the distribution of the responses received, removed, and usable for each sample. The final set of usable data contained 256 consumer panel responses,[13] 165 student responses, 193 for the worldwide OCM, 222 for the U.S. OCM, and 212 for the non-U.S. OCM.

We specifically utilize these *minimally* cleaned data sets for the entirety of the following analyses. We removed individuals who (1) didn't complete the entire survey and thus did not provide the needed data for the complete analyses or (2) failed the attention and response quality questions, indicating potentially falsified responses. These actions are based on prior recommendations in the use of online research (Aust et al. 2012; Oppenheimer et al. 2009), which we feel all researchers should examine and address before conducting their own statistical analyses regardless of the recruitment techniques.[14] However, to retain the quality level of the data that was utilized within each analytical step, we made no further changes to remove any potential outliers or adjust the structural model parameters to increase model fit indices. Therefore, the following results indicate what one could expect to find if the data cleaning procedure was *minimal* and focused purely on a response completion and initial data quality checks without conducting an extensive outlier analysis or model adjustment procedure as could be expected in a typical analysis procedure.

[10]The features were Windows Gadgets, Jump Lists, Windows Live Movie Maker, Aero Peek, Windows Shake, Windows Snap, multimedia streaming, and taskbar previews.

[11]The features of Amazon's Mechanical Turk at the time of data collection were limited to simple restrictions such as those based on physical location of the participants. We utilize a U.S. restriction to conceptually mirror the constraints of the consumer panel and college student origins, the non-U.S. sample blocked all participants located in the United States, while the worldwide listing had no restrictions in place for location. Detailed settings for the task listings are available from authors upon request.

[12]The actual amount paid to each participant was not disclosed.

[13]The consumer panel provided additional responses to cover the potential for faulty responses.

[14]In addition, we ran the analyses utilizing the full data sets without removing participants and found a relatively consistent pattern of results; therefore, we present these minimally cleaned data sets for our primary analyses.

Table 1. Participant Response Removal

	Students	Worldwide OCM	U.S. OCM	Non-U.S. OCM	Consumer Panel
Total Responses	178	209	236	237	268
Failure to Finish	6	2	4	13	11
Failed Quality Questions	7	14	10	12	1
Final Usable Responses	165	193	222	212	256
Percent Usable	93%	92%	94%	89%	96%

We also measured the total amount of time required for the entire data collection in each sample. In the worldwide OCM survey, we received 209 responses in only 2 days. The non-U.S. OCM study took 3 days to achieve 237 responses. Alternatively, in the U.S. OCM study, 236 participants had engaged after 60 days.[15] The significantly longer period in the U.S. OCM sample may be a function of the consistent, relatively low payment level for each sample as the Study 2 U.S. OCM sample only took 7 days to receive 239 responses for an incentive of $0.50. The student sample took approximately 20 days to achieve 178 responses while the consumer panel only took 4 days to receive 268 responses. The mean response times ranged between 10 and 13.5 minutes in all five samples; the minimum time across all samples was 2 minutes, which clearly indicated fictitious answers. However, because the results were relatively equivalent across samples, and our goal was to examine the initial differences, we retained these complete, yet clearly fictitious, responses for the analysis.

Analysis and Results

We begin with a comparison of the demographic differences, and then undertake exploratory factor analysis to examine the initial structure of the measures. In addition, we describe our confirmatory factor analysis, which establishes the psychometric properties of the measures using covariance-based structural equation modeling (CB-SEM). We also conduct a structural path analysis using CB-SEM to test the hypothesized relationships within our TAM models. Finally, we describe the CB-SEM group invariance tests, conducted to examine any further differences in the measurement characteristics (e.g., loadings, variances, and means) of each sampling frame. These analyses were conducted using the statistical package R (R Core Team 2012). The entire set of analyses was replicated with partial least squares (PLS) to address any concerns of sample sizes below the recommended CB-SEM threshold of 200 responses.[16] The full results of our PLS analyses are presented in Appendix D with consistent results. Table 2 presents an overview of the study procedure detailing the focus and empirical tests examined in each step of the analysis.

Demographics

The demographics in the samples differed in a few ways, as anticipated (Meyerson and Tryon 2003; Ross et al. 2010), and we list these details in Tables 3 and 4. The three OCM samples were older than the student sample yet still younger than the consumer panel. In terms of gender, the consumer panel and U.S. OCM sample consisted of more women (55.4% and 57.2%, respectively), whereas the worldwide OCM, non-U.S. OCM, and student samples all consisted of more men (70.5%, 72.1%, and 56%, respectively). The education levels of participants in the worldwide and non-U.S. OCM samples tended to be higher than all other samples while those of the consumer panel and U.S. OCM sample were not significantly different. Interestingly, the demographic distributions from Study 2 (see Appendix C) were similar to those found in Study 1, which may indicate the distributions that may be expected from each sampling frame.

In Table 4, we provide the statistical assessment of the demographic differences and survey completion times via a series of t-tests of means, chi-square tests of proportions, and Wilcoxon sum–rank tests for categorical rank differences. As expected, we found significant differences in many of the comparisons among samples; however, interestingly, we found the student, consumer panel, and U.S. OCM samples were highly similar in many of the demographic distributions, especially given the drastic differences of the worldwide and non-U.S. samples. The demographic distributions of the worldwide and non-U.S. OCM samples are highly similar and

[15] At 60 days the Mechanical Turk listing was closed due to the time allowed for tasks to be posted on the marketplace at the time this study was conducted.

[16] We thank the anonymous reviewer for this recommendation.

Table 2. Analysis Overview		
Analysis Step	**Focus of Tests**	**Empirical Tests**
1. Demographics	Differences in sample compositions across demographic attributes.	T-tests, chi square, and Wilcoxon sum–rank.
2. Exploratory Factor Analysis	Differences in underlying factor structure of measurement items utilizing multiple factor rotation methods.	Principal components and maximum likelihood analyses with varimax and oblimin rotation.
3. Confirmatory Factor Analysis	Differences in the theoretical measurement models in reference to multiple model fit indices.	Lambda values, CFI, SRMR, RMSEA, Cronbach's alpha, composite rho, average variance extracted, and Fornell-Larcker test.
4. Structural Models	Differences in the theoretical relationships between constructs in reference to model fit indices and t-tests of coefficient differences.	Covariance-based and partial least squares structural equation modeling comparing CFI, SRMR, RMSEA, R^2, and t-test of coefficient comparison.
5. Group Invariances	Differences in sample intercepts, factor loadings, residual variances, and scale means utilizing a series of constrained structural models, ANOVAs, and pairwise comparison of each sample.	CFI, ΔCFI, Chi Square, ΔChi Square, ANOVA, and Scheffe's pair-wise comparisons of scale means.

exhibit what we might expect to find as the majority of OCM participants based on previous demographic evidence obtained from Mechanical Turk (Ross et al. 2010). Additionally, the demographic data for the students, consumer panel, and U.S. OCM samples were all fairly similar across many of the categories with the primary differences being the age, education levels, income, and family compositions. This may indicate that the U.S. OCM sample is a much closer approximation to the demographic compositions of the consumer panel on many of the collected attributes than the current student sample. However, comparison of demographic categories alone does not rule out the potential for coverage and nonresponse biases within web-based survey techniques (Couper 2000) and, therefore, requires the continued examination of measurement error, scale differences, and model relationships to determine the level and extent of biases.

Exploratory Factor Analysis

For the exploratory factor analysis, we used both principal components analysis and maximum likelihood with varimax and oblimin rotations ((DeVellis 2003; Kim and Mueller 1978; Nunnally and Bernstein 1994). Table 5 contains the results from the maximum likelihood varimax rotation.[17] In all five samples, the items loaded as expected on their focal constructs and less than 0.40 on any other construct. Thus, we retained all item indicators within each model to ensure consistency in the analyses between each sample. Interestingly, while the strength of some loadings varied across samples, based on this initial test one might infer that there are no significant differences among the samples. However, we conducted confirmatory factor analysis, structural model analysis, and group invariance tests to further explore any potential differences between the samples in regard to their psychometrics, structural models, and measurement invariances.

Confirmatory Factor Analysis

Establishing convergent validity requires meeting three criteria: (1) adequate model fit, (2) significant lambda values greater than 0.30, and (3) an average variance extracted (AVE) greater than 0.50 (Byrne 2010; Hair et al. 2006). Hu and Bentler (1999) further recommend comparative fit indices (CFI) of at least 0.95, standardized root mean square residual (SRMR) values less than 0.08, and root mean square error of approximation (RMSEA) of less than 0.06 to achieve acceptable model fit. Table 6 presents the loadings and model fit criteria for the CFA. The consumer panel provided the largest CFI (0.959), followed by the student sample (0.954), whereas the U.S. OCM sample is close to meeting this threshold (0.939); however, the non-U.S. OCM and worldwide OCM

[17]The other methods of extraction and rotation shared the same pattern of results.

Table 3. Demographics Distribution

		Distribution				
		Student	**Worldwide OCM**	**U.S. OCM**	**Non-U.S. OCM**	**Consumer Panel**
Gender*	Male	0.56	0.70	0.43	0.72	0.43
	Female	0.42	0.30	0.57	0.28	0.55
Age	Mean	23.00	29.00	32.00	29.00	44.00
	Median	21.00	26.00	28.00	27.00	46.00
	Minimum	18.00	18.00	16.00	17.00	19.00
	Maximum	48.00	62.00	68.00	63.00	69.00
Education Level*	Less than High School	0.00	0.01	0.01	0.00	0.01
	High School/GED	0.08	0.10	0.11	0.07	0.15
	Some College	0.63	0.13	0.35	0.12	0.26
	2-Year College Degree	0.08	0.10	0.12	0.07	0.13
	4-Year College Degree	0.14	0.36	0.32	0.42	0.28
	Masters Degree	0.05	0.29	0.07	0.31	0.13
	Doctoral Degree	0.00	0.01	0.01	0.02	0.02
	Professional Degree (JD, MD)	0.00	0.01	0.01	0.02	0.03
Race*	White/Caucasian	0.73	0.25	0.69	0.11	0.81
	African American	0.04	0.01	0.09	0.00	0.06
	Hispanic	0.04	0.02	0.03	0.02	0.06
	Asian	0.11	0.61	0.07	0.83	0.07
	Native American	0.01	0.01	0.00	0.00	0.01
	Pacific Islander	0.00	0.01	0.00	0.00	0.00
	Mixed/Other	0.05	0.10	0.10	0.04	0.02
Family Structure*	Single, no children	0.84	0.48	0.50	0.51	0.25
	Single, with children	0.00	0.02	0.07	0.01	0.11
	Married, no children	0.07	0.13	0.14	0.18	0.13
	Married, with children	0.04	0.28	0.19	0.26	0.45
	Life partner, no children	0.04	0.06	0.05	0.01	0.04
	Life partner, with children	0.00	0.03	0.05	0.02	0.02
Annual Income Range (in U.S. dollars)*	$19,999 >	0.73	0.52	0.29	0.58	0.09
	$20,000 – $29,999	0.07	0.18	0.21	0.18	0.07
	$30,000 – $39,999	0.06	0.07	0.17	0.10	0.13
	$40,000 – $49,999	0.01	0.10	0.08	0.05	0.10
	$50,000 – $59,999	0.01	0.04	0.10	0.01	0.12
	$60,000 – $69,999	0.02	0.03	0.03	0.02	0.07
	$70,000 – $79,000	0.01	0.02	0.02	0.03	0.08
	$80,000 – $89,999	0.04	0.02	0.04	0.02	0.20
	$90,000 <	0.01	0.02	0.04	0.02	0.12
Time Elapsed	Mean	12.68	11.01	9.99	12.27	13.54
	Median	11.00	10.00	10.00	11.00	10.59
	Std. Dev.	7.67	6.29	3.67	7.33	9.70
	Min.	2.00	2.00	2.00	2.00	2.00
	Max.	56.00	46.00	41.00	52.00	55.12

*Value displayed as percentage of total responses.

Table 4. Demographic Differences

		Demographic Comparisons									
		Student Vs. Worldwide	Student Vs. U.S.	Student Vs. Non-U.S.	U.S. Vs. Worldwide	Non-U.S. Vs. Worldwide	U.S. vs. Non-U.S.	Consumer Panel Vs. Student	Consumer Panel Vs. U.S.	Consumer Panel Vs. Non-U.S.	Consumer Panel Vs. Worldwide
Gender¹	Male	5.814***	7.959***	8.044***	30.926***	0.0722	37.025***	6.559*	0.084	35.034***	28.952***
	Female										
Age²	Mean	8.294***	11.043***	9.223***	3.912***	0.07	4.033***	25.337***	10.419***	16.704***	15.973***
Education Level²	Education Rank	9.654***	4.020***	12.874***	5.577***	2.202*	8.289***	5.54***	1.574	6.43***	3.924***
Race¹	White/ Caucasian	80.284***	0.344	148.941***	80.061***	12.785***	151.361***	3.502	7.881**	224.634***	138.302***
	African American	2.944	3.177	5.566*	12.851***	0.002	16.986***	0.694	0.934	11.014***	7.647**
	Hispanic	0.296	0.000	0.154	0.142	0.000	0.045	0.694	1.415	2.671	3.015
	Asian	92.605***	1.500	187.501***	134.902***	22.084***	246.75***	2.277	0.004	273.037***	152.859***
	Native American	0.015	0.059	0.773	0.000	0.002	0.000	0.000	0.130	0.999	0.050
	Pacific Islander	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.000
	Mixed/ Other	3.194	3.367	0.000	0.000	0.000	5.076*	2.702	15.656***	2.177	15.149***
Family Structure¹	Single, no children	54.441***	51.176***	49.184***	0.131	0.207	0.000	149.143***	33.106***	33.556***	25.95***
	Single, with children	0.000	4.298*	0.000	4.169*	0.016	6.498*	11.905***	2.451	16.367	12.518***
	Married, no children	1.914	2.072	6.51*	0.000	1.190	1.281	1.415	0.037	2.273	0.026
	Married, with children	30.453***	13.54***	25.938***	4.771*	0.224	2.670	72.232***	33.937***	16.34***	11.142***
	Life partner, no children	0.015	0.000	3.561	0.120	5.256*	3.293	0.218	0.112	1.822	0.814
	Life partner, with children	0.550	1.935	0.000	0.177	0.593	2.209	0.002	1.664	0.000	0.269
Annual Income Range	Mean Difference²	2.849**	6.661***	1.888	3.860***	1.087	5.085***	16.682***	9.988***	15.494***	13.925***
	Categorical Rank³	11798***	9299.5***	13949***	26550***	21655	30927***	35639***	41276***	45386***	40012***
Time Elapsed	Mean	2.246*	4.188***	0.522	1.958	1.874	4.067***	1.014	5.412***	1.604	3.346***

Notes: ¹Chi-square proportion, ²Mean difference t-test, ³Wilcoxon Sum-Rank test, *p-value < .05; **p-value < .01; ***p-value < .001

Table 5. Exploratory Factor Analysis

	Students			Worldwide OCM			U.S. OCM			Non-U.S. OCM			Consumer Panel		
	Perceived Usefulness	Perceived Ease of Use	Behavioral Intention	Perceived Usefulness	Perceived Ease of Use	Behavioral Intention	Perceived Usefulness	Perceived Ease of Use	Behavioral Intention	Perceived Usefulness	Perceived Ease of Use	Behavioral Intention	Perceived Usefulness	Perceived Ease of Use	Behavioral Intention
Usefulness 1	0.709			0.745			0.845			0.699			0.820		
Usefulness 2	0.790			0.756			0.839			0.722			0.866		
Usefulness 3	0.883			0.803			0.870			0.775			0.872		
Usefulness 4	0.777			0.776			0.838			0.798			0.869		
Usefulness 5	0.736			0.751			0.767			0.734			0.774		
Usefulness 6	0.669			0.712			0.737			0.697			0.704		
Ease of Use 1		0.725			0.678			0.825			0.720			0.868	
Ease of Use 2		0.798			0.629			0.854			0.644			0.843	
Ease of Use 3		0.825			0.707			0.831			0.739			0.854	
Ease of Use 4		0.662			0.674			0.706			0.750			0.764	
Ease of Use 5		0.642			0.764			0.845			0.702			0.852	
Ease of Use 6		0.773			0.731			0.789			0.766			0.865	
Intention 1			0.960			0.764			0.879			0.764			0.929
Intention 2			0.843			0.746			0.773			0.742			0.844
Intention 3			0.929			0.846			0.911			0.712			0.923

Note: Maximum Likelihood extraction with varimax rotation. Loadings less than 0.40 removed for clarity.

Table 6. Confirmatory Factor Analysis

Measurement Item	Students	Worldwide OCM	U.S. OCM	Non-U.S. OCM	Consumer Panel
Usefulness 1	0.832	0.817	0.912	0.812	0.917
Usefulness 2	0.898	0.814	0.923	0.846	0.939
Usefulness 3	0.738	0.879	0.921	0.785	0.945
Usefulness 4	0.798	0.841	0.903	0.865	0.933
Usefulness 5	0.784	0.797	0.810	0.785	0.840
Usefulness 6	0.710	0.765	0.789	0.808	0.794
Ease of Use 1	0.810	0.735	0.889	0.732	0.922
Ease of Use 2	0.858	0.808	0.916	0.801	0.916
Ease of Use 3	0.675	0.835	0.826	0.766	0.937
Ease of Use 4	0.744	0.790	0.818	0.836	0.879
Ease of Use 5	0.715	0.773	0.853	0.778	0.915
Ease of Use 6	0.814	0.833	0.852	0.870	0.930
Intention 1	0.884	0.910	0.880	0.751	0.981
Intention 2	0.977	0.810	0.987	0.815	0.908
Intention 3	0.965	0.869	0.967	0.866	0.981
CFI	0.954	0.901	0.939	0.916	0.959
SRMR	0.059	0.047	0.052	0.050	0.035
RMSEA	0.078	0.160	0.109	0.103	0.098

Note: All loadings at $p < 0.001$

samples achieved the lowest values with 0.916 and 0.901, respectively. All five samples were adequately below the SRMR threshold. For RMSEA, the student sample came closest, with a value of 0.078, and the consumer panel, U.S. OCM, and non-U.S. OCM samples indicated similar values (0.098, 0.109, and 0.103, respectively). However, the worldwide OCM sample's RMSEA of 0.16 was far greater than the recommended reliability threshold, potentially indicating some significant issues within the current measurement model. Our analysis thus indicates that the samples did not meet all of the CB-SEM fit indices perfectly; however, many of the estimates are close to the recommended thresholds, which is promising given that there were no model adjustments within our CB-SEM. As mentioned previously, we did not conduct any supplemental analyses to determine the potential for outliers. Additionally, we did not make any changes to the CB-SEM models via recommended modification indices, which would further increase the fit indices of the model (Byrne 2010). Therefore, some deviations from the recommended thresholds should be expected, to a certain extent. While we believe that further refinement of the theoretical models and parameter estimates would increase the model fit to the point of achieving adequate standards in some samples (students, consumer panels, and U.S. OCM), the repeated lack of model fit for the worldwide and non-U.S. OCM samples causes hesitation in the use of these specific samples. Within the PLS analyses, each sample met all recommended thresholds for reliability as well as convergent and divergent validity without any further alterations to the data or models.

Another criterion for convergent validity is that the lambda values should be at least 0.30 and significant (Hair et al. 2006). The lambda values in Table 6 indicate that the loadings of each item within each sample exceed these assessments at $p < 0.001$. A third criterion holds that the AVE should be greater than or equal to 0.5 (Hair et al. 2006; Kline 2010), to ensure that the shared variance of the measures is greater than the variance associated with measurement error (see Table 7). All constructs in each sample met these criteria, in empirical support of convergent validity.

To assess discriminant validity, we used Fornell and Larcker's (1981) test and compared the square root of the AVE with the off-diagonal correlations where the square root of the AVE exceeded all off-diagonal correlations in all samples' correlation matrices. The correlation matrices in Table 7, therefore, support both convergent and divergent validity.

Scale reliability was evaluated via both Cronbach's alpha and composite rho values for each construct (see Table 7). Reliability estimates exceeding 0.70 are sufficient and recommended (Hair et al. 2006). Because the measurement scales of all samples achieved reliability scores greater than 0.70, we found adequate support for measurement reliability for each sample. The results were consistent for the PLS analyses where each criterion was met or exceeded (see Appendix D). Thus, from an examination of simply the demographics and psychometrics through common research practices, one could assume some level of validity for all samples. While the worldwide and non-U.S. OCM samples did show weaker fit indices than the student, consumer, and U.S. OCM samples, there is the possibility that some alterations to the modification indices or outlier analysis may address these issues. Additionally, the robustness analysis within PLS passed all convergent and divergent validity as well as reliability thresholds without any issues, leading a researcher who utilizes PLS to assume similar data quality across all samples. However, we continued in the following structural model analysis to conduct a deeper examination and uncovered an interesting pattern of differences that exist in some samples, which would have been overlooked had we stopped at the psychometrics in our exploration as in some prior research.

Structural Model

In Figure 2 we present the estimated structural path models utilizing CB-SEM[18] and a test of the path coefficient differences for each of the five samples. The SRMR fit indices were acceptable in each sample (Hu and Bentler 1999), whereas the CFI and RMSEA estimates were slightly beyond the preferred limits. The consumer panel, student, and U.S. OCM samples met or approached the CFI threshold; however, the worldwide and non-U.S. OCM samples again strayed further from preferred acceptability levels. The RMSEA values exceeded acceptability thresholds in all the samples; however, they were relatively similar, indicating comparable model fit indices between the samples. Similarly as in our CFA, our structural models were not adjusted for outliers or modification indices to improve model fit; therefore, they provided a direct indication of the original data quality from the separate data collections. We could potentially alleviate the deviations from the recommended model fit indices by cleaning outliers, transforming variables, or adjusting the SEM model via the recommended modification indices (Byrne 2010; Kline 2010). For example, both the worldwide

[18]We also ran the entire set of analyses utilizing PLS to account for smaller sample sizes in some of the samples collected. Appendix D provides the full details of this analysis, which closely matches that of our primary CB-SEM analysis. We thank the anonymous reviewer for suggesting this robustness analysis.

Table 7. Correlations and Reliabilities

Students	Mean	SD	Cronbach's Alpha	Composite Rho	AVE	PU	PEOU	BI
Perceived Usefulness	5.589	0.839	0.910	0.911	0.63	**0.796**		
Perceived Ease of Use	5.678	0.736	0.897	0.898	0.60	0.479	**0.772**	
Behavioral Intention	5.776	1.609	0.959	0.960	0.89	0.384	0.331	**0.943**
Worldwide OCM								
	Mean	**SD**	**Cronbach's Alpha**	**Composite Rho**	**AVE**	**PU**	**PEOU**	**BI**
Perceived Usefulness	5.783	0.907	0.925	0.925	0.672	**0.820**		
Perceived Ease of Use	5.688	0.762	0.912	0.912	0.635	0.695	**0.797**	
Behavioral Intention	5.749	0.995	0.891	0.898	0.746	0.554	0.689	**0.864**
U.S. OCM								
	Mean	**SD**	**Cronbach's Alpha**	**Composite Rho**	**AVE**	**PU**	**PEOU**	**BI**
Perceived Usefulness	5.549	1.190	0.953	0.953	0.771	**0.878**		
Perceived Ease of Use	5.953	1.189	0.943	0.944	0.739	0.553	**0.860**	
Behavioral Intention	5.232	1.723	0.959	0.962	0.894	0.608	0.501	**0.946**
Non-U.S. OCM								
	Mean	**SD**	**Cronbach's Alpha**	**Composite Rho**	**AVE**	**PU**	**PEOU**	**BI**
Perceived Usefulness	5.566	0.869	0.923	0.923	0.668	**0.817**		
Perceived Ease of Use	5.151	0.704	0.913	0.913	0.638	0.696	**0.799**	
Behavioral Intention	5.750	1.008	0.855	0.853	0.660	0.624	0.608	**0.812**
Consumer Panel								
	Mean	**SD**	**Cronbach's Alpha**	**Composite Rho**	**AVE**	**PU**	**PEOU**	**BI**
Perceived Usefulness	5.730	1.288	0.961	0.961	0.804	**0.897**		
Perceived Ease of Use	5.670	1.211	0.969	0.969	0.841	0.652	**0.917**	
Behavioral Intention	5.613	1.374	0.970	0.970	0.916	0.518	0.481	**0.957**

Note: Square-root of the AVE on diagonal

and non-U.S. OCM samples have multiple modification indices exceeding 10.0 which indicate potential improvements for model fit. However, as our focus is the differences in the initial data samples, we have left the models identical across all samples to identify any further differences that may exist.

Based on the path coefficients, most of the theorized relationships in each of the models were significant and in the predicted directions at the $p < 0.001$ level. However, the worldwide OCM sample produced an interesting non-significant path between perceived usefulness and behavioral intention, which contradicts most TAM studies (King and He 2006). This result refuted one of the most well-justified relationships in all of MIS, potentially indicating that if Davis (1989) had utilized this worldwide OCM sample we may not have the TAM model as it exists today.[19] Table 8 also reveals a few significant differences between the structural path coefficients of the worldwide OCM sample and the additional

[19]We thank the AE for their insightful discussion of this issue.

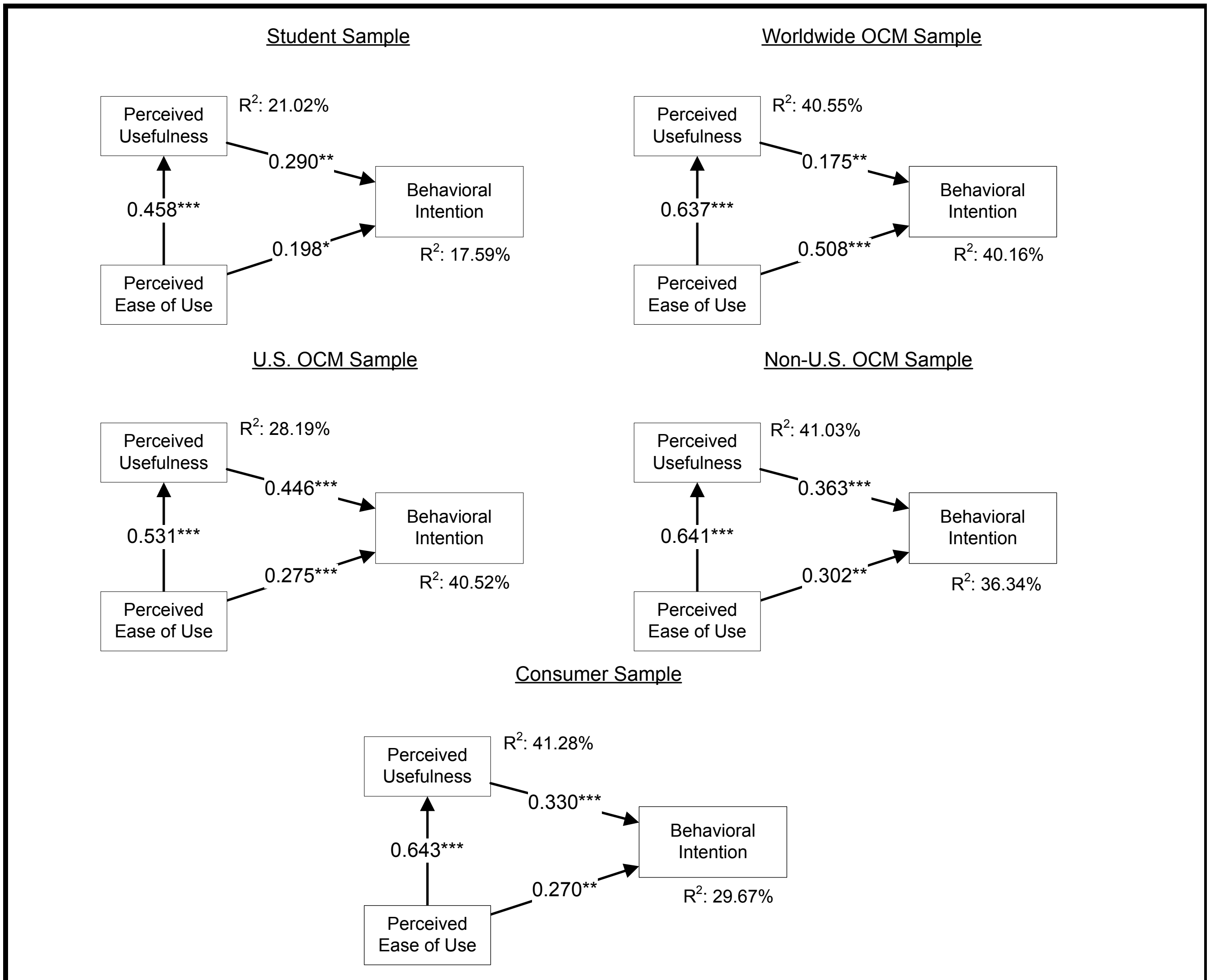


Figure 2. Structural Model Results

Table 8. Path Coefficient Differences

Comparison	PEOU ➜ BI	PEOU ➜ PU	PU ➜ BI
Student vs. U.S.	-0.046(0.815)	-0.074(0.490)	-0.184(0.311)
Student vs. Non-U.S.	-0.146(0.520)	-0.217(0.094)	-0.097(0.614)
Student vs. Worldwide	-0.397(0.073)	-0.216(0.085)	0.147(0.432)
Student vs. Consumer	-0.059(0.786)	-0.173(0.180)	-0.063(0.729)
U.S. vs. Non-U.S.	-0.100(0.548)	-0.143(0.189)	0.087(0.557)
U.S. vs. Worldwide	**-0.351(0.025)**	-0.142(0.175)	**0.331(0.019)**
U.S. vs. Consumer	-0.013(0.917)	-0.099(0.243)	0.121(0.325)
Non-U.S. vs. Worldwide	-0.251(0.127)	0.001(0.993)	0.244(0.078)
Non-U.S. vs. Consumer	0.087(0.573)	0.044(0.669)	0.034(0.797)
Worldwide vs. Consumer	**0.338(0.020)**	0.043(0.665)	-0.210(0.092)

Notes: p-value for two-tailed test in parentheses.

Table 9. Worldwide Robustness Analysis

Structural Path	Worldwide n = 193	Worldwide Without U.S. n = 173	India Only n = 125
PEOU ➜ BI	0.588***	0.560***	0.371**
PEOU ➜ PU	0.695***	0.764***	0.776***
PU ➜ BI	0.146	0.168†	0.373**
Model Fit			
PU – R^2	0.483	0.584	0.602
BI – R^2	0.486	0.486	0.492
CFI	0.901	0.897	0.890
RMSEA	0.116	0.113	0.113
SRMR	0.047	0.050	0.053

Note: †*p* < 0.1; **p* < .05; ***p* < .01; *** < .001

four samples.[20] Interestingly, while only 3 out of 30 path comparisons differed significantly across all 5 samples at $p < 0.05$, the differences all involve the worldwide OCM samples such that the effect of perceived usefulness and perceived ease of use on behavioral intentions differs from that of the U.S. OCM sample and the impact of perceived ease of use on behavioral intentions differs with the consumer panel.

Therefore, in an attempt to further investigate the worldwide OCM sample's deviation from prior theory, we took a closer look at the data in this sample to determine possible reasons for the nonsignificant findings of perceived usefulness. The differences and nonsignificance of the worldwide OCM may be in part due to coverage, nonresponse, or selection bias influencing the underlying estimates (Groves 1989). We began by examining in detail the location of each respondent within the worldwide OCM sample to determine any potential demographic differences between respondents. Interestingly, although the call for participation was worldwide, the majority of the participants in the worldwide OCM sample were located in a single country, India, which may potentially have led to varying interpretations of the (1) questionnaire items, (2) focal technology, and (3) perceptions compared to individuals in other countries. Therefore, we began our robustness analysis by eliminating all of the respondents from the sample who were from the U.S. (20 respondents), making the subset of data conceptually similar to the non-U.S. OCM sample. We reanalyzed the structural models utilizing this subset and found that path coefficients became slightly closer to the expected effects found in prior literature yet only barely significant ($p < 0.10$). We then further reduced this subset to only those individuals located in India to examine the effects of a sample with a consistent country of origin. This subset would be conceptually similar to the single-country restriction of our U.S. OCM sample albeit with an alternative country of origin. This final subset provided results that were significant and in line with prior theory, indicating that the potential differences in individuals in various countries may have masked the relationships within the model. Table 9 provides the results from this robustness analysis which we ran in both CB-SEM as well as PLS with similar results.[21] Running this robustness analysis and uncovering the theoretical relationships required the capture of alternative respondent attributes and propensity variables to adjust for potential biases (Couper 2000; Malhotra 2008).

Group Invariance Tests

The final step in our comparisons involved CB-SEM group invariance tests, which we used to examine measurement variances for each model in detail. That is, we tested where the samples show invariance, or no difference, in their various measurement estimates (Pentz and Chou 1994; Steenkamp and Baumgartner 1998). To assess differences between models, we used the ΔCFI statistic, which is robust and more accurate than the χ^2 difference test (Cheung and Rensvold 2002; Kline 2010), which suffers from sensitivity to the sample size.

[20]Using a two-tailed t-test of differences (Chin 2000):

$$t = \frac{Path_{sample\_1} - Path_{sample\_2}}{\left[\sqrt{\frac{(m-1)^2}{(m+n-2)} * S.E._{sample1}^2 + \frac{(n-1)^2}{(m+n-2)} * S.E._{sample2}^2}\right] * \left[\sqrt{\frac{1}{m} + \frac{1}{n}}\right]}$$

[21]We additionally utilized dummy variables for country categories and found similar results.

We examined differences in increasingly constrained measurement models, on the basis of the factor loadings, intercepts, residuals, and means, to determine at which level the measurements varied. Appendix E contains an in-depth description and the results of this multistep, iterative process. The models were equal in their factor loadings and intercepts across all samples, similar to those results found in the EFA. However, as we began to examine the variances and covariances within each sample, subtle differences emerged in the sample comparisons. At this level of analysis, we found differences in the comparisons with the exception being the worldwide and U.S. OCM samples remaining invariant to each other as well as the student and U.S. OCM samples remaining invariant. Interestingly, the consumer panel produced differences in the variances compared to all the other samples collected, indicating potential dissimilarities in the response measurement for this sample as well.

Furthermore, while some comparisons showed invariances between samples, many showed differences in scale means within some aspect of the models. Therefore, to provide a clearer picture of which scales are actually differing across each sample, we conducted a series of ANOVA and pair-wise comparison tests with a Scheffe's correction. The results indicate that perceived usefulness does not differ between the five samples ($p > 0.05$) while perceived ease of use ($p < 0.001$) and behavioral intention ($p < 0.001$) do show slight differences. The mean value of perceived ease of use for the non-U.S. OCM sample differs from those of all other samples ($p < 0.001$) while the U.S. OCM sample's mean value only slightly differs from those of the consumer panel ($p < 0.05$). Additionally, the mean value of behavioral intention for the U.S. OCM sample differs from that of all other samples ($p < 0.01$) except for the consumer panel.[22] One thing to note from this subset analysis is that while many of the comparisons show no differences, the differences that do exist typically involve the worldwide and non-U.S. OCM samples differing significantly from the student, consumer panel, and U.S. OCM samples. For detailed results refer to Appendix E.

The results from the series of tests conducted within this study provide important insights into the viability of OCMs for use in academic research. Specifically, we found repeated issues with the worldwide and non-U.S. OCM samples in regard to their CFA model fit indices, structural model fit indices, measurement and scale invariances, and, most importantly, the complete lack of significance of the perceived usefulness–behavioral intention relationship in the TAM model. However, we typically noticed consistent results between the student, consumer panel, and U.S. OCM samples, indicating the potential for interchangeability among these samples. One interesting aspect is that had we not explored the details of these differences beyond demographics and psychometrics, these results might not have been identified as the psychometrics (reliability, convergent validity, divergent validity, and the factor loadings) did not generally differ among the samples and met all validity thresholds within PLS compared to CB-SEM. This is aptly evident as prior research has found that "the data obtained are at least as reliable as those obtained via traditional methods" (Buhrmester et al. 2011, p. 3), which is obviously not the case in our studies. Thus, while the decision to utilize OCMs as a viable recruitment technique is dependent on the focus of the research parameters to be estimated (i.e., mean values or theoretical relationships) and the potential differences that may exist based on the composition of the sample, we must caution researchers in utilizing the worldwide and non-U.S. OCM samples at this time until further research is conducted to explore the underlying cause of the present differences.

Discussion

The purpose of this research was to assess potential differences in the demographics, psychometrics, and structural properties across online crowdsourcing markets (OCM) and more traditional samples such as consumer panels and college students. Researchers have begun to increase their use of OCMs to collect high quality data, rather than relying only on more traditional sampling frames (Leonhardt et al. 2011; Mason and Suri 2012; Paolacci et al. 2011; Paolacci et al. 2010). Accordingly, we analyzed both the technology acceptance model and the expectation–disconfirmation model with five and four distinct samples, respectively, to assess the validity and reliability of using OCMs as a sampling frame.

The initial results indicated that the psychometric properties of the measurement scales across all TAM and EDT samples remain largely valid and reliable in both CB-SEM and PLS techniques. Specifically, the scales showed similar results in terms of convergent validity, discriminant validity, and reliability. However, during a deeper examination, we found significant differences in the path coefficients and the group invariance tests such that the theorized paths for our worldwide OCM sample in Study 1 provided inconsistent results with theory and the non-U.S. OCM sample in Study 2 provided differences across multiple paths. While in Study 1, after controlling for sample differences within our worldwide OCM sample, the structural models and corresponding coefficients were not significantly different, in Study 2 this

[22] U.S. OCM differs from consumer panel ($p > 0.05$).

same *post hoc* analysis actually created more problems than it solved when exploring the differing non-U.S. sample. Therefore, based on these issues with potential data quality, we currently caution researchers about using an unrestricted worldwide and non-U.S. OCM sample due to a variety of biases that may exist and are not captured within this research.

Despite the issues that arose in utilizing worldwide and non-U.S. samples, our data suggest that the use of U.S. OCMs for academic research of U.S. populations is a potentially promising and acceptable method of data collection if researchers follow a few quality precautions. To provide consistent, comparable, and valid results, though, studies must address and report on certain characteristics of their OCM samples. In particular, we found strong demographic differences in our worldwide and non-U.S. OCM samples compared with the student, consumer panel, and U.S. OCM samples. This could indicate specific demographic measures to capture in order to ensure interpretable and comparable results. For example, if the underlying theory suggests that cultural or demographic factors may have an effect, researchers should implement additional demographic controls and restrictions to address and control for such issues (Leidner and Kayworth 2006; Straub et al. 2002).

Research Implications

This research provides the initial empirical evidence that OCMs may be viable sources of potentially large samples of participants for conducting academic research and can reach a much larger variety of demographics within the population. However, such studies utilizing OCMs must undertake much more stringent examinations of the demographics, validity, and reliability to ensure the consistency of the results and procedures with previous research (Meyerson and Tryon 2003). Our study provides some guidelines and procedures for the initial steps for validating the psychometrics such as those of the original TAM and EDT models. Although we find initial support for employing U.S.-based OCM participants to measure both the TAM and EDT constructs, further studies must continue to address unexamined constructs and experimental techniques in OCM environments to provide similar empirical support. Validation tests should include comparisons with previous studies, sampling frames, and techniques to offer supplementary evidence of the validity of using OCM to support research (Buchanan and Smith 1999).

Another implication is that researchers may need to be selective regarding the demographic attributes of OCM members who participate in their studies. As our findings indicate, demographic differences among OCM participants are much more varied than those of a typical college student population. Demographics can have potentially significant influences on relationships within TAM models (King and He 2006; Venkatesh et al. 2003) and such demographic information must be addressed and reported in detail by researchers who use OCMs. A failure to address this issue could lead to masked or biased data, results, and interpretations, such as in our worldwide OCM sample within Study 1, especially if demographic or cultural factors are likely to influence the theoretical relationships. For example, while many theories may be culturally sensitive, researchers often fail to examine the data across multiple cultures. This use of OCMs could potentially provide the ability to expand and contextualize the theoretical contributions by strategically targeting differing cultures and examining the varying relationships that may emerge. While demographics may play some part in the differences that we found, we cannot conclude that demographics alone are the problem with the worldwide OCM sample in Study 1 and the non-U.S. OCM sample in Study 2. For example, the *post hoc* analysis in Study 2, which limited responses to a single country of origin, created more model differences than were originally present.

Due to the issues that have emerged from this research we urge caution in the use of non-U.S. OCM samples based on two general reasons. First, the responses provided by non-U.S. OCM participants clearly provide different conclusions than those of the U.S. populations collected in this study. To remain objective and clarify our interpretations, we acknowledge that while we cannot claim that the U.S. responses are the "correct" responses within this study, the focus of our comparison for generalization was to our U.S. student and consumer samples. Recent research has shown that there exist exceptional culture differences across individuals, with U.S. participants being "exceptional even within the unusual population of Westerners—outliers among outliers" (Henrich et al. 2010, p. 76). That being said, had we collected a student and consumer panel from India, we may have found that the conclusions between the non-U.S. OCM samples and these alternatives may have not significantly differed. While we caution the use of these samples, we cannot claim that they are "wrong" as our research has not fully explored their potential and is left as an avenue for future research. Second, the demographics of the non-U.S. OCM samples do not seem to generalize even to their own countries by appearing more highly educated, wealthier, and generally more male than the countries in which they live. Thus, it is not clear exactly which populations non-U.S. OCM samples can be generalized and until it is clear, we caution researchers in their use. It is clear that future research must be conducted to explore the factors creating differences between samples other than simply demographic compositions. However, as a unique advantage, OCMs provide extensive controls over the selection of participants from a larger population to fit a defined sampling frame, which is often not the case for traditional student samples or cost effective for targeted consumer panels.

OCMs also provide researchers with the potential ability to create highly detailed filters and requirements to solicit only those respondents that directly fit the researcher's intended sampling frame and reduce the concerns that have arisen from unrestricted participant access. While in our study we utilize only a simple country of origin filter, advancements in OCMs have grown to include much more detailed filters that can be implemented in a research design such as reading comprehension, background checks, and knowledge pretesting. Thus, while researchers have the ability to control their participant selection, they also have the responsibility to report and justify any restrictions set in place by each OCM selection criterion. Most studies reveal only a limited selection of demographic properties, such as gender or age (e.g., Bagchi and Li 2011; Leonhardt et al. 2011; Sakamoto and Bao 2011; Yu and Nickerson 2011), but our results show that the variety of demographic differences is much greater for an OCM frame compared with student sampling frames. Similarly, if studies have detailed filters in place, the resulting demographic composition should be closely examined and reported to confirm the success of these mechanisms. As our study shows, measures and responses from people in various demographic categories and locations exhibit differences in item variances as well as potential interpretations across cultures (Henrich et al. 2010), which could lead to inconclusive theoretical evidence, such as exhibited in our worldwide OCM sample in Study 1. Due to these current results and the ability to limit the sampling frame within OCMs, we recommend that researchers exhibit caution when using worldwide and non-U.S. OCM samples until further research has provided more information on the potential biases that may still exist.

Researchers should clearly report specifications and designs of their recruiting efforts in OCM frames, including the variety of settings to select and motivate participants. For example, some studies have explored the potential for differences among online samples such as incentive levels, which have been shown to significantly influence response results in both traditional samples (Groves et al. 2000) and OCMs (Mason and Suri 2012). Therefore, researchers should clearly report the incentive level they offered to allow for consistent replications and comparisons across studies. Due to the inability to capture response rates in many web surveys (Couper 2000), researchers should also conduct detailed analyses of the completion rates and time needed to complete each response, along with quality, attention, and human verification controls (Aust et al. 2012; Malhotra 2008). Furthermore, the great potential for fictitious or automated answers in online studies suggests the need for comparisons with previous tests of known, valid responses, to define reliable response times. Finally, web-based studies require more scrutiny than laboratory studies because researchers typically cannot control the environment (Vadillo and Matute 2011). Thus, publications should include descriptions of all participant recruitment procedures, participant restrictions, survey procedures, and data cleaning efforts to provide evidence of reliable results from OCM samples. To support consistent replications and interpretations of empirical results, in Table 10 we offer an overview of information that ideally should be captured and reported when studies utilize OCMs.

From a logistics point of view, using OCMs to conduct academic research also increases a researcher's ability to conduct repetitive iterations of model testing and scale development, thus generating stronger methodological techniques and theory. Because OCMs are considerably cheaper and support quicker response rates than traditional student sampling (Mason and Suri 2012), the execution of multiple, iterative studies becomes much more feasible. Moreover, the OCM population is much larger than student populations, so the chance of tainted sampling pools that have completed previous surveys or experiments diminishes. With this support, researchers can generate stronger empirical tools with extensive validation before attempting to employ more mainstream or limited sampling frames, such as organizational employees.

Limitations and Future Research

Although our examination provides the initial support for the validity of using U.S. OCM sampling frames for academic research, it also has several limitations. First, we have examined only two theoretical models (TAM and EDT), although we chose extensively validated, well-known models that have appeared in multiple contexts (King and He 2006), which allowed us to achieve clear expectations and comparisons of the psychometric properties (Buchanan and Smith 1999). This approach also gives us confidence that the differences among samples were due to the method and technique, not the theoretical model (Burton-Jones 2009). Therefore, we have provided an initial validation and justification of the usefulness of OCMs, specifically U.S. OCMs, by providing evidence of their consistency and reliability compared to our traditional student and consumer panels when utilizing specific OCM filtering mechanisms and capturing or controlling for specific OCM attributes. However, a single study

Table 10. OCM Reporting Recommendations

1. Participant demographics
    a. Country of origin distribution
    b. Income
    c. Age
    d. Gender
    e. Marital status
    f. Employment status
2. Participation restrictions
    a. Location or country of origin
    b. Computer system requirements
    c. Survey experience
3. Payment incentives
    a. Monetary specifications
    b. Bonus incentive conditions
4. Task timeline
    a. Average time to completion
    b. Minimum and maximum time to completion
5. Data quality questions and checks
    a. Human verification tasks
    b. Attention verification tasks
    c. Embedded software techniques (e.g., CAPTCHA)
    d. Description of correct vs. failed responses
6. Detailed data cleaning procedures
    a. Number of responses received before cleaning
    b. Checks for compliance with participant restrictions (e.g., country of origin validation)
    c. Description of protection from previous survey responses

cannot completely alleviate validity questions, which require repeated examinations by researchers who employ programmatic research to provide further support and evidence (Burton-Jones 2009). Future research should examine the consistency of their theoretical models and relationships with those that have been previously validated in traditional sampling frames before accepting the results at face value. Additionally, without capturing a variety of alternative participant attributes, the ability to control for differences via *post hoc* adjustments is limited (Couper 2000). Therefore, future research utilizing OCMs should capture extensive participant attributes to allow for a deeper examination of the responses that may have masked the theoretical results, as in our worldwide OCM sample.

Second, we investigated only the use of a survey questionnaire, which did not require experimental techniques or controls for participation. Our results thus are specific to questionnaire studies, although researchers can run experimental and quasi-experimental tasks online through OCMs (i.e., Sprouse 2011; Yu and Nickerson 2011) and may conclude similar findings from a carefully designed procedure. Utilizing more controlled experimental techniques within OCMs may require significant additions and design considerations to address differences in computer hardware, software, and Internet connectivity (Benfield and Szlemko 2006). Future research should examine the validity and merit of utilizing OCMs for alternative experimental and qualitative designs, which may include their own intricacies and provide insight into additional required reporting recommendations.

Third, we examined OCM responses to only a relatively small and consistent payment level within each study. The leverage–saliency theory (Groves et al. 2000) indicates that various design features can have significant impacts on the composition of respondents. These design features have implications on the level of coverage, selection, and nonresponse error, in addition to the quality, validity, and reliability of the results. While increases in the incentive levels may increase overall response rates for respondents in various locations, they may also alter the self-selection of individuals participating by attracting participants uninterested in the topic (Groves 2006), thus altering the motivation within the study. In addition, payment levels lower than those we offered generally may not work for longer survey questionnaires, as lower payment levels may generate greater participant dropout rates. Interestingly, changes to the incentives for online crowdsourcing participation show initial evidence of increases in the number of tasks completed, but not in the quality of work (Mason and Suri 2012). Additionally, while the payment level among the OCMs remained constant in Study 1, the incentives between the consumer panels ($5.25), student sample (prize drawing), U.S. OCM ($0.20), and non-U.S. OCM ($0.20) sample varied while their structural model results did not significantly differ despite the range of incentive levels. It may be that incentive levels do not have a significant impact on all research topics. However, due to our limited examination of incentive effects through OCMs, we believe that future research should more closely examine the effect of incentive levels on not only the quantity and quality of responses but also the psychometric and structural proprieties.[23]

Finally, researchers should examine OCMs' techniques for addressing various coverage, nonresponse, and selection biases, allowing more for generalizable research. To address the intricacies of the demographic differences, a researcher should examine the possibility to utilize design controls, data partitioning, or *post hoc* adjustments such as sample weighting to a target population (Malhotra 2008). Unfortunately, using nonprobability samples limits the ability of researchers to draw statistical inferences; however, the use of weighted adjustments may improve the representiveness of the sample and thus the ability to make inferences (Bethlehem and Biffignandi 2012). While *post hoc* weighting has seen extensive use, it has been met with questionable success (Rookey et al. 2008). While the use of *post hoc* adjustments may be beneficial in some scenarios, they still rely on the researcher's collection of alternative variables that allow for adjustment as well as those that may influence response propensity and selection (Couper 2000).

While this study has provided a foundation for the use of OCMs in academic research, it also brings to light the need for additional research to ensure the consistency of results across a variety of conditions. The goal of this study was ***not*** to be an all-inclusive justification of OCMs but to introduce, examine, and provide support for further investigation and testing of these recruitment techniques, which are emerging within our own and allied disciplines (see Appendix A). We challenge researchers to evaluate and make their own decisions on the potential of OCMs for their own theoretical work, which may improve from repeated iterations of data collection with larger and more diverse samples through OCMs.

[23]We thank the anonymous reviewer for this suggestion.

Practical Implications

This study and those that have previously utilized OCMs for academic research provide the potential to influence the development and improvement of the OCMs themselves by indicating tools and techniques that can be utilized for better research designs. For example, using filters which select participants based on attributes such as age, gender, income, or marital status may improve result generalizability by generating probability samples targeted toward specific demographic distributions. Survey and online experimental designs developed by researchers could lead to improved tool integration within OCMs to allow for easier development and inclusion of such tools in future research endeavors.

Furthermore, this study provides insights into the demographic differences that may be relevant for practitioners who utilize OCMs in their own outsourcing of tasks. The most current OCM demand is by practitioners and it would be highly beneficial for the users of these services to know of potential biases in worker responses. Hopefully, this research will help requesters gain better quality results and understand the potential differences that may exist with different study designs and demographic restrictions. More importantly, we see the continued and growing use of OCMs within academic research as a reciprocal development between researcher needs and OCM capabilities to lead to higher quality technologies and tools across OCMs.[24]

Conclusion

As publication pressures have grown for researchers in a variety of fields (Dean et al. 2011), competition for participant resources has increased as well. The ability to gather external data thus is limited, especially for researchers in smaller universities (Mason and Suri 2012), who therefore need to explore alternative methods for participant recruitment (Smith and Leigh 1997). Although student samples are convenient and often accessible, they may be less demographically diverse and exhibit lowered motivation, and the increased competition among researchers has strained even student resources (Gordon et al. 1986; Peterson 2002; Thomas 2011). Additionally, the use of expensive online consumer panels may not be fiscally possible for many researchers. Therefore, online crowdsourcing markets (OCM) may be a viable alternative: They provide instant access to a large, highly diverse (Ross et al. 2010), and motivated (Kaufmann et al. 2011) population of participants from across the globe, with relatively minimal recruitment costs (Paolacci et al. 2010).

[24]We thank an anonymous reviewer for this addition.

By examining the well-known technology acceptance model across five distinct samples (consumer panels, college students, worldwide OCM, U.S. OCM, and non-U.S. OCM) as well as the expectation–disconfirmation model across four distinct samples (college students, worldwide OCM, U.S. OCM, and non-U.S. OCM) through nine data collections, we provide initial evidence of the validity and reliability of OCMs. We found, specifically, that U.S. OCMs are a viable alternative sampling frame for the recruitment of U.S. participants. The demographics of each sample differed significantly, as expected (Ross et al. 2010), but interestingly their psychometric properties were fundamentally the same. In our CB-SEM confirmatory factor analysis, we evaluated reliability, convergent validity, and divergent validity; we also performed a CB-SEM path analysis and group invariance tests between samples.[25] While our group invariance tests indicate differences in the variance and means between some samples, we found no major differences in the initial psychometrics of the measurement models. Had our exploration stopped at this point, one might have inferred that quality levels between samples were equivalent; however, a deeper exploration uncovered some important and interesting differences.

When we began to examine the structural models implied by each theory in Study 1 and Study 2, the differences between samples became much more evident. In Study 1, while only 3 out of 30 path comparisons were different among samples, the worldwide OCM sample produced these differences with the relationship between perceived usefulness and behavioral intention being nonsignificant. This finding alone was unexpected because this relationship is found in nearly all TAM studies. Through a *post hoc* analysis in Study 1, we isolated these differences to potential demographic issues masking the relationship; however, in Study 2, this same procedure created more problems with the differing non-U.S. OCM sample than were originally present. Alternatively, the students, consumer panel, and U.S. OCM samples were all highly similar in the majority of the tests utilized, with the U.S. OCM sample exhibiting a closer approximation to the consumer panel than a student sample did. Both Study 1 and Study 2 provide the initial support for the use of U.S. OCMs for participant recruitment by providing results and statistical conclusions similar to that of both college students and consumer panels from the United States. However, based on the unexplained biases and differences that exist within the worldwide and non-U.S. OCM samples, we express caution to researchers in utilizing these unrestricted sampling approaches until further research is conducted to determine the causality of these issues, especially when generalizing to a specific population.

[25]We replicated our analysis in PLS with consistent results presented in Appendix D.

As research requirements continue to mount for faculty and Ph.D. students, the use of OCMs can provide a good foundation for extensive theory testing and refinement, especially if multiple iterative studies continue to strengthen and validate our research methodologies related to high quality data collection. The requirements for using OCMs relate to the researcher's ability to design the study, measurement scales, and survey development effectively, in terms of validity of the constructs, technical development of the questionnaire, and careful design of the OCM recruitment. From the potential differences across samples, we identify in Table 10 several procedural and sample attributes that should be clearly reported by researchers that use OCMs to support the justification, validation, replication, and consistent comparison of their reported results.

Acknowledgments

We would like to thank the senior editor, Rajiv Kohli, the associate editor, Eric Walden, and two anonymous reviewers for their insightful and supportive guidance during the review process. We also would like to extend our gratitude to Fred Davis, Rajiv Sabherwal, Traci Carte, and the University of Arkansas colloquium for their feedback on earlier versions of this manuscript.

References

Aust, F., Diedenhofen, B., Ullrich, S., and Musch, J. 2012. "Seriousness Checks Are Useful to Improve Data Validity in Online Research," *Behavior Research Methods* (45), pp. 527-535.

Bagchi, R., and Li, X. 2011. "Illusionary Progress in Loyalty Programs: Magnitudes, Reward Distances, and Step-Size Ambiguity," *Journal of Consumer Research* (37), pp. 888-901.

Benfield, J., and Szlemko, W. 2006. "Internet-Based Data Collection: Promises and Realities," *Journal of Research Practice* (2:2), pp. 1-15.

Bethlehem, J., and Biffignandi, S. 2012. *Handbook of Web Surveys*, Hoboken, NJ: Wiley.

Birnbaum, M. (ed.). 2000. *Psychological Experiments on the Internet*, San Diego, CA: Academic Press.

Birnbaum, M. H. 2004. "Human Research and Data Collection via the Internet," *Annual Review of Psychology* (55:1), pp. 803-832.

Bohannon, J. 2011. "Social Science for Pennies," *Science* (334:21), p. 307.

Buchanan, T., and Smith, J. L. 1999. "Research on the Internet: Validation of a World-Wide Web Mediated Personality Scale," *Behavior Research Methods* (31:4), pp. 565-571.

Buhrmester, M., Kwang, T., and Gosling, S. D. 2011. "Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-

Quality, Data?," *Perspectives on Psychological Science* (6:1), pp. 3-5.

Bursztein, E., Bethard, S., Fabry, C., Mitchell, J. C., and Jurafsky, D. 2010. "How Good Are Humans at Solving CAPTCHAs? A Large Scale Evaluation," in *Proceedings of the 31st IEEE Symposium on Security and Privacy*, Oakland, CA, May 16-19, pp. 399-413.

Burton-Jones, A. 2009. "Minimizing Method Bias Through Programmatic Research," *MIS Quarterly* (33:3), pp. 445-471.

Byrne, B. M. 2010. *Structural Equation Modeling with AMOS*, New York: Routledge.

Cheung, G. W., and Rensvold, R. B. 2002. "Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance," *Structural Equation Modeling* (9:2), pp. 233-255.

Chin, W. W. 2000. "Frequently Asked Questions—Partial Least Squares & PLS-Graph. Home Page" (http://disc-nt.cba.uh.edu/chin/plsfaq.htm).

Conway, C. M., and Limayem, M. 2011. "'You Want It When?' How Temporal Dissonance in IT Workers Contributes to Project Failures," in *Proceedings of the 32nd International Conference on Information Systems*, Shanghai, China, December 4-7.

Couper, M. 2000. "Review: Web Surveys: A Review of Issues and Approaches," *Public Opinion Quarterly* (64:4), pp. 464-494.

Curtin, R., Presser, S., and Singer, E. 2000. "The Effects of Response Rate Changes on the Index of Consumer Sentiment.," *Public Opinion Quarterly* (64:4), pp. 413-428.

Davis, F. 1989. "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology," *MIS Quarterly* (13:3), pp. 319-340.

Davis, F., Bagozzi, R., and Warshaw, P. R. 1989. "User Acceptance of Computer Technology: A Comparison of Two Theoretical Models," *Management Science* (35:8), pp. 982-1003.

Dean, D. L., Lowry, P. B., and Humpherys, S. 2011. "Profiling the Research Productivity of Tenured Information Systems Faculty at U.S. Institutions," *MIS Quarterly* (35:1), pp. 1-15.

DeVellis, D. R. F. 2003. *Scale Development: Theory and Applications* (2nd ed.), Newbury Park, CA: Sage.

Fornell, C., and Larcker, D. F. 1981. "Evaluating Structural Equation Models with Unobservable Variables and Measurement Error," *Journal of Marketing Research* (18:1), pp. 39-50.

Gordon, M., Slade, L. A., and Schmitt, N. 1986. "The 'Science of the Sophomore' Revisited: From Conjecture to Empiricism," *The Academy of Management Review* (11:1), pp. 191-207.

Gosling, S. D., Vazire, S., Srivastava, S., and John, O. P. 2004. "Should We Trust Web-Based Studies? A Comparative Analysis of Six Preconceptions about Internet Questionnaires," *American Psychologist* (59:2), pp. 93-104

Groves, R. 1989. *Survey Errors and Survey Costs*, New York: Wiley.

Groves, R. 2006. "Nonresponse Rates and Nonresponse Bias in Household Surveys," *Public Opinion Quarterly* (70:5), pp. 646-675.

Groves, R., Singer, E., and Corning, A. 2000. "Leverage-Saliency Theory of Survey Participation: Description and an Illustration," *Public Opinion Quarterly* (64:3), pp. 299-308.

Hair, J. F., Black, B., Babin, B., Anderson, R. E., and Tatham, R. L. 2006. *Multivariate Data Analysis* (6th ed.), Upper Saddle River, NJ: Pearson Prentice Hall.

Henrich, J., Heine, S. J., and Norenzayen, A. 2010. "The Weirdest People in the World?," *Behavioral and Brain Sciences* (33:2-3), pp. 61-83

Honaker, L. M. 1988. "The Equivalency of Computerized and Conventional MMPI Administration: A Critical Review," *Clinical Psychology Review* (8:6), pp. 561-577.

Howe, J. 2006. "The Rise of Crowdsourcing," *Wired Magazine* (14:14), pp. 1-5.

Hu, L., and Bentler, P. 1999. "Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives," *Structural Equation Modeling* (6:1), pp. 1-55.

Kaufmann, N., Schulze, T., and Veit, D. 2011. "More Than Fun and Money. Worker Motivation in Crowdsourcing: A Study on Mechanical Turk," in *Proceedings of the 17th Americas Conference on Information Systems*, Detroit, MI, August 4-7.

Keeter, S., Miller, C., Kohut, A., Groves, R. M., and Presser, S. 2000. "Consequences of Reducing Nonresponse in a National Telephone Survey," *Public Opinion Quarterly* (64:2), pp. 125-148.

Kim, J. O., and Mueller, C. W. 1978. *Introduction to Factor Analysis: What It Is and How to Do It*, Beverly Hills, CA: Sage.

King, W., and He, J. 2006. "A Meta-Analysis of the Technology Acceptance Model," *Information & Management* (43:6), pp. 740-755.

Kittur, A., Chi, E. H., and Suh, B. 2008. "Crowdsourcing User Studies with Mechanical Turk," in *Proceedings of 26th Conference on Human Factors in Computing Systems*, Florence, Italy, April 5-10, pp. 453-456.

Kline, R. B. 2010. *Principles and Practice of Structural Equation Modeling*, New York: The Guilford Press.

Leidner, D., and Kayworth, T. 2006. "*Review*: A Review of Culture in Information Systems Research: Toward a Theory of Information Technology Culture Conflict," *MIS Quarterly* (30:2), pp. 357-399.

Leonhardt, J. M., Keller, L. R., and Pechmann, C. 2011. "Avoiding the Risk of Responsibility by Seeking Uncertainty: Responsibility Aversion and Preference for Indirect Agency When Choosing for Others," *Journal of Consumer Psychology* (21:4), pp. 405-413.

Lewis, I., Watson, B., and White, K. M. 2009. "Internet Versus Paper-and-Pencil Survey Methods in Psychological Experiments: Equivalence Testing of Participant Responses to Health-Related Messages," *Australian Journal of Psychology* (61:2), pp. 107-116.

Malhotra, N. 2008. "Completion Time and Response Order Effects in Web Surveys," *Public Opinion Quarterly* (72:5), pp. 914-934.

Marge, M., Banerjee, S., and Rudnicky, A. I. 2010. "Using the Amazon Mechanical Turk for Transcription of Spoken Language," in *2010 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, Dallas, TX< March 14-19, pp. 5270-5273.

Marks, P. 2011. "Can Site's Low Pay Skew Research?," *New Scientist* (211:2820), p. 21.

Mason, W., and Suri, S. 2012. "Conducting Behavioral Research on Amazon's Mechanical Turk," *Behavior Research Methods* (44:1), pp. 1-23.

Meyerson, P., and Tryon, W. W. 2003. "Validating Internet Research: A Test of the Psychometric Equivalence of Internet and In-Person Samples," *Behavior Research Methods, Instruments, & Computers* (35:4), pp. 614-620.

Miller, J., Daly, J., Wood, M., Brooks, A., and Roper, M. 1996. "Electronic Bulletin Board Distributed Questionnaires for Exploratory Research," *Journal of Information Science* (22:2), pp. 107-115.

Mogilner, C., Aaker, J., and Kamvar, S. D. 2012. "How Happiness Affects Choice," *Journal of Consumer Research* (39:2), pp. 429-443.

Nunnally, J., and Bernstein, I. 1994. *Psychometric Theory* (3rd ed.), New York: McGraw-Hill.

Oliver, R. L. 1980. "A Cognitive Model for the Antecedents and Consequences of Satisfaction," *Journal of Marketing Research* (17:4), pp. 460-469.

Oppenheimer, D. M., Meyvis, T., and Davidenko, N. 2009. "Instructional Manipulation Checks: Detecting Satisficing to Increase Statistical Power," *Journal of Experimental Social Psychology* (45:4), pp. 867-872.

Paolacci, G., Burson, K., and Rick, S. 2011. "The Intermediate Alternative Effect: Considering a Small Tradeoff Increases Subsequent Willingness to Make Large Tradeoffs," *Journal of Consumer Psychology* (21:4), pp. 384-392.

Paolacci, G., Chandler, J., and Ipeirotis, P. G. 2010. "Running Experiments on Amazon Mechanical Turk," *Judgment and Decision Making* (5:5), pp. 411-419.

Pentz, M. A., and Chou, C. 1994. "Measurement Invariance in Longitudinal Clinical Research Assuming Change from Development and Intervention," *Clinical Psychology* (62:3), pp. 450-462.

Peterson, R. A. 2002. "On the Use of College Students in Social Science Research: Insights from a Second-Order Meta-Analysis," *Journal of Personality* (28), pp. 450-461.

R Core Team. 2012. "R: A Language and Environment for Statistical Computing," Vienna, Austria: R Foundation for Statistical Computing.

Rookey, B. D., Hanway, S., and Dillman, D. A. 2008. "Does a Probability-Based Household Panel Benefit from Assignment to Postal Response as an Alternative to Internet-Only?," *Public Opinion Quarterly* (72:5), pp. 962-984.

Ross, J., Irani, L., Zaldivar, A., and Tomlinson, B. 2010. "Who Are the Crowdworkers? Shifting Demographics in Mechanical Turk," in *CHI EA '10: Extended Abstracts on Human Factors in Computing Systems*, Atlanta, GA, April 10-15, pp. 2863-2872.

Sakamoto, Y., and Bao, J. 2011. "Testing Tournament Selection in Creative Problem Solving Using Crowds," in *Proceedings of the 32nd International Conference on Information Systems*, Shanghai, China, December 4-7.

Schmidt, W. 1997. "World-Wide Web Survey Research: Benefits, Potential Problems, and Solutions," *Behavior Research Methods* (29:2), pp. 274-279.

Shen, W., Kiger, T. B., Davies, S. E., Rasch, R. L., Simon, K. M., and Ones, D. S. 2011. "Samples in Applied Psychology: Over a Decade of Research in Review," *The Journal of Applied Psychology* (96:5), pp. 1055-1064.

Smith, M. A., and Leigh, B. 1997. "Virtual Subjects: Using the Internet as an Alternative Source of Subjects and Research Environment," *Behavior Research Methods* (29:4), pp. 496-505.

Sorokin, A., Berenson, D., Srinivasa, S. S., and Hebert, M. 2010. "People Helping Robots Helping People: Crowdsourcing for Grasping Novel Objects," in *Proceedings of the 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Taipei, Taiwan, October 18-22, pp. 2117-2122.

Spiller, S. A. 2011. "Opportunity Cost Consideration," *Journal of Consumer Research* (38:4), pp. 595-610.

Sprouse, J. 2011. "A Validation of Amazon Mechanical Turk for the Collection of Acceptability Judgments in Linguistic Theory," *Behavior Research Methods* (43), pp. 155-167.

Steenkamp, J. B. E., and Baumgartner, H. 1998. "Assessing Measurement Invariance in Cross-National Consumer Research," *Journal of Consumer Research* (25:1), pp. 78-107.

Straub, D., Loch, K., Evaristo, R., Karahanna, E., and Strite, M. 2002. "Toward a Theory-based Measurement of Culture," *Journal of Global Information Management* (10:1), pp. 13-23.

Sun, Y., Wang, N., and Peng, Z. 2011. "Working for One Penny: Understanding Why People Would Like to Participate in Online Tasks with Low Payment," *Computers in Human Behavior* (27:2), pp. 1033-1041.

Thomas, R. W. 2011. "When Student Samples Make Sense in Logistics Research," *Supply Chain Management* (32:3), pp. 287-290.

Toepoel, V., Das, M., and Van Soest, A. 2008. "Effects of Design in Web Surveys: Comparing Trained and Fresh Respondents," *Public Opinion Quarterly* (72:5), pp. 985-1007.

Trochim, W. M. K., and Donnelly, J. P. 2006. *The Research Methods Knowledge Base*, Cinncinati, OH: Atomic Dog.

Vadillo, M. A., and Matute, H. 2011. "Further Evidence on the Validity of Web-Based Research on Associative Learning: Augmentation in a Predictive Learning Task," *Computers in Human Behavior* (27:2), pp. 750-754.

Venkatesh, V., Morris, M. G., Davis, G. B., and Davis, F. D. 2003. "User Acceptance of Information Technology: Toward a Unified View," *MIS Quarterly* (27:3), pp. 425-478.

Ward, M. K., and Broniarczyk, S. M. 2011. "It's Not Me, It's You: How Gift Giving Creates Giver Identity Threat as a Function of Social Closeness," *The Journal of Consumer Research* (38:1), pp. 164-181.

Yu, L., and Nickerson, J. V. 2011. "Generating Creative Ideas Through Crowds: An Experimental Study of Combination," in *Proceedings of the 32nd International Conference on Information Systems*, Shanghai, China, December 4-7.

About the Authors

Zachary R. Steelman is an assistant professor in the Management Science and Information Systems Department at the Spears School of Business, Oklahoma State University. His research interests include knowledge management systems and the selection, development, and management of IT portfolios for individuals, teams, and organizations. He has served as an editorial assistant for *IEEE Transactions on Engineering Management*. He has authored or coauthored articles appearing in proceedings of the Americas Conference on Information Systems and of the Hawaii International Conference on System Sciences.

Bryan I. Hammer is a faculty member at the Spears School of Business, Oklahoma State University. His research interests include privacy and security, social media usage, and the interrelationship between information systems and neuroscience. Outlets for research he has authored and coauthored include *MIS* Quarterly, *Journal of the Association for Information Systems*, proceedings of the International Conference on Information Systems, and proceedings of the NeuroIS Gmunden Retreat.

Moez Limayem has served as dean of the College of Business at the University of South Florida since moving there in 2012 from the Sam M. Walton College of Business at the University of Arkansas, where he served simultaneously as the associate dean for research and graduate programs, the executive director of the Information Technology Research Institute and the Radio Frequency Identification Center, and the Edwin & Karlee Bradberry Chair in Information Systems. The recipient of numerous professional awards, Moez has published numerous articles, many of which focus on the intersection of technology with the consumer, academic, and business worlds. His research appears in journals such as *MIS Quarterly, Information Systems Research*, and *Management Science*.

MIS Quarterly

RESEARCH ESSAY

DATA COLLECTION IN THE DIGITAL AGE: INNOVATIVE ALTERNATIVES TO STUDENT SAMPLES

Zachary R. Steelman
Spears School of Business, Oklahoma State University, Tulsa, OK 74106 U.S.A. {zach@steelman.com}

Bryan I. Hammer
Spears School of Business, Management Science and Information Systems, Oklahoma State University, Stillwater, OK 74078 U.S.A. {bryan.hammer@okstate.edu}

Moez Limayem
College of Business, University of South Florida, Tampa, FL 33620 U.S.A. {mlimayem@usf.edu}

Appendix A

Amazon's Mechanical Turk Research Examples

Information Systems Examples

Andrea, C., and Lorenzo, N. 2010. "'We Don't Need No Education': Required Abilities in Online Labor Markets," in *Proceedings of the 31st International Conference on Information Systems*, St. Louis, MO, December 12-15.
Avital, M., Grover, V., Bjørn-Andersen, N., Li, E. Y., and Aviv, T. 2011. "Nurturing a Thriving Information Systems Discipline: A Call To Action," in *Proceedings of the 32nd International Conference on Information Systems*, Shanghai, China, December 4-7.
Chen, L., and Liu, D. 2012. "Comparing Strategies for Winning Expert-rated and Crowd-rated Crowdsourcing Contests: First Findings," in *Proceedings of the 18th Americas Conference on Information Systems*, Seattle, WA, August 9-11.
Conway, C. M., and Limayem, M. 2011. "'You Want It When?' How Temporal Dissonance in IT Workers Contributes to Project Failures," in *Proceedings of the 32nd International Conference on Information Systems*, Shanghai, China, December 4-7.
Ghose, A., Ipeirotis, P. G., and Li, B. 2010. "Designing Ranking Systems for Hotels on Travel Search Engines to Enhance User Experience," in *Proceedings of the 31st International Conference on Information Systems*, St. Louis, MO, December 12-15.
Grigore, M., and Rosenkranze, C. 2011. "Increasing the Willingness to Collaborate Online: An Analysis of Sentiment-Driven Interactions in Peer Content Production," in *Proceedings of the 32nd International Conference on Information Systems*, Shanghai, China, December 4-7.
Kern, R., Bauer, C., Thies, H., and Satzger, G. 2010. "Validating Results of Human-based Electronic Services Leveraging Multiple Reviewers," in *Proceedings of the 16th Americas Conference on Information Systems*, Lima, Peru, August 12-15.
Kim, J. 2012. "The Effect of Design Characteristics of Mobile Applications on User Retention: An Environmental Psychology Perspective," in *Proceedings of the 18th Americas Conference on Information Systems*, Seattle, WA, August 9-11.
Kumi, R., and Limayem, M. 2012. "Impact of Online Content on Attitudes and Buying Intentions," in *Proceedings of the 18th Americas Conference on Information Systems*, Seattle, WA, August 9-11.
Müller, R. M., Thoring, K., and Oostinga, R. 2010. "Crowdsourcing with Semantic Differentials: A Game to Investigate the Meaning of Form," in *Proceedings of the 16th Americas Conference on Information Systems*, Lima, Peru, August 12-15.
Sakamoto, Y., and Bao, J. 2011. "Testing Tournament Selection in Creative Problem Solving Using Crowds," in *Proceedings of the 32nd International Conference on Information Systems*, Shanghai, China, December 4-7.
Soliman, W., and Tuunainen, V. 2012. "Crowdsourcing as a Mobile Service*Case Study: Publishing Photography," in *Proceedings of the 18th Americas Conference on Information Systems*, Seattle, WA, August 9-11.

Steelman, Z., Soror, A., Limayem, M., and Worrell, D. 2012. "Obsessive Compulsive Tendencies as Predictors of Dangerous Mobile Phone Usage," in *Proceedings of the 18th Americas Conference on Information Systems*, Seattle, WA, August 9-11.

Sun, Y., Fang, Y., and Lim, K. H. 2010. "Understanding Sustained Participation in Transactional Virtual Communities," in *Proceedings of the 31st International Conference on Information Systems*, St. Louis, MO, December 12-15.

Sun, Y., Wang, N., Yin, C., and Che, T. 2012. "Investigating the Non-Linear Relationships in the Expectancy Theory: The Case of Crowd-sourcing Marketplace," in *Proceedings of the 18th Americas Conference on Information Systems*, Seattle, WA, August 9-11.

Van der Heijden, H. 2012. "A New Look at User Commitment Towards Information Systems: Evidence From Amazon's Mechanical Turk," in *Proceedings of the 20th European Conference on Information Systems*, Barcelona, Spain, June 10-13.

Wati, Y., and Davis, F. 2012. "The Relations Among Executive Functions and Users' Perceptions Toward Using Technologies to Multitask," in *Proceedings of the 11th Annual Workshop on SIGHCI Research in MIS*, Orlando, FL, December 16.

Woodall, P., Oberhofer, M., and Borek, A. 2012. "A Preliminary Study on Methods for Retaining Data Quality Problems in Automatically Generated Test Data," in *Proceedings of the 18th Americas Conference on Information Systems*, Seattle, WA, August 9-11.

Yu, L., and Nickerson, J. V. 2011. "Generating Creative Ideas Through Crowds: An Experimental Study of Combination," in *Proceedings of the 32nd International Conference on Information Systems*, Shanghai, China, December 4-7.

Zhang, N. A., Wang, C. A., and Xu, Y. 2011. "Privacy in Online Social Networks," in *Proceedings of the 32nd International Conference on Information Systems*, Shanghai, China, December 4-7.

Marketing Examples

Sela, A., and Berger, J. 2012. "Decision Quicksand: How Trivial Choices Suck Us In," *Journal of Consumer Research* (39:2), pp. 360-370.

Galak, J., Kruger, J., and Loewenstein, G. 2013. "Slow Down! Insensitivity to Rate of Consumption Leads to Avoidable Satiation," *Journal of Consumer Research* (39:5), pp. 993-1009.

Gershoff, A. D., and Koehler, J. J. 2011. "Safety First? The Role of Emotion in Safety Product Betrayal Aversion," *Journal of Consumer Research* (38:1), pp. 140-150.

Goodman, J. K., and Malkoc, S. A. 2012. "Choosing Here and Now versus There and Later: The Moderating Role of Psychological Distance on Assortment Size Preferences," *Journal of Consumer Research* (39:4), pp. 751-768.

Leonhardt, J. M., Keller, L. R., and Pechmann, C. 2011. "Avoiding the Risk of Responsibility by Seeking Uncertainty: Responsibility Aversion and Preference for Indirect Agency When Choosing for Others," *Journal of Consumer Psychology* (21:4), pp. 405-413.

Mogilner, C., Aaker, J., and Kamvar, S. D. 2012. "How Happiness Affects Choice," *Journal of Consumer Research* (39:2), pp. 429-443.

Paolacci, G., Burson, K., and Rick, S. 2011. "The Intermediate Alternative Effect: Considering a Small Tradeoff Increases Subsequent Willingness to Make Large Tradeoffs," *Journal of Consumer Psychology* (21:4), pp. 384-392.

Rick, S. I., and Schweitzer, M. E. 2013. "The Imbibing Idiot Bias: Consuming Alcohol Can be Hazardous to Your (Perceived) Intelligence," *Journal of Consumer Psychology* (23:2), pp. 212-219.

Spiller, S. A. 2011. "Opportunity Cost Consideration," *Journal of Consumer Research* (38:4), pp. 595-610.

Thomas, M., and Tsai, C. I. 2012. "Psychological Distance and Subjective Experience: How Distancing Reduces the Feeling of Difficulty," *Journal of Consumer Research* (39:2), pp. 324-340.

Townsend, C., and Sood, S. 2012. "Self-Affirmation Through the Choice of Highly Aesthetic Products," *Journal of Consumer Research* (39:2), pp. 415-428.

Ward, M. K., and Broniarczyk, S. M. 2011. "It's Not Me, It's You: How Gift Giving Creates Giver Identity Threat as a Function of Social Closeness," *The Journal of Consumer Research* (38:1), pp. 164-181.

Weaver, K., Garcia, S. M., and Schwarz, N. 2012. "The Presenter's Paradox," *Journal of Consumer Research* (39:3), pp. 445-460.

Xu, A. J., and Wyer, R. S. 2012. "The Role of Bolstering and Counterarguing Mind-Sets in Persuasion," *Journal of Consumer Research* (38:5), pp. 920-932.

Appendix B

TAM Measurement Scales

Perceived Usefulness: Seven-point Likert scales, “strongly disagree” to “strongly agree.”

- Using Windows 7 would enable me to accomplish tasks more quickly.
- Using Windows 7 would improve my performance.
- Using Windows 7 would increase my productivity.
- Using Windows 7 would enhance my effectiveness.
- Using Windows 7 would make it easier to do my job.
- I would find Windows 7 useful in my job.

Perceived Ease of Use: Seven-point Likert scales, “strongly disagree” to “strongly agree.”

- Learning to operate Windows 7 would be easy for me.
- I would find it easy to get Windows 7 to do what I want it to do.
- My interaction with Windows 7 would be clear and understandable.
- I would find Windows 7 to be flexible to interact with.
- It would be easy for me to become skillful at using Windows 7.
- I would find Windows 7 easy to use.

Behavioral Intention: Seven-point Likert scales, “strongly disagree” to “strongly agree.”

- I intend to use the system.
- I predict I would use the system.
- I plan to use the system.

Demographic Variables:
Gender (1 = male, 2 = female)

Age (number specified by participant)

Please select the highest level of education you received:

- Less than High School
- High School / GED
- Some College
- 2-year College Degree
- 4-year College Degree
- Master’s Degree
- Doctoral Degree
- Professional Degree (JD, MD)

What is your race?

- White/Caucasian
- African American
- Hispanic
- Asian
- Native American
- Pacific Islander
- Mixed/Other

Please specify the country of your primary citizenship (listed in alphabetical order).

Did you serve in your country's military? If yes, please specify:
- Air
- Maritime/Naval/Sea
- Land/Army
- Other

Please indicate your family structure:
- Single without children
- Single with children
- Married without children
- Married with children
- Life partner without children
- Life partner with children

What is your annual income range?
- $19,999 and below
- $20,000 – $29,999
- $30,000 – $39,999
- $40,000 – $49,999
- $50,000 – $59,999
- $60,000 – $69,999
- $70,000 – $79,000
- $80,000 – $89,999
- $90,000 or more

Of the color choices presented, which one do you prefer the most?
- Blue
- Yellow
- Green
- Red
- Orange
- Violet
- Purple
- Black
- White

Human Verification Questions: The next question is to ensure you are not a robot application. Please select "[letter]" as your answer.
- A
- B
- C
- D
- E

What is your religious affiliation? (Note: Chinese Religion is defined as the combined beliefs of Chinese folk religion, Confucianism, Taoism, Buddhism, and ancestor worship).
- Atheist
- Aladura
- Asatru
- Baha'i Faith
- Bon
- Buddhism
- Cao Dai
- Chinese Religion
- Chopra Center
- Christianity
- Christian Science
- Confucianism
- Eckankar
- Epicureanism
- Falun Gong
- Greco-Roman Religion
- Hare Krishna
- Hinduism
- Islam
- Jainism
- Judaism
- Kemetic Reconstructionism
- Mayan Religion
- Mithraism
- Neopaganism
- New Thought
- The Occult
- Rastafari
- Satanism
- Scientology
- Shinto
- Sikhism
- Stoicism
- Taoism (Daoist)
- Unification Church
- Unitarian Universalism
- Vampirism
- Wicca
- Zoroastrianism

Which occupational category best describes your employment? (based on the U.S. Census, 40 categories)

- Management: professional or related occupations
- Management: business or financial operations occupations
- Management occupations, except farmers and farm managers
- Farmers and farm managers
- Business and financial operations
- Business operations specialists
- Financial specialists
- Computer or mathematical
- Architects, surveyors, cartographers, or engineers
- Drafters, engineering, or mapping technicians
- Life, physical, or social science
- Community and social services
- Legal
- Education, training, or library
- Arts, design, entertainment, sports, or media
- Health diagnosing or treating practitioners & technical occupations
- Health technologists or technicians
- Health care support
- Fire fighting, prevention or law enforcement workers (including supervisors)
- Other protective service workers (including supervisors)
- Food preparation or serving-related
- Building, grounds cleaning or maintenance
- Personal care or service
- Sales or related occupations
- Office or administrative support
- Farming, fishing, or forestry
- Supervisors, construction or extraction
- Construction trades workers
- Extraction workers
- Installation, maintenance, or repair occupations
- Production
- Supervisors, transportation or material moving
- Aircraft or traffic control
- Motor vehicle operators
- Rail, water or other transportation
- Material moving

Appendix C

Study 2: Expectation–Confirmation Theory Analyses

To provide robustness to our primary analyses and explore the validity of utilizing online crowdsourcing markets (OCM) for academic research we conducted a second study utilizing a larger, alternative model from Marketing, an allied discipline. We selected the expectation–disconfirmation theory (EDT) (Oliver 1980), which examines the antecedents and consequences of satisfaction decisions. We chose the EDT because we needed a secondary model that has been utilized consistently both in its measurement as well as its relationships in prior research (Buchanan and Smith 1999; Meyerson and Tryon 2003) and that is a slightly more complex model compared to our TAM model in Study 1. The original EDT model as described by Oliver (1980) examines the attitudes and intentions across multiple time periods to investigate the impacts of expectations and the disconfirmation of expectations. Again, as the focus of this article is not to propose new theory, we refer readers to Oliver's original work for a discussion of the rationale for this model, which has seen repeated support.

In an attempt to strengthen the results within Study 1, we conduct a replication of our primary analyses utilizing the EDT model to determine any differences that may arise between OCMs and alternative samples in a larger, more complex structural model.

Participants and Demographics

For this study, we recruited two sets of participants: (1) college students from a major midwestern U.S. university and (2) users of a popular OCM, Amazon's Mechanical Turk. To directly mirror the design utilized in Study 1 and ensure consistency, we collected responses from four specific samples: college student, U.S., non-U.S., and worldwide OCM respondent. In all four samples, the participant data was collected using an identical survey questionnaire.

Procedure

The empirical test of the EDT model was conducted utilizing data collected via an online survey. After giving their consent, participants were given a series of informational pages depicting information about the recently released version of Microsoft Office Web Apps, which can be utilized as a potential replacement for the traditional Microsoft Office Desktop Suite. After familiarizing themselves with the technology description, participants responded to a series of questions at time 1 (expectations, attitude, and behavioral intention). Following the submission of their time 1 responses the participants were asked to utilize a trial version of the Microsoft Office Web Apps software, which could be accessed via the Internet for free. After interacting with the software for a short period of time (average time was approximately 10 minutes), participants were asked a series of follow-up questions in reference to the technology (disconfirmation, satisfaction, attitude, behavioral intention). The final aspect of the survey consisted of a set of questions used in Study 1 to determine respondent demographics. All measurement items and scales utilized in Study 2 were adapted from prior literature and are available from the authors by request.

The participant restrictions put in place for the OCM samples were similar to those of Study 1: (1) users within the U.S. only, (2) users ***not*** within the U.S., and (3) an unrestricted, worldwide participation.[1] Due to the need for actual participation with the software and consequently the increased time of participation, all OCM participants were paid $0.50 for their complete and valid participation. The college students were recruited via a campus wide survey listing and were entered into a drawing for one of ten $10.00 gift cards for complete responses.

In Table C1, we display the distribution of responses received, removed, and usable for each sample. As in the TAM study, we are unable to calculate the actual response rate for the OCM samples and, therefore, the examination of the completed responses act as a proxy for response rate quality (Bethlehem and Biffignandi 2012). The data cleaning process mirrored that within Study 1, which we consider a *minimal* cleaning (refer to Study 1 for details). One interesting finding in this study is that the OCM respondents all completed the entire survey while the student sample exhibited some participants that dropped out of the survey. This could potentially be a function of the payment incentive increase compared to the incentive in Study 1. While the OCM participants all completed the survey, they failed more of the quality response checks than the student participants.

Table C1. Participant Response Removal

	Students	Worldwide OCM	U.S. OCM	Non-U.S. OCM
Total Responses	244	288	251	275
Failure to Finish	38	0	0	0
Failed Quality Questions	5	26	12	32
Final Usable Responses	201	262	239	243
Percent Usable	82%	91%	95%	88%

[1]The worldwide listing was collected first and any respondents who had participated in this survey were removed from U.S. or non-U.S. participation.

Analysis and Results

To conduct the primary analysis for Study 2, we used partial least squares (PLS) as it allows for easier handling of the second-order formative construct for the modeling of expectations (Chin 1998; Ringle et al. 2012). To remain consistent with Oliver's conceptualization of individual expectations as a summated value of a set of expectations in reference to a product or technology we modeled the two focal expectations (perceived ease of use and perceived usefulness; Davis, 1989) as a second-order formative construct. The discussion below depicts the results from the PLS analysis. We additionally ran the analysis utilizing a second-order reflective model with highly similar results.

Demographics

We began our analysis in Study 2 by examining the differences in the composition of each sample by empirically comparing the distributions across each sample. In Table C2, we provide the distribution of demographic attributes across all four samples. The demographic distributions from Study 2 are highly similar to those collected during Study 1, which could indicate the type of demographic distributions that may be expected utilizing each respective sampling frame. To further examine the differences among samples we empirically compared each sample; we provide the detailed results of our demographic comparisons between each utilizing a series of t-tests of means, chi-square tests of proportions, and Wilcoxon sum-rank tests for categorical rank differences in Table C3. As expected, the samples differed across various demographic attributes such as age, education, family structure, and income. Therefore, the selection and use of a sampling frame in a researcher's study should take into account the type of demographic distributions that may exist based on the technique. The collection of these attributes will allow for the *post hoc* adjustment of weighting responses if required as well as controlling for potential demographic differences that may exist within the theoretical relationships. However, while the demographics do differ between samples, a further look at the measurement and structural properties of the theoretical models is required to determine the extent of various biases.

Confirmatory Factor Analysis

To examine the differences in the measurement item structures with regard to convergent and divergent validity as well as reliabilities, we examine a series of psychometric tests. To establish convergent validity of each construct within our PLS analysis, we examine the factor loadings and cross-loadings as well as the average variance explained (AVE) for each construct (Hair et al. 2006). Across all samples, each item loaded primarily on its focal construct and less on the other constructs in the model,[2] providing evidence of convergent validity (Chin 1998; Gefen and Straub 2005). Additionally, the AVE for each construct (see Table C4) exceeded the recommended 0.50 threshold ranging from 0.62 to 0.89 providing further support for convergent validity (Hair et al. 2006).

To establish discriminant validity, we again examined the factor loadings and cross-loadings as well as the square root of the AVE of each construct in relation to all other constructs in the model (Fornell and Larcker 1981). The results indicate that, across all samples, the AVEs exceed all correlations among the variables and the measurement items load primarily on their focal construct and less so on all others (Chin 1998; Gefen and Straub 2005).

Finally, to determine the reliability and consistency of the scales utilized with the model, we examine both the Cronbach's alpha and composite rho of each scale, which should meet or exceed 0.70 for adequate reliability (Hair et al. 2006). The reliability estimates all exceed these thresholds within all samples aside from disconfirmation in the worldwide OCM with a value of 0.68. However, the associated composite rho for this construct is 0.82, which provides some evidence of a reliable measure. Therefore, we have evidence that the psychometrics of the scales utilized in the EDT models are valid and consistent across all samples by meeting or exceeding our series of validation tests. Additionally, we have found no significant differences or threats to validity between the samples to indicate that the measures were interpreted differently.

Structural Model

In Figure C1, we present the structural path models estimated utilizing partial least squares (PLS) with a recommended bootstrapping estimation of 1,000 resamples (Chin 2010). As discussed earlier, ***expectations*** was modeled as a second-order formative construct in line with Oliver's conceptualization of expectations to be a summation of all associated expectations. Interestingly, when examining the models as a whole, we find relatively consistent results within all samples, which indicate a level of confidence that the samples do not differ on many aspects, similar to the results from Study 1. Additionally, all the theorized paths within the model were significant across all samples, even in the presence of a more complex model.

---

[2]Due to space constraints, the factor loading matrices are not printed here but are available from the authors upon request.

Table C2. Demographics Distribution					
		Distribution			
		Student	Worldwide OCM	U.S. OCM	Non-U.S. OCM
Gender*	Male	0.55	0.66	0.51	0.70
	Female	0.45	0.35	0.49	0.30
Age	Mean	23.34	29.98	32.85	28.46
	Median	22.00	27.00	29.00	26.00
	Minimum	19.00	18.00	19.00	16.00
	Maximum	45.00	69.00	65.00	63.00
Education Level*	Less than High School	0.00	0.00	0.00	0.01
	High School/GED	0.06	0.05	0.13	0.05
	Some College	0.63	0.14	0.31	0.10
	2-Year College Degree	0.10	0.05	0.14	0.10
	4-Year College Degree	0.20	0.47	0.30	0.46
	Master's Degree	0.01	0.26	0.09	0.23
	Doctoral Degree	0.00	0.01	0.01	0.00
	Professional Degree (JD, MD)	0.00	0.02	0.01	0.03
Race*	White/Caucasian	0.67	0.10	0.70	0.07
	African American	0.03	0.02	0.10	0.00
	Hispanic	0.05	0.01	0.05	0.01
	Asian	0.13	0.76	0.05	0.84
	Native American	0.01	0.00	0.01	0.00
	Pacific Islander	0.00	0.00	0.00	0.00
	Mixed/Other	0.09	0.10	0.09	0.08
Family Structure*	Single, no children	0.88	0.48	0.45	0.51
	Single, with children	0.03	0.02	0.08	0.02
	Married, no children	0.02	0.11	0.09	0.11
	Married, with children	0.04	0.26	0.24	0.27
	Life partner, no children	0.01	0.03	0.07	0.01
	Life partner, with children	0.01	0.05	0.05	0.04
Annual Income Range (in U.S. dollars)*	$19,999 >	0.80	0.52	0.26	0.53
	$20,000 – $29,999	0.08	0.18	0.16	0.22
	$30,000 – $39,999	0.01	0.12	0.14	0.08
	$40,000 – $49,999	0.01	0.07	0.13	0.05
	$50,000 – $59,999	0.02	0.05	0.10	0.04
	$60,000 – $69,999	0.01	0.02	0.05	0.04
	$70,000 – $79,000	0.00	0.00	0.02	0.01
	$80,000 – $89,999	0.01	0.01	0.08	0.01
	$90,000 <	0.02	0.02	0.04	0.02
Time Elapsed	Mean	21.63	22.71	18.51	25.31
	Median	14.37	15.84	12.75	14.50
	Std. Dev.	40.90	21.29	18.26	84.84
	Min.	3.48	2.88	2.72	2.50
	Max.	488.82	144.00	155.83	1288.02

*Value displayed as percentage of total responses.

Table C3. Demographic Differences

		Demographic Comparisons					
		Student vs. Worldwide	**Student vs. U.S.**	**Student vs. Non-U.S.**	**U.S. vs. Worldwide**	**Non-U.S. vs. Worldwide**	**U.S. vs. Non-U.S.**
Gender†	Male	4.773*	0.478	10.247**	9.809**	1.082	17.317***
	Female						
Age††	Mean	10.415***	12.368***	9.076***	3.166**	2.053*	5.126***
Education Level††	Education Rank	54.024***	4.723***	12.744***	45.817***	57.948***	6.883***
Race†	White/Caucasian	159.166***	0.257	171.294***	184.411***	0.972	196.172***
	African American	0.236	6.206*	4.257*	12.000***	2.026	20.808***
	Hispanic	6.411*	0.000	5.719*	6.847**	0.000	6.114*
	Asian	187.869***	4.116*	221.331***	252.686***	3.214	288.807***
	Native American	0.053	0.000	0.717	0.000	0.000	0.519
	Pacific Islander	0.018	0.008	0.009	0.000	0.000	0.000
	Mixed/Other	0.000	0.006	0.196	0.080	0.453	0.048
Family Structure†	Single, no children	65.949***	82.630***	57.200***	1.463	0.219	2.984
	Single, with children	1.289	4.483*	0.977	0.921	0.019	1.529
	Married, no children	20.572***	11.292***	17.628***	1.664	0.071	0.779
	Married, with children	49.095***	36.617***	47.139***	1.054	0.000	0.801
	Life partner, no children	7.112**	10.722***	3.015	0.401	0.804	2.727
	Life partner, with children	11.593***	6.706**	7.279**	0.670	0.489	0.000
Annual Income Range	Mean Difference††	3.762***	9.968***	3.242**	6.954***	0.485	7.295***
	Categorical Rank†††	18660***	10277***	17812***	42186***	32616	40019***
Time Elapsed	Mean	0.341	0.998	0.598	2.371*	0.465	1.222

Notes: †Chi-square proportion; ††mean difference t-test, †††Wilcoxon sum-rank test, *p < .05; **p < .01; ***p < .001.

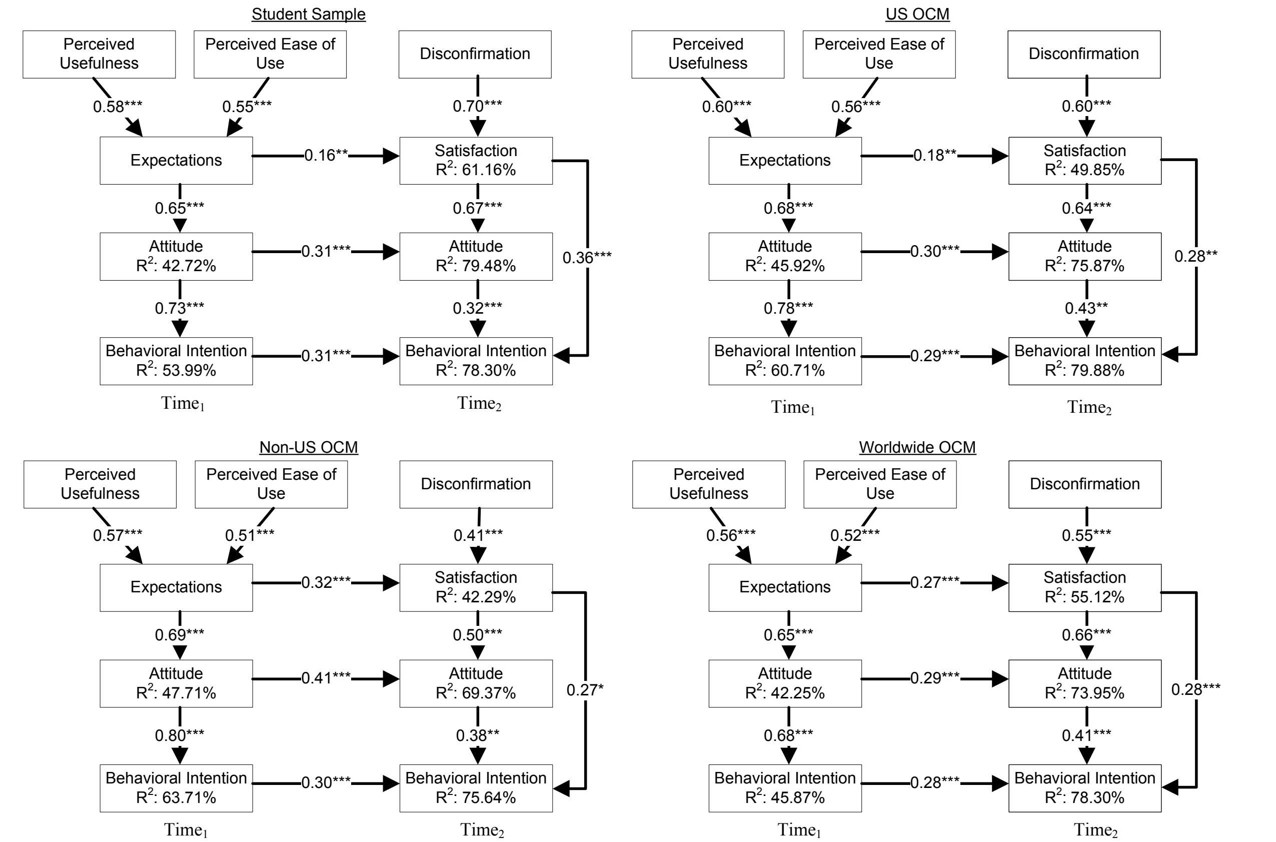
To dive further into the analysis and determine any differences across specific coefficients we conducted a series of t-tests of differences (Chin 2000). Table C5 provides the results of these comparisons which show that only 3 of 66 relationships differ. Interestingly, it is not the worldwide OCM sample that provides differences in this study but the non-U.S. OCM sample. The results indicate that the student, U.S., and worldwide OCM samples do not differ significantly across their theoretical relationships. Thus, based on the results so far, we have provided additional evidence in support of OCMs providing similar results to those of student samples. However, to further mirror our analyses in Study 1, we continue with a comparison of differences between mean scale levels as well seeking further clarity.

To provide a clearer picture of which measurement scales may differ across samples, we conducted a series of ANOVA and pair-wise comparison tests with a Scheffe's correction. The results presented in Table C6 indicate that the variables collected in time 1 (expectation, attitude, and behavioral intention) did not significantly differ among samples ($p > 0.05$). However, the samples did exhibit differences in the time 2 variables (disconfirmation, $p < 0.001$; satisfaction, $p < 0.001$; attitude, $p < 0.01$; and behavioral intention, $p < 0.05$). To determine which specific samples may have caused the differences present in our ANOVA analyses, we examined the pair-wise comparisons in further detail. It appears that within attitude (t2) the differences exist between the worldwide and student samples ($p < 0.05$) while the student, U.S. and non-U.S. samples do not significantly differ. Behavioral intention (t2) significantly differs only between the worldwide and U.S. OCM samples ($p < 0.05$) while all other comparisons do not differ. However, the differences within disconfirmation and satisfaction become more prominent with the student and U.S. OCM samples differing significantly ($p < 0.01$) from both the worldwide and non-U.S. OCM samples.

Table C4. Correlations and Reliabilities

		AVE	Composite Rho	Cronbach's Alpha	1	2	3	4	5	6	7
Worldwide OCM											
1	Attitude (t2)	0.74	0.92	0.88	**0.86**						
2	Attitude (t1)	0.76	0.93	0.89	0.66	**0.87**					
3	BI (t2)	0.74	0.92	0.89	0.84	0.63	**0.86**				
4	BI (t1)	0.77	0.93	0.90	0.74	0.68	0.77	**0.88**			
5	Disconfirmation	0.62	0.82	0.68	0.61	0.44	0.61	0.54	**0.78**		
6	Expectation	0.62	0.93	0.91	0.66	0.65	0.68	0.74	0.61	**0.79**	
7	Satisfaction	0.72	0.91	0.87	0.83	0.56	0.80	0.68	0.71	0.60	**0.85**
Non-U.S. OCM											
		AVE	**Composite Rho**	**Cronbach's Alpha**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
1	Attitude (t2)	0.76	0.93	0.89	**0.87**						
2	Attitude (t1)	0.77	0.93	0.90	0.74	**0.88**					
3	BI (t2)	0.82	0.95	0.93	0.82	0.70	**0.90**				
4	BI (t1)	0.80	0.94	0.92	0.79	0.80	0.79	**0.89**			
5	Disconfirmation	0.63	0.83	0.72	0.61	0.58	0.62	0.61	**0.79**		
6	Expectation	0.65	0.94	0.92	0.66	0.69	0.61	0.70	0.59	**0.81**	
7	Satisfaction	0.76	0.93	0.90	0.78	0.67	0.78	0.73	0.60	0.56	**0.87**
U.S. OCM											
		AVE	**Composite Rho**	**Cronbach's Alpha**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
1	Attitude (t2)	0.85	0.96	0.94	**0.92**						
2	Attitude (t1)	0.79	0.94	0.91	0.72	**0.89**					
3	BI (t2)	0.89	0.97	0.96	0.86	0.69	**0.94**				
4	BI (t1)	0.86	0.96	0.94	0.70	0.78	0.75	**0.92**			
5	Disconfirmation	0.72	0.88	0.80	0.64	0.49	0.61	0.42	**0.85**		
6	Expectation	0.62	0.93	0.91	0.54	0.68	0.56	0.65	0.47	**0.79**	
7	Satisfaction	0.83	0.95	0.93	0.84	0.65	0.80	0.59	0.69	0.46	**0.91**
Students											
		AVE	**Composite Rho**	**Cronbach's Alpha**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
1	Attitude (t2)	0.86	0.96	0.95	**0.93**						
2	Attitude (t1)	0.80	0.94	0.92	0.72	**0.89**					
3	BI (t2)	0.87	0.96	0.95	0.83	0.67	**0.93**				
4	BI (t1)	0.85	0.96	0.94	0.67	0.73	0.74	**0.92**			
5	Disconfirmation	0.70	0.87	0.78	0.70	0.47	0.65	0.45	**0.84**		
6	Expectation	0.67	0.94	0.93	0.52	0.65	0.53	0.59	0.41	**0.82**	
7	Satisfaction	0.83	0.95	0.93	0.86	0.63	0.82	0.60	0.77	0.45	**0.91**

Notes: BI = Behavioral Intention, t1 = time period 1, t2 = time period 2, AVE = Average Variance Explained, square-root of the AVE on diagonal.



Notes: **p* < 0.05, ***p* < 0.01, ****p* < 0.001

Figure C1. Structural Model Results

Table C5. Path Coefficient Differences

	Student vs. U.S.	Student vs. Non-U.S.	Student vs. Worldwide	U.S. vs. Non-U.S.	U.S. vs. Worldwide	Non-U.S. vs. Worldwide
Attitude (t1) -> Attitude (t2)	0.006(0.933)	-0.102(0.424)	0.018(0.836)	-0.108(0.374)	0.012(0.892)	0.119(0.336)
Attitude (t1) -> BI (T1)	-0.044(0.427)	-0.063(0.132)	0.058(0.419)	-0.019(0.705)	0.102(0.158)	0.121(0.060)
Attitude (t2) -> BI (t2)	-0.106(0.567)	-0.06(0.715)	-0.087(0.477)	0.046(0.812)	0.019(0.909)	-0.027(0.850)
BI (t1) -> BI (t2)	0.020(0.849)	0.010(0.921)	0.025(0.800)	-0.011(0.923)	0.004(0.969)	0.015(0.883)
PEOU -> Expectation	-0.013(0.651)	0.039(0.105)	0.023(0.326)	**0.052(0.029)**	0.036(0.122)	-0.016(0.431)
PU -> Expectation	-0.022(0.592)	0.007(0.838)	0.019(0.552)	0.029(0.410)	0.042(0.223)	0.012(0.644)
Expectation -> Attitude (t1)	-0.024(0.741)	-0.037(0.549)	0.004(0.964)	-0.013(0.844)	0.028(0.729)	0.041(0.573)
Expectation -> Satisfaction	-0.018(0.828)	-0.156(0.130)	-0.105(0.255)	-0.138(0.173)	-0.087(0.344)	0.051(0.626)
Disconfirmation -> Satisfaction	0.098(0.219)	**0.293(0.006)**	0.154(0.094)	**0.195(0.046)**	0.056(0.510)	-0.139(0.181)
Satisfaction -> Attitude (t2)	0.022(0.761)	0.164(0.205)	0.002(0.983)	0.142(0.251)	-0.020(0.805)	-0.162(0.192)
Satisfaction -> Attitude (t2)	0.082(0.577)	0.089(0.596)	0.081(0.507)	0.007(0.963)	-0.001(0.993)	-0.008(0.953)

Notes: p-value for two-tailed tests in parentheses, BI = Behavioral Intention, PEOU = Perceived Ease of Use, PU = Perceived Usefulness

Table C6. Scale Mean Differences

Attitude (t1) (p = 0.309)							
		Mean	**SD**	**1**	**2**	**3**	**4**
1	Student	5.96	0.91	-			
2	U.S. OCM	5.99	1.04	0.03	-		
3	Non-U.S. OCM	6.06	0.93	0.10	0.07	-	
4	Worldwide OCM	6.11	0.98	0.16	0.12	0.06	-

Attitude (t2) (p = 0.007)							
		Mean	**SD**	**1**	**2**	**3**	**4**
1	Student	5.85	1.04	-			
2	U.S. OCM	5.92	1.08	0.07	-		
3	Non-U.S. OCM	6.02	0.95	0.16	0.09	-	
4	Worldwide OCM	6.15	0.91	0.30*	0.23	0.14	-

Behavioral Intention (t1) (p = 0.534)							
		Mean	**SD**	**1**	**2**	**3**	**4**
1	Student	6.10	0.96	-			
2	U.S. OCM	5.99	1.13	-0.10	-		
3	Non-U.S. OCM	6.05	0.98	-0.05	0.06	-	
4	Worldwide OCM	6.12	0.96	0.02	0.13	0.07	-

Behavioral Intention (t2) (p = 0.017)							
		Mean	**SD**	**1**	**2**	**3**	**4**
1	Student	5.88	1.09	-			
2	U.S. OCM	5.80	1.27	-0.08	-		
3	Non-U.S. OCM	5.99	1.05	0.11	0.19	-	
4	Worldwide OCM	6.09	0.94	0.22	0.29*	0.11	-

Disconfirmation (p = 0.000)							
		Mean	**SD**	**1**	**2**	**3**	**4**
1	Student	5.07	0.84	-			
2	U.S. OCM	5.11	0.96	-0.08	-		
3	Non-U.S. OCM	5.42	0.88	0.11***	0.19**	-	
4	Worldwide OCM	5.53	0.92	0.22***	0.29***	0.11	-

Expectation (p = 0.133)							
		Mean	**SD**	**1**	**2**	**3**	**4**
1	Student	5.79	0.95	-			
2	U.S. OCM	5.65	0.95	-0.14	-		
3	Non-U.S. OCM	5.73	0.86	-0.07	0.08	-	
4	Worldwide OCM	5.83	0.85	0.03	0.18	0.10	-

Perceived Ease of Use (p = 0.463)							
		Mean	**SD**	**1**	**2**	**3**	**4**
1	Student	5.79	1.09	-			
2	U.S. OCM	5.71	1.08	-0.08	-		
3	Non-U.S. OCM	5.72	0.87	-0.07	0.01	-	
4	Worldwide OCM	5.83	0.89	0.04	0.12	0.11	-

Perceived Usefulness (p = 0.044)							
		Mean	**SD**	**1**	**2**	**3**	**4**
1	Student	5.79	1.05	-			
2	U.S. OCM	5.57	1.14	-0.22	-		
3	Non-U.S. OCM	5.73	1	-0.06	0.16	-	
4	Worldwide OCM	5.82	0.95	0.03	0.25*	0.09	-

Satisfaction (p = 0.000)							
		Mean	**SD**	**1**	**2**	**3**	**4**
1	Student	5.48	1.04	-			
2	U.S. OCM	5.71	1.03	0.23	-		
3	Non-U.S. OCM	5.86	0.96	0.38***	0.15	-	
4	Worldwide OCM	6.00	0.92	0.53***	0.29*	0.15	-

Note: t1 = time period 1, t2 = time period 2, p-value of ANOVA in parentheses.

Discussion

Based on our series of analyses, it is evident that the samples collected do have some slight differences across demographics, a few theoretical relationships, and their scale levels within the model. However, when taking a closer look at the individual samples themselves we find strong evidence that differences do not exist between the student and U.S. OCM samples or the majority of the relationships within the worldwide and non-U.S. OCM samples despite having slight differences. Thus, based on the results from Study 1 and Study 2, we have confidence that the use of OCMs is a potential alternative to student samples and provides results that are very similar to what one would expect to capture from a student sample (if one were intending to measure that unique perspective) while providing a quicker, more diverse, and cheaper alternative for participant recruitment. In addition, by indicating the types of demographic differences that may exist, relevant variables have been identified by these studies that can be utilized for controls to provide further support within a researcher's examination.

References

Bethlehem, J., and Biffignandi, S. 2012. *Handbook of Web Surveys*, New York: Wiley.

Buchanan, T., and Smith, J. L. 1999. "Research on the Internet: Validation of a World-Wide Web Mediated Personality Scale," *Behavior Research Methods* (31:4), pp. 565-571.

Chin, W. W. 1998. "The Partial Least Squares Approach to Structural Equation Modeling," in *Modern Methods for Business Research: Methodology for Business and Management*, G. A. Marcoulides (ed.), London: Lawrence Erlbaum Associates, pp. 295-336.
Chin, W. W. 2000. "Frequently Asked Questions—Partial Least Squares & PLS-Graph" (http://disc-nt.cba.uh.edu/chin/plsfaq.htm).
Chin, W. W. 2010. "How to Write Up and Report PLS Analyses," in *Handbook of Partial Least Squares*, V. Esposito Vinzi, W. W. Chin, J. Henseler, and H. Wang (eds.), Berlin: Springer, pp. 655-690.
Davis, F. 1989. "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology," *MIS Quarterly* (13:3), pp. 319-340.
Fornell, C., and Larcker, D. F. 1981. "Evaluating Structural Equation Models with Unobservable Variables and Measurement Error," *Journal of Marketing Research* (18:1), pp. 39-50.
Gefen, D., and Straub, D. 2005. "A Practical Guide to Factorial Validity Using PLS-Graph: Tutorial and Annotated Example," *Communications of the Association for Information Systems* (16:1), pp. 91-109.
Hair, J. F., Black, B., Babin, B., Anderson, R. E., and Tatham, R. L. 2006. *Multivariate Data Analysis* (6th ed.), Upper Saddle River, NJ: Pearson Prentice Hall.
Meyerson, P., and Tryon, W. W. 2003. "Validating Internet Research: A Test of the Psychometric Equivalence of Internet and In-Person Samples," *Behavior Research Methods, Instruments, & Computers* (35:4), pp. 614-620.
Oliver, R. L. 1980. "A Cognitive Model for the Antecedents and Consequences of Satisfaction," *Journal of Marketing Research* (17:4), pp. 460-469.
Ringle, C. M., Sarstedt, M., and Straub, D. W. 2012. "A Critical Look at the Use of PLS-SEM in MIS Quarterly," *MIS Quarterly* (36:1), pp. iii-xiv.

Appendix D

PLS Robustness Analysis

In addition to the primary covariance-based structural equation modeling (CB-SEM) analysis of the technology acceptance model (Davis 1989; Davis et al. 1989) presented within the paper, we replicated the entire analysis utilizing partial least squares (PLS) analysis. Some of our samples are slightly under the recommended minimum of 200 responses for robust estimates of CB-SEM. Therefore, to address these issues we have replicated the analysis utilizing PLS which is recommended for smaller sample sizes that do not meet the minimum sample sizes of CB-SEM while still incorporating a structural equation estimation technique (Chin 1998).

Table D1 presents the reliability estimates, average variance extracted (AVE), and correlation matrices for each of the samples collected. All of the constructs for each of the samples have reliability estimates >0.85, AVEs > 0.50, and square roots of the AVEs greater than any off-diagonal correlations provided evidence of reliability, convergent validity, and divergent validity (Hair et al. 2006). Additionally, Table D2 provides the loadings and cross-loadings for each measurement item utilized in each of the data samples. All of the data samples provide further evidence of convergent and divergent validity as each item loads primarily on its focal construct while having lesser loadings on all other constructs within the analysis (Gefen and Straub 2005). Therefore, the PLS analysis of each of the constructs provides adequate evidence of reliability, convergent, and divergent validity for each of the samples that remain consistent with the primary analysis.

After determining the validity and reliability of our measurement model, we estimated the structural model within SmartPLS 2.0 M3 (Ringle et al. 2005) utilizing the recommended bootstrapping estimation with 1,000 resamples to provide robust estimates and significance levels for each parameter (Chin 2010). Figure D1 depicts the results from each of our analyses for the structural models. The pattern of results provides estimates consistent with those found in the primary analysis utilizing CB-SEM techniques; however the variance maximization procedures utilized within PLS provide significant coefficients for even the worldwide OCM sample in this study. To test for significant differences between the path coefficients of each model we utilize a two-tailed t-test of differences (Chin 2000) presented in Table D3. We found that only 6 of the 30 comparisons were significantly different at the $p < 0.05$ level. Interestingly, in the CB-SEM analysis, only 3 of the 30 relationships were significantly different at the $p < 0.05$ level. More importantly, the majority of these differences come from comparisons with the worldwide and non-U.S. OCM samples.

Finally, to examine the potential explanation for the lower path coefficient between PU and BI for the worldwide OCM sample, as in our CB-SEM analysis, we estimated two additional models. First, we removed all responses that were from the United States to provide a sample that would be theoretically similar to the non-U.S. OCM sample. The results in Table D4 indicate that doing this creates a model that follows more closely the additional structural models in the analysis with PU playing a slightly stronger role in the influence of BI. If we further reduce this sample to the majority respondent country, India, we find results that provide an even closer relationship to the additional models in the analysis.

Table D1. Correlations and Reliabilities

	Mean	SD	Cronbach's Alpha	Composite Reliability	AVE	BI	PEOU	PU
Students								
Behavioral Intention	5.692	1.543	0.959	0.973	0.924	**0.961**		
Perceived Ease of Use	5.862	0.761	0.897	0.920	0.657	0.331	**0.811**	
Perceived Usefulness	5.583	0.856	0.910	0.930	0.691	0.381	0.458	**0.831**
U.S. OCM								
	Mean	**SD**	**Cronbach's Alpha**	**Composite Reliability**	**AVE**	**BI**	**PEOU**	**PU**
Behavioral Intention	5.340	1.683	0.960	0.974	0.927	**0.963**		
Perceived Ease of Use	5.641	1.120	0.944	0.955	0.781	0.512	**0.884**	
Perceived Usefulness	5.368	1.177	0.953	0.963	0.811	0.592	0.531	**0.900**
Non-U.S. OCM								
	Mean	**SD**	**Cronbach's Alpha**	**Composite Reliability**	**AVE**	**BI**	**PEOU**	**PU**
Behavioral Intention	5.754	1.122	0.855	0.911	0.773	**0.879**		
Perceived Ease of Use	5.912	0.818	0.913	0.932	0.697	0.534	**0.835**	
Perceived Usefulness	5.780	0.943	0.923	0.940	0.723	0.557	0.641	**0.850**
Worldwide OCM								
	Mean	**SD**	**Cronbach's Alpha**	**Composite Reliability**	**AVE**	**BI**	**PEOU**	**PU**
Behavioral Intention	5.813	1.093	0.899	0.937	0.831	**0.912**		
Perceived Ease of Use	5.902	0.801	0.913	0.932	0.696	0.619	**0.834**	
Perceived Usefulness	5.770	0.945	0.925	0.941	0.727	0.498	0.637	**0.853**
Consumer Panel								
	Mean	**SD**	**Cronbach's Alpha**	**Composite Reliability**	**AVE**	**BI**	**PEOU**	**PU**
Behavioral Intention	5.653	1.373	0.970	0.980	0.943	**0.971**		
Perceived Ease of Use	5.575	1.195	0.969	0.975	0.866	0.482	**0.931**	
Perceived Usefulness	5.406	1.242	0.961	0.969	0.838	0.504	0.643	**0.915**

Note: Square-root of the AVE on diagonal.

These results are directly similar to those found in the robustness analysis provided in the CB-SEM analysis and discussed previously in the paper. Therefore, based on the results of our analysis, we find highly consistent results between the CB-SEM technique and the PLS technique, providing increased robustness to our findings.

In addition to the structural model analysis we also conducted a supplementary analysis of the difference between scale levels utilized within the model. This procedure is similar to the steps taken in our CB-SEM group invariance tests; however, due to the inability to control parameter estimates in PLS as in CB-SEM, we examined only the differences in latent variable mean scores as an indication of differences. In Table D5 we depict the differences in the latent variable mean scores utilizing Scheffe's pairwise comparisons. Interestingly, we do find some important differences in the worldwide and non-U.S. OCM samples while the student, consumer, and U.S. OCM samples were highly similar. We found that behavioral intention differs only between the U.S. OCM sample and the worldwide and non-U.S. OCM samples ($p < 0.05$); the student, consumer, and U.S. OCM samples did not significantly differ ($p > 0.05$). For perceived usefulness only the consumer panel and the worldwide OCM differed ($p < 0.05$) while all other samples were similar. Additionally, for perceived ease of use, only the consumer panel differed from both the worldwide ($p < 0.05$) and non-U.S. OCM ($p < 0.01$) samples. Overall, we find that across all theoretical constructs the student, consumer, and U.S. OCM samples did not differ, indicating the potential to utilize U.S. OCMs as a viable alternative to homogenous student samples and expensive consumer panels. However, and more importantly, this analysis also provides evidence of our caution to researchers on the use of worldwide and non-U.S. OCM samples until further research is conducted exploring the causes of these differences among samples.

Table D2. PLS Loadings and Crossloadings

	Students			Worldwide OCM			U.S. OCM			Non-U.S. OCM			Consumer Panel		
	Behavioral Intention	Perceived Ease of Use	Perceived Usefulness	Behavioral Intention	Perceived Ease of Use	Perceived Usefulness	Behavioral Intention	Perceived Ease of Use	Perceived Usefulness	Perceived Usefulness	Perceived Ease of Use	Behavioral Intention	Behavioral Intention	Perceived Ease of Use	Perceived Usefulness
Intention 1	0.968	0.288	0.319	0.929	0.633	0.522	0.970	0.481	0.578	0.907	0.521	0.582	0.980	0.464	0.489
Intention 2	0.943	0.323	0.371	0.887	0.511	0.398	0.940	0.515	0.555	0.842	0.441	0.335	0.954	0.478	0.473
Intention 3	0.973	0.339	0.400	0.919	0.537	0.429	0.978	0.482	0.578	0.887	0.439	0.511	0.979	0.463	0.506
Ease of Use 1	0.142	0.696	0.144	0.443	0.792	0.466	0.375	0.845	0.347	0.429	0.818	0.466	0.474	0.933	0.547
Ease of Use 2	0.217	0.826	0.357	0.602	0.843	0.564	0.447	0.900	0.432	0.428	0.788	0.514	0.444	0.930	0.601
Ease of Use 3	0.287	0.862	0.383	0.512	0.859	0.586	0.483	0.925	0.540	0.388	0.841	0.556	0.446	0.948	0.625
Ease of Use 4	0.339	0.826	0.432	0.481	0.830	0.578	0.539	0.861	0.547	0.469	0.861	0.563	0.471	0.906	0.646
Ease of Use 5	0.294	0.803	0.411	0.473	0.820	0.431	0.367	0.883	0.412	0.419	0.813	0.527	0.413	0.929	0.576
Ease of Use 6	0.258	0.841	0.383	0.566	0.858	0.536	0.458	0.885	0.481	0.532	0.886	0.576	0.441	0.938	0.584
Usefulness 1	0.277	0.335	0.777	0.423	0.550	0.850	0.569	0.510	0.927	0.550	0.503	0.829	0.486	0.640	0.927
Usefulness 2	0.359	0.384	0.857	0.425	0.523	0.848	0.580	0.488	0.915	0.476	0.537	0.850	0.506	0.585	0.935
Usefulness 3	0.303	0.423	0.901	0.454	0.594	0.890	0.531	0.506	0.925	0.449	0.568	0.867	0.475	0.606	0.942
Usefulness 4	0.231	0.398	0.832	0.451	0.562	0.864	0.560	0.493	0.915	0.440	0.581	0.882	0.457	0.594	0.939
Usefulness 5	0.380	0.381	0.841	0.378	0.523	0.845	0.472	0.433	0.868	0.415	0.498	0.828	0.425	0.536	0.892
Usefulness 6	0.332	0.362	0.774	0.412	0.498	0.817	0.475	0.430	0.850	0.500	0.575	0.844	0.411	0.561	0.855

Table D3. Path Coefficient Differences

Path Comparison	PEOU ➜ BI	PEOU ➜ PU	PU ➜ BI
Student vs. U.S.	0.077(0.486)	-0.073(0.468)	-0.157(0.142)
Student vs. Non-U.S.	-0.103(0.490)	**-0.182(0.046)**	-0.074(0.620)
Student vs. Worldwide	**-0.310(0.011)**	**-0.178(0.050)**	0.115(0.300)
Student vs. Consumer	-0.072(0.567)	**-0.184(0.043)**	-0.041(0.715)
U.S. vs. Non-U.S.	-0.027(0.838)	-0.110(0.212)	0.083(0.517)
U.S. vs. Worldwide	**-0.233(0.029)**	-0.106(0.231)	**0.272(0.004)**
U.S. vs. Consumer	0.005(0.970)	-0.112(0.184)	0.116(0.301)
Non-U.S. vs. Worldwide	-0.206(0.178)	0.004(0.961)	0.189(0.193)
Non-U.S. vs. Consumer	0.032(0.836)	-0.002(0.980)	0.033(0.814)
Worldwide vs. Consumer	0.238(0.084)	-0.006(0.941)	-0.156(0.188)

Note: p-value for two-tailed tests in parentheses.

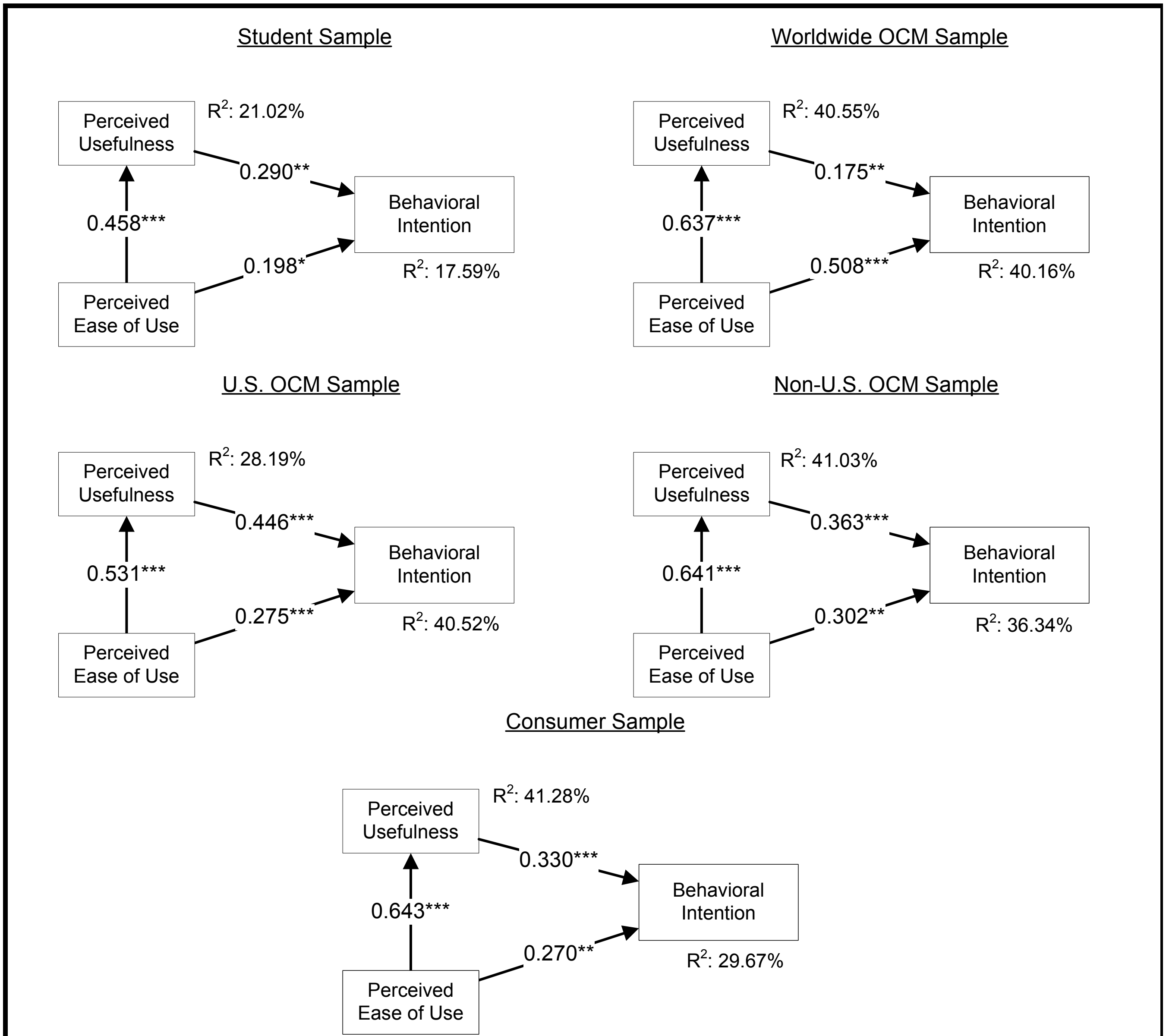


Figure D1. PLS Path Model Results

Table D4. Worldwide Robustness Analysis

	Worldwide	Worldwide Without U.S.	India Only
	n = 193	n = 173	n = 125
PEOU ➜ BI	0.508***	0.470***	0.339**
PEOU ➜ PU	0.637***	0.694***	0.710***
PU ➜ BI	0.175**	0.203*	0.325**
PU – R^2	0.4055	0.4816	0.5038
BI – R^2	0.4016	0.3939	0.3765

Notes: *p < .05; **p < .01; ***p < .001.

Table D5. Scale Mean Differences

		Mean	SD	1	2	3	4	5
Behavioral Intention								
		Mean	**SD**	**1**	**2**	**3**	**4**	**5**
1	Student	5.692	1.543	-				
2	U.S. OCM	5.340	1.683	-0.353	-			
3	Non-U.S. OCM	5.754	1.122	0.061	0.414*	-		
4	Worldwide OCM	5.813	1.093	0.121	0.474*	0.060	-	
5	Consumer Panel	5.653	1.373	-0.039	0.314	-0.100	-0.160	-
Perceived Usefulness								
		Mean	**SD**	**1**	**2**	**3**	**4**	**5**
1	Student	5.583	0.856	-				
2	U.S. OCM	5.641	1.12	-0.219	-			
3	Non-U.S. OCM	5.780	0.943	0.198	0.413	-		
4	Worldwide OCM	5.770	0.945	0.187	0.402	-0.011	-	
5	Consumer Panel	5.406	1.242	-0.176	0.039	-0.374	-0.363*	-
Perceived Ease of Use								
		Mean	**SD**	**1**	**2**	**3**	**4**	**5**
1	Student	5.862	0.761	-				
2	U.S. OCM	5.368	1.177	-0.222	-			
3	Non-U.S. OCM	5.912	0.818	0.050	0.272	-		
4	Worldwide OCM	5.902	0.801	0.039	0.261	-0.011	-	
5	Consumer Panel	5.575	1.195	-0.287	-0.066	-0.337**	-0.326*	-

Notes: *p < .05; **p < .01; ***p < .001.

References

Chin, W. W. 1998. "The Partial Least Squares Approach to Structural Equation Modeling," in *Modern Methods for Business Research: Methodology for Business and Management*, G. A. Marcoulides (ed.), London: Lawrence Erlbaum Associates, pp. 295-336.

Chin, W. W. 2000. "Frequently Asked Questions—Partial Least Squares & PLS-Graph" (http://disc-nt.cba.uh.edu/chin/plsfaq.htm).

Chin, W. W. 2010. "How to Write Up and Report PLS Analyses," in *Handbook of Partial Least Squares*, V. Esposito Vinzi, W. W. Chin, J. Henseler, and H. Wang (eds.), Berlin: Springer, pp. 655-690.

Davis, F. 1989. "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology," *MIS Quarterly* (13:3), pp. 319-340.

Davis, F., Bagozzi, R., and Warshaw, P. R. 1989. "User Acceptance of Computer Technology: A Comparison of Two Theoretical Models," *Management Science* (35:8), pp. 982-1003.

Gefen, D., and Straub, D. 2005. "A Practical Guide to Factorial Validity Using PLS-Graph: Tutorial and Annotated Example," *Communications of the Association for Information Systems* (16:1), pp. 91-109.

Hair, J. F., Black, B., Babin, B., Anderson, R. E., and Tatham, R. L. 2006. *Multivariate Data Analysis* (6th ed.,) Upper Saddle River, NJ: Pearson Prentice Hall.

Ringle, C. M., Wende, S., and Will, A. 2005. "SmartPLS 2.0 M3" (http://www.smartpls.de).

Appendix E

Group Invariance Tests

In the group invariance tests, we used five configuration models (Table E1) for the incremental tests. Model 1 served as the baseline, and we compared each subsequent model with it. If the comparison did not reveal invariance, we delved more deeply to assess where divergence occurred. In determining invariance, we required a ΔCFI less than or equal to 0.01 to indicate invariance (Cheung and Rensvold 2002; Kline 2010). In Table E2 we provide the results of our staged analyses from our group invariance comparisons.

Interestingly, the results show no differences between the student and U.S. OCM samples, the student and worldwide OCM samples, and the worldwide OCM and non-U.S. OCM samples when examining loadings, intercepts, residuals, and means based upon ΔCFI > 0.01 criteria. Across all comparisons the five samples were invariant in their loadings and intercepts, indicating similarity between the various recruitment methods. However, once the models evolved to include variances and covariances within the model, the consumer panel showed differences compared to all other samples. Additionally, the non-U.S. OCM sample differed from the student and U.S. OCM samples. However, the group invariance tests did indicate that some of the samples were invariant, specifically the student and worldwide OCM, student and U.S. OCM, and worldwide and non-U.S. OCM. The worldwide and U.S. OCM samples just barely exceeded the threshold on model 5, indicating a slight difference in means.

While these tests provide an indication of how the models are invariant between samples as a whole, a deeper understanding of the differences in specific scale values will provide further insights into potential biases. Utilizing the latent variable scores provided by the associated loadings for each model, we examine the difference in means to determine sample differences in further detail. When examining the structural models as a whole, the means were not significantly different between the models of the worldwide and non-U.S. samples, while others indicated variation. Therefore, to provide a clearer picture of which scales are differing across each sample, we conducted a series of ANOVA and pair-wise comparison tests with a Scheffe's correction (see Table E3). Results indicate that the perceived usefulness does not differ among the five samples ($p > 0.05$) while perceived ease of use ($p < 0.001$) and behavioral intention ($p < 0.001$) do show some differences. The mean values of perceived ease of use for the non-U.S. OCM sample differs from all other samples ($p < 0.001$) while the U.S. OCM sample only slightly differs from the consumer panel ($p < 0.05$). Additionally, the mean value of behavioral intention for the U.S. OCM sample differs from all other samples ($p < 0.01$).[3] One thing to note from this subset analysis is that the majority of the comparisons show no differences between the worldwide and non-U.S. OCM samples as well as the student and U.S. OCM samples, indicating the potential for interchangeability.

Therefore, based on our group invariance and pairwise comparisons, it appears that there are indeed differences in the scale variances and means. However, the majority of the comparisons show invariance between the samples with only a few skewing the comparisons indicated in the overall analyses. Specifically, we find that the student, consumer panel, and U.S. OCM samples are fairly consistent within their estimations, indicating the potential for interchangeability.

References

Cheung, G. W., and Rensvold, R. B. 2002. "Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance," *Structural Equation Modeling A Multidisciplinary Journal* (9:2), pp. 233-255.

Kline, R. B. 2010. *Principles and Practice of Structural Equation Modeling*, New York: The Guildford Press.

Table E1. Group Invariance Testing Models

Step	Model Description	Constrained Parameters
Model 1	Baseline Invariance	None
Model 2	Weak Invariance	Factor loadings
Model 3	Strong Invariance	Factor loadings and intercepts
Model 4	Strict Invariance	Loadings, intercepts, and residuals (variances and covariances)
Model 5	Very Strict Invariance	Loadings, intercepts, residuals, and means

[3]U.S. OCM differs from consumer panel ($p < 0.05$).

Table E2. Group Invariance Testing

Model Statistics		Student vs. Worldwide	Student vs. U.S.	Student vs. Non-U.S.	Worldwide vs. U.S.	Worldwide vs Non-U.S.	U.S. vs. Non-U.S.	Consumer Panel vs. U.S.	Consumer Panel vs. Non-U.S.	Consumer Panel vs. Worldwide	Consumer Panel vs. Student
Model 1	CHISQ	484.157	490.811	458.415	624.239	591.842	598.497	617.49	585.09	610.83	477.41
	DF	174	174	174	174	174	174	174	174	174	174
	PVALUE	0	0	0	0	0	0	0	0	0	0
	CFI	0.925	0.944	0.933	0.920	0.908	0.930	0.95	0.946	0.942	0.958
	RMSEA	0.100	0.097	0.093	0.112	0.109	0.106	0.104	0.1	0.106	0.091
	BIC	12656.264	14155.51	13663.53	14849.481	14357.6	15856.05	16566.1	16074.467	15068.3	14374.8
Model 2	CHISQ	497.132	501.663	487.16	636.862	604.532	631.8	643.35	610.42	617.52	497.446
	DF	186	186	186	186	186	186	186	186	186	186
	PVALUE	0	0	0	0	0	0	0	0	0	0
	CFI	0.925	0.944	0.929	0.920	0.908	0.926	0.949	0.944	0.942	0.956
	RMSEA	0.097	0.094	0.092	0.109	0.106	0.105	0.102	0.099	0.102	0.089
	BIC	12598.673	14094.86	13620.99	14789.94	14298.33	15816.56	16518	16026	15001.8	14322.2
Model 1 vs. Model 2	ΔCFI	0.000	0.000	0.004	0.000	0.000	0.004	0.002	0.002	0.001	0.001
	ΔCHISQ	12.976	10.852	28.745	12.624	12.69	33.303	25.86	25.333	6.682	20.041
Model 3	CHISQ	531.103	507.074	522.343	670.091	614.059	667.306	681.19	640.14	640.47	539.461
	DF	198	198	198	198	198	198	198	198	198	198
	PVALUE	0	0	0	0	0	0	0	0	0	0
	CFI	0.920	0.945	0.923	0.921	0.909	0.922	0.946	0.942	0.941	0.952
	RMSEA	0.097	0.090	0.093	0.108	0.102	0.105	0.101	0.098	0.1	0.09
	BIC	12562.076	14028.77	13584.89	14751	14235.9	15779.28	16481.9	15982	14951.5	14291.6
Model 1 vs. Model 3	ΔCFI	0.006	0.001	0.009	0.001	0.000	0.007	0.004	0.004	0.001	0.005
	ΔCHISQ	46.946	16.263	63.927	45.852	22.216	68.809	63.7	55.051	29.635	62.056
Model 4	CHISQ	565.993	558.408	614.915	746.524	644.093	815.58	806.88	976.083	827.3	655.627
	DF	213	213	213	213	213	213	213	213	213	213
	PVALUE	0	0	0	0	0	0	0	0	0	0
	CFI	0.915	0.939	0.905	0.910	0.905	0.900	0.934	0.899	0.918	0.938
	RMSEA	0.096	0.092	0.100	0.111	0.100	0.115	0.108	0.124	0.114	0.099
	BIC	12508.759	13990.72	13588.36	14737.231	14175.99	15836.56	16515.1	16225.7	15046.8	14317.1
Model 1 vs Model 4	ΔCFI	0.010	0.005	0.028	0.010	0.003	0.029	0.017	0.047	0.024	0.019
	ΔCHISQ	81.836	67.597	156.5	122.285	52.251	217.083	189.39	390.99	216.44	178.223
Model 5	CHISQ	570.331	566.056	619.452	764.886	644.704	892.017	818.15	993.74	841.52	667.032
	DF	216	216	214	216	216	220	216	216	216	216
	PVALUE	0	0	0	0	0	0	0	0	0	0
	CFI	0.915	0.938	0.904	0.908	0.906	0.889	0.933	0.897	0.916	0.937
	RMSEA	0.096	0.082	0.100	0.111	0.099	0.119	0.108	0.124	0.114	0.099
	BIC	12495.455	13980.5	13586.96	14737.552	14158.61	15870.53	16507.919	16224.9	15042.8	14310.3
Model 1 vs Model 5	ΔCFI	0.010	0.006	N/A	0.013	0.002	N/A	N/A	N/A	N/A	N/A
	ΔCHISQ	86.174	75.244	N/A	140.647	52.862	N/A	N/A	N/A	N/A	N/A
Model 4 vs Model 5	ΔCFI	N/A	N/A	0.000	N/A	N/A	0.002	0.001	0.002	0.002	0.001
	ΔCHISQ	N/A	N/A	4.538	N/A	N/A	16.051	11.622	17.654	14.266	11.404

Table E3. Scale Mean Differences								
Behavioral Intention								
		Mean	**SD**	**1**	**2**	**3**	**4**	**5**
1	Student	5.776	1.609	-				
2	U.S. OCM	5.232	1.723	-0.544**	-			
3	Non-U.S. OCM	5.750	1.008	-0.021	0.518**	-		
4	Worldwide OCM	5.749	0.995	-0.267	0.518**	-0.001	-	
5	Consumer Panel	5.613	1.374	-0.163	0.381	-0.137	-0.136	-
Perceived Usefulness								
		Mean	**SD**	**1**	**2**	**3**	**4**	**5**
1	Student	5.589	0.839	-				
2	U.S. OCM	5.549	1.190	-0.040	-			
3	Non-U.S. OCM	5.566	0.869	-0.232	0.017	-		
4	Worldwide OCM	5.783	0.907	0.194	0.234	0.217	-	
5	Consumer Panel	5.730	1.288	0.141	0.182	0.165	-0.052	-
Perceived Ease of Use								
		Mean	**SD**	**1**	**2**	**3**	**4**	**5**
1	Student	5.678	0.736	-				
2	U.S. OCM	5.953	1.189	0.275	-			
3	Non-U.S. OCM	5.151	0.704	-0.526***	-0.802***	-		
4	Worldwide OCM	5.688	0.762	0.010	-0.265	0.537***	-	
5	Consumer Panel	5.670	1.211	-0.008	-0.283*	0.518***	-0.018	-

Notes: *p < .05; **p < .01; ***p < .001.