Topical Article

SOCIETY FOR THE TEACHING OF PSYCHOLOGY

Teaching of Psychology
39(4) 245-251

DOI: 10.1177/0098628312456615
http://top.sagepub.com

SAGE

Participants at Your Fingertips: Using Amazon's Mechanical Turk to Increase Student–Faculty Collaborative Research

Dan R. Johnson[1] and Lauren A. Borden[1]

Abstract

The literature suggests students gain important skills when directly involved with faculty in research. However, students at smaller institutions are often faced with limited research opportunities and faculty members are faced with limited participant-pools, funding, and space to perform research. Amazon's Mechanical Turk (MTurk) may provide a solution to many of these problems. MTurk provides an online human participant-pool, along with tools to build experiments, and it allows data to be collected quickly and inexpensively. In this study of narrative fiction and empathy, data was collected using the traditional, laboratory-based approach, and on MTurk using identical measures and protocols. Results indicated MTurk data exhibits comparable reliability, gender and ethnicity composition to data collected in the laboratory. Two important differences emerged: MTurk participants were 10 years older, on average, and they demonstrated higher scores on trait measures of empathy and state measures of involvement into the story presented in the study. A brief user's guide to MTurk is presented that caters to first-time users. Finally, common pitfalls and their solutions are presented with the hope that faculty and students can begin doing research on MTurk immediately.

Keywords

student research, internet data collection, online, research skills

Direct involvement in the research process has many benefits for students including gaining technical and interpersonal skills, analytical and independent learning skills, and improving their chances of being accepted into competitive graduate programs (Ishiyama, 2002; Kierniesky, 2005; Landrum & Nelsen, 2002). Although student–faculty collaborative research has increased recently in smaller psychology departments, significant challenges remain (Kierniesky, 2005). Challenges to students include a lack of awareness and limited access to research opportunities (Perlman & McCann, 2005; Wayment & Dickson, 2008). In addition, for faculty at smaller institutions, the traditional approaches to data collection often exceed the financial resources, participant-pool availability, facilities, and staff resources necessary for successful research (Marek, Christopher, & Walker, 2004). Amazon's Mechanical Turk (MTurk) offers a promising solution to many of the aforementioned challenges to performing research at smaller institutions. MTurk is an internet-based platform that provides: (1) an online and diverse participant-pool (Buhrmester, Kwang, & Gosling, 2011; Pontin, 2007), (2) tools for survey and experiment creation, and (3) a method of rapid and inexpensive data collection.

At smaller institutions, one primary constraint on student–faculty research collaboration is a small or nonexistent participant-pool. Without a large participant-pool, data collection can proceed so slowly that collecting enough data for even one study can take years. This prohibitively slow pace makes class research projects, senior theses with data collection, and directed research burdensome if not impossible at smaller schools. Often a student and faculty member will have only one semester to design and complete the data collection process for a study. The unfortunate consequence is that student-learning about research ends with a review of the literature and perhaps a research proposal that never comes to fruition. Through MTurk's large and readily accessible participant-pool, data can easily be collected in one semester, often in less than a week, thereby making these short-term collaborative research projects feasible.

An additional constraint for faculty at smaller schools is often limited funding and space to perform research. MTurk offers a solution to both of these problems as well because the participants on MTurk will reliably complete studies for minimal payment (e.g., as little at 2 cents, see Buhrmester et al., 2011). In addition, lab space for data collection is unnecessary

[1] Psychology Department, Washington and Lee University, Lexington, VA, USA

Corresponding Author:
Dan R. Johnson, Psychology Department, Washington and Lee University, Parmly Hall, 230, Lexington, VA 24450, USA
Email: johnsondr@wlu.edu

because data is collected entirely on the internet. The faculty member and student researcher merely need a computer to create their MTurk study and to collect and analyze the data.

Given the promise of MTurk as a vehicle for student–faculty collaborative research at smaller institutions, we hope to promote its use for research by providing: (1) an introduction to MTurk and its terminology, (2) a brief user's guide, (3) a comparison of data collected in the laboratory and on MTurk, and (4) a list of common pitfalls and their solutions (see Discussion).

An Introduction to Amazon's MTurk

The researcher (i.e., requester) creates a study (i.e., human intelligence task, HIT) by modifying free templates that are available on the MTurk website (https://www.mturk.com) that allow researchers to replicate almost any task that can be performed on a computer. The participant-pool on MTurk consists of workers, or individuals over 18 years of age who have signed up to receive payment for completing studies. The workers can browse through thousands of studies (i.e., HITs) and search by compensation amount, time allotted to complete the HIT, date posted, and/or keyword. The financial burden to researchers is quite low as workers complete HITs for as little compensation as 2 cents. Once Workers complete the HIT and the researcher reviews and accepts the data, Workers are paid online by Amazon from an account funded by the researcher. For use of their website and participant-pool, Amazon's commission is 10% of the total compensation amount.

A Brief User's Guide to MTurk

Getting started. To get started on MTurk, the researcher should sign up for an account by entering an email address and creating a password. Then, the researcher should take the informative free tour of the website for a general overview of MTurk protocol. Next, the researcher should visit the Requester Sandbox (https://requestersandbox.mturk.com), where one can create a HIT, perform troubleshooting, and collect pilot data (with lab members first) without the HIT being visible to Workers[1] or distributing payment. There are six main tabs at the Requester Sandbox including Home, Design, Publish, Manage, Developer, and Help. Under the Design tab, the researcher will want to examine the freely available templates and determine which suits the study best. After choosing a template from which to start, the researcher then chooses how many Workers are needed for the study, the time the researcher would like to allot for each Worker to complete the study, and how much Workers are to be paid.

The researcher can also impose nearly any type of screening criteria (e.g., women only, 70 years and older) to select which Workers will complete the HIT. These participant screening mechanisms, called Qualifications, are quite valuable and flexible. For example, an aging researcher might want to collect data from 18 to 30 years old and adults over 70. A customized Qualification that meets these screening criteria can be created. Another Qualification may be to select Workers from a specific region, such as the U.S. or Michigan. In addition, researchers may want to select only "high quality" Workers by requiring that they have a 95% or above approval rating. The approval rating is simply the percentage of HITs the Workers have completed reliably in the past in the judgment of other researchers, who of course are also concerned with quality data. If the Worker has participated in 20 studies, they would need to be approved in 19 of them to meet this approval rating.

Designing your HIT/study. When building a pilot HIT in the Requester Sandbox, the researcher will want to ensure the HIT is user-friendly. Workers have thousands of options and if a HIT is cumbersome to take, they will not complete additional HITs from that researcher. Strategies for creating HITs easy to complete include making text large and easy to read, minimizing the amount of scrolling, and making sure the HIT is organized and each item is well spaced. Consider using text size and/or color differences to highlight important instructions and distinguish sections from each other.

Next, the researcher should sign up as a Worker in the Worker Sandbox (https://workersandbox.mturk.com) and complete a number of HITs to get ideas on how to address formatting and spacing issues in addition to any problems specific to the researcher's study. For more sophisticated solutions, learning some HTML programming is highly recommended (e.g., *HTML for Dummies*, Titel & Noble, 2011). To view and edit the HTML code responsible for generating the template, the Requester can simply click the "Edit HTML source" function in the Design Layout section of the Design tab.

Publishing your HIT/study. Before paying Workers to complete the HIT (at https://www.mturk.com), the researcher should collect a few pilot data points by publishing the HIT in the Requester Sandbox. Allow only a few Workers to participate by specifying how many Workers will be piloting the study in the "Number of assignments per HIT" section of the template. This will allow members of the researcher's lab team to sign-up and produce pilot data, but will prevent other Workers from having access to the HIT. Then, check for clarity, errors, and ease of completion. Next, generate keywords that Workers will use to search for the HIT. It is critical to generate numerous keywords so that a diverse set of Workers will be able to find and select the HIT and, of course, beware of selection bias. Next, the researcher determines how much to pay each Worker to complete the HIT and the time allotted. The typical amount to pay a Worker for a 30-min survey, for example, appears to be 50 cents or less (see Buhrmester et al., 2011) and still results in rapid data collection.

Importantly, Buhrmester, Kwang, and Gosling (2011) showed that compensation amount does not influence data reliability, but smaller compensation amounts can slow data collection. However, even a 30-min survey in which Workers were paid only 10 cents resulted in over 6 participants being collected per hour so that a 150-person study

could potentially be performed for $15 in approximately 1 day (Buhrmester et al., 2011). The researcher then uses a credit card to fund the requisite amount for the study. Lastly, MTurk will allow the HIT to be published and release it to Workers.

Managing your HIT/study. Once the HIT is published, Workers can begin completing it immediately. View the percentage of participants who have completed the HIT out of the total number in the study using the Manage tab. As Workers complete the HIT, the researcher should screen the raw data to ensure completeness and accuracy. After screening each Worker's data, it is entirely up to the researcher whether to "accept" or "reject" their data. If the researcher rejects the data, the Worker is not paid, and it negatively impacts the Worker's approval rating. As a result, Workers are highly motivated to maintain a high approval rating and reliably complete the study. It is recommended the researcher create a contact e-mail where Workers can report any technical difficulties or points of confusion. In the authors' experience, Workers often offer thoughtful suggestions on how to improve the HIT (formatting, layout, etc.). This feedback can be useful in designing future studies. When data collection is complete, it is available in an Excel file that can easily be imported into SPSS or some other statistics package for data analysis.

Comparison of Data Collected in the Laboratory and MTurk Data

Although MTurk has only recently been used to perform psychological research, it has been used fruitfully for research in social psychology (Fishbach, Henderson, & Koo, 2011; Gómez et al., 2011), cognitive psychology (Eriksson & Simpson, 2011), and linguistics (Sprouse, 2011). However, direct comparisons between data collected in the laboratory and MTurk are sparse. A recent study compared the reliability of MTurk data to reliabilities observed in previous studies and found MTurk data to be as reliable and in some cases more reliable than data collected in the laboratory (Buhrmester et al., 2011). This is notable given the significant loss of control when collecting data outside of the laboratory.

The first author (faculty member) and the second author (an undergraduate student) performed two collaborative studies that were identical in every way except in one study, data were collected in the laboratory and in other, data was collected on MTurk. This provides a rare, direct comparison between MTurk and lab data with identical protocols. Sixty-two participants were collected for each sample and two Qualifications were used to screen MTurk participants including selecting only those living in the United States and those holding at least a 95% HIT approval rating. In these studies, the relationship between reading narrative fiction and empathy was investigated.[2] After reading a fictional story, participants reported the empathy they felt for the story's characters and the degree to which they were transported into the story. Two overarching types of empathy were measured: trait and state empathy. Trait empathy consists of four smaller facets including one's general tendency to feel tender, concerned feelings for another (Empathic Concern), become immersed in a story and connect with its characters (Fantasy), become anxious in stressful situations (Personal Distress), and often step into someone else's shoes to try and understand their feelings (Perspective-Taking) (Interpersonal Reactivity Index; Davis, 1983). State empathy was assessed by asking participants how much they felt sympathetic, soft-hearted, warm, compassionate, tender, and moved for the characters in the story while reading (Batson, Early, & Salvarni, 1997). Finally, transportation into the story was measured with three facets including cognitive engagement, emotional involvement, and the level of imagery experienced (Green & Brock, 2000). Please see Johnson (2012) for further details.

To determine the comparability between lab and MTurk data, Table 1 presents demographics, descriptive statistics, correlations between primary measures, and reliability of the primary measures from both data sources. The demographic characteristics of MTurk participants revealed that two-thirds were women. This proportion was marginally more than the proportion of women in the lab sample.[3] Although the age range was similar to lab data, the mean age was significantly higher by 10 years in MTurk sample.[4] In addition, the age standard deviation of the MTurk sample was nearly double that of the lab sample. The ethnic diversity was comparable across MTurk and lab data, where nearly 80% of both samples were White.

All four facets of trait empathy in addition to state empathy and transportation were compared on their internal consistency (Cronbach's α) across lab and MTurk data. Lab and MTurk data for all measures were comparable in that the largest different between reliabilities was .143 and all values were within the range of those reported in previous studies of trait empathy (Davis, 1983, range = .71–.77; Fernandez, Dufey, & Kramp, 2011, range = .66–.78) and transportation (Green & Brock 2000, $\alpha = .76$).

The means of the measures tended to vary more between lab and MTurk samples as compared to internal consistency. Independent t tests were performed and as Table 1 shows, compared to the lab sample, the MTurk sample exhibited similar levels of State Empathy and Perspective-Taking and significantly higher Fantasy, Personal Distress, and Transportation, with marginally higher Empathic Concern. While the MTurk sample tended to have higher standard deviations across measures, the additional variability did not prevent comparable or significantly higher scores for the MTurk sample. These results are consistent with Buhrmester et al.'s (2011) findings that MTurk participants may be more intrinsically motivated than lab participants given their willingness to participate for minor compensation. This helps to explain MTurk participants' somewhat higher levels of trait empathy and enhanced willingness or ability to become transported into the study's story. These results provide an important extension of Buhrmester et al.'s (2011) findings because this study provides a direct

Table 1. Comparison of Data Collected on MTurk and Lab Data

	Lab Data	MTurk Data
Demographics	[3]Men = 30, Women = 32 Mdn_{age} = 20, Range = 18–53 [4]M_{age} = 21, SD_{age} = 6 White = 81%, Black = 13%[a]	Men = 20, Women = 42 Mdn_{age} = 29, Range = 18–60 M_{age} = 33, SD_{age} = 11 White = 75%, Black = 13%[a]
Empathic Concern		
M	26.63	28.06
SD	5.00	4.40
Cronbach's α	.82	.80
Independent *t* test	$t(122) = 1.96$, $p = .094$, $d = .30$	
Fantasy		
M	24.45	26.35
SD	4.96	5.07
Cronbach's α	.71	.81
Independent *t* test	$t(122) = 2.11$, $p = .037$, $d = .38$	
Personal Distress		
M	16.52	19.19
SD	5.25	6.46
Cronbach's α	.81	.88
Independent *t* test	$t(122) = 2.53$, $p = .0128$, $d = .45$	
Perspective taking		
M	24.44	25.31
SD	5.57	4.73
Cronbach's α	.84	.83
Independent *t* test	$t(122) = .94$, $p = .350$, $d = .17$	
State empathy		
M	3.63	3.65
SD	.81	.88
Cronbach's α	.86	.83
Independent *t* test	$t(122) = .13$, $p = .896$, $d = .02$	
Transportation		
M	48.82	59.32
SD	7.35	11.37
Cronbach's α	.75	.89
Independent *t* test	$t(122) = 6.11$, $p < .001$, $d = 1.10$	
Correlation	Empathic Concern State Empathy	Empathic Concern State Empathy
Fisher's *r* to *z*	$r = .26$, $p = .038$, $p = .179$,	$r = .41$, $p < .001$
Correlation	Transportation State Empathy	Transportation State Empathy
Fisher's *r* to *z*	$r = .39$, $p < .001$; $p = .054$	$r = .61$, $p < .001$

Note. [a]The remaining ethnicity percentages consisted mainly of Asian-Americans and Latinos.

comparison between two identical protocols given in the lab and on MTurk. In addition, this study required participants to perform an involved protocol including reading a story, completing both trait and state measures that took many participants 45 min to complete. Consequently, this data demonstrate MTurk participants' willingness and indeed potentially greater engagement in a longer study protocol, as shown in significantly higher scores on state measures, compared to a lab sample. It is possible that the larger values for MTurk could be due to the age difference between the lab and MTurk samples, but this is speculative. In addition, one would not expect any two samples to demonstrate equivalent descriptive statistics regardless of the data source.

Finally, using Fisher's *r* to *z* transformation to compare correlation coefficients, Table 1 shows that the correlation between Empathic Concern and State Empathy was directionally higher in the MTurk sample ($r = .41$), but not significantly higher than the lab sample ($r = .26$). However, the correlation between the two state measures, Transportation and State Empathy, was marginally higher in the MTurk sample

($r = .61$), compared the lab sample ($r = .39$). These results provide further evidence of the viability of MTurk data in that the effect size of interrelationships is also comparable or higher in the MTurk sample.

To summarize the results, MTurk and lab data exhibited an overall comparability. This is notable and perhaps even surprising given the researcher's lack of control over MTurk participants who could be completing the HIT from home, work, or some other location. However, two important differences between the samples were: (1) a significantly higher average age in the MTurk sample and (2) significantly higher scores on some trait empathy facets and the state measure of transportation. Regarding reliability, descriptive statistics, and correlations of the primary dependent measures, similarity appeared to outweigh differences. In addition, given the older age of MTurk participants compared to lab participants, one might predict some differences between the samples regardless of the online versus lab format.

General Discussion

There are many challenges to performing research at smaller institutions including limited research opportunities for students and limited participant-pool availability, time, funding, and space for faculty (Marek et al., 2004; Perlman & McCann, 2005; Wayment & Dickson, 2008). MTurk offers a potential solution to these problems by providing a very large online participant-pool and a way to rapidly and inexpensively collect data.

While previous studies have indicated MTurk provides reliable data (Buhrmester et al., 2011), the current study examined this issue further by directly comparing lab and MTurk data collected using identical and time-consuming research protocols and observing interrelationships among measures. Overall, there was very good comparability between lab and MTurk data and dependent measures on MTurk demonstrated reliability at or better than shown in previous studies. However, the MTurk sample was about 10 years older, on average, than the lab sample and demonstrated greater variability in age as well. The important difference in age distribution has advantages and disadvantages. For studies looking to improve external validity, MTurk may have an advantage in representativeness of the adult population. For studies focused on a small experimental effect, age variability may add unwanted noise to the data and make it harder to observe condition differences.[5] In addition, MTurk participants' scores were significantly higher on some trait and state measures, indicating potentially more engaged or intrinsically motivated participants. While speculative, this could be due to the older age of the MTurk participants or something inherent to their willingness to complete a study for minimal payment.

A primary goal of this article is both to introduce faculty at smaller institutions to MTurk and save them valuable time and difficulty in using MTurk with their students. The brief user's guide was provided to introduce faculty to MTurk so they can determine whether it is a tool that may serve them. For those who are seriously considering using MTurk, a list of practical limitations inherent to MTurk, common pitfalls, and their potential solutions are outlined below.

MTurk Limitations/Pitfalls and Solutions

Workers rushing through HITs. Unlike the laboratory, it is difficult to control how much time participants spend on each section of the protocol. The primary concern is often that the Worker simply answers questions randomly or rushes through the study. One possible strategy for preventing this from occurring is to be sure to check the "Time Completed" data column for each Worker to make sure they took a reasonable amount of time to complete the study. Additionally, if researchers use a Qualification to select Workers with a 95% approval rating or above, this generally minimizes the likelihood that this will occur. A third solution is that if surveys are used, one can capitalize on the reverse-coded items by computing a difference score between similar items on a scale that are positively coded and reverse-coded. The larger the difference score, the more reliable the Worker is likely responding. Finally, just like a lab study, build in instructional manipulation checks to ensure the Worker is thoughtfully following instructions. If Workers fail any of these screening criteria, the researcher can simply reject their data without paying them.

Workers returning to completed sections. A significant limitation of MTurk's freely available templates is that they are essentially like one continuous document, so the researcher cannot separate the study's components into different pages. The problem is that Workers can always return to a previously completed section. If one is doing a memory study, for example, Workers can go back and look at the word list presented earlier in the study. One solution that was utilized in the present MTurk study involves creating hyperlinks to navigate to each section of the study. Then, the researcher simply adds considerable space in between each section of the study so that it is very cumbersome for the participant to scroll back to completed sections. Workers do have a limited time to complete the HIT, and the more difficult one makes it for them to return to completed sections, the harder it will be for the Worker to return to completed sections and still complete the HIT on time. A second solution is to link to other survey tools (e.g., SurveyMonkey) that allow separate page creation. A third, more sophisticated solution is to link the HIT to a separate webpage in another domain and then add in separate pages via HTML programming. This requires more advanced knowledge of HTML programming but allows the researcher to build the HIT to solve this problem and any others one may encounter.

Workers completing multiple between-participant conditions or the same HIT on multiple occasions. MTurk prevents Workers from taking the same HIT multiple times, so intervention on the researcher's part is not required to prevent this from occurring. However, if one is doing an experiment with multiple between-participant conditions and does not want Workers to complete the control and experimental conditions, one can publish each

condition under the same name. It is critical that the researcher changes the HIT template name (e.g., Control vs. Experimental) but publishes the identical HIT title for each condition (e.g., Take Survey). Given that the Workers' identification numbers are recorded with the rest of their data, researchers can check to make sure that each Worker only completed one of the study conditions.

What type of study does not work well on MTurk? Currently, the freely available templates do not allow one to collect response time data. As a result, experiments requiring millisecond precision timing of stimulus presentation or data collection cannot be performed using default templates. As a result, MTurk templates are best suited for studies that rely on survey methods and picture ratings. However, with more sophisticated HTML programming and a link to another domain, nearly any experiment can be created on MTurk, including response time data collection. In fact, many of the studies posted by universities link to their own domain to perform this type of research.

Conclusions

It is hoped that the reliability of MTurk data shown here encourages psychology instructors and students to add it to their existing research toolbox. MTurk offers an especially promising solution to many of the challenges faced by faculty at smaller institutions, like a small participant-pool and limited research space. In addition, research performed on MTurk could play a new role in the psychology course curriculum for courses like research methods, senior thesis, and upper division content-specific courses. Given the extremely large and diverse participant-pool available on MTurk and the Qualification screening tool, researchers at smaller institutions now have access to cross sections of people previously thought impossible or prohibitively burdensome to recruit. Adding further credence to MTurk data, the current results indicated MTurk participants may be even more likely to engage in the research protocol than lab participants, although the reason for this remains unknown. Armed with a basic user's guide and solutions to pitfalls provided in this article, it is hoped psychology faculty and students can benefit from research performed on MTurk right now.

Acknowledgments

The author would like to thank the editor and two anonymous reviewers for their exceptionally thoughtful and constructive comments on a previous version of this article.

Declaration of Conflicting Interests

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

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Financial support from the Lenfest Research Fund from Washington and Lee University exclusively for compensation of participants.

Notes

1. The HIT is visible to Workers in the Worker Sandbox, but it is not visible to Workers for payment. This does not pose a problem as long as lab members sign up shortly after your pilot HIT is posted in the Requester Sandbox.
2. A portion of the lab data is used for a different purpose in another study (Johnson, 2012).
3. Fisher's Exact test, $p = .10$.
4. Independent t test, $t(122) = 7.54$, $p < .0001$, $d = 1.35$.
5. Thanks to a reviewer for this insight.

References

Batson, C. D., Early, S., & Salvarni, G. (1997). Perspective taking: Imagining how another feels versus imagining how you would feel. *Personality and Social Psychology Bulletin*, *23*, 751–758. doi:10.1177/0146167297237008

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's mechanical turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, *6*, 3–5. doi:10.1177/1745691610393980

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, *44*, 113–126. doi:10.1037/0022-3514.44.1.113

Eriksson, K., & Simpson, B. (2011). Perceptions of unfairness in allocations between multiple recipients. *Cognitive Psychology*, *62*, 225–244. doi:10.1016/j.cogpsych.2011.01.001

Fernandez, A. M., Dufey, M., & Kramp, U. (2011). Testing the psychometric properties of the interpersonal reactivity index (IRI) in Chile: Empathy in a different cultural context. *European Journal of Assessment, 27*, 179–185. doi: 10.1027/1015-5759/a000065

Fishbach, A., Henderson, M. D., & Koo, M. (2011). Pursuing goals with others: Group identification and motivation resulting from things done versus things left undone. *Journal of Experimental Psychology: General*, *140*, 520–534. doi:10.1037/a0023907

Gómez, A., Brooks, M. L., Buhrmester, M. D., Vásquez, A., Jetten, J., & Swann Jr., W. B. (2011). On the nature of identify fusion: Insights into the construct and a new measure. *Journal of Personality and Social Psychology*, *100*, 918–933. doi:10.1037/a0022642

Green, M. C., & Brock, T. C. (2000). The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology*, *79*, 701–721. doi:10.1037/0022-3514.79.5.701

Ishiyama, J. (2002). Does early participation in undergraduate research benefit social science and humanities students? *College Student Journal*, *36*, 380–386.

Johnson, D. R. (2012). Transportation into a story increases empathy, prosocial behavior, and perceptual bias toward fearful expressions. *Personality and Individual Differences*, *52*, 150–155. doi:10.1016/j.paid.2011.10.005

Kierniesky, N. C. (2005). Undergraduate research in small psychology departments: Two decades later. *Teaching of Psychology*, *32*, 84–90. doi:10.1207/s15328023top3202_1

Landrum, R. E., & Nelson, L. R. (2002). The undergraduate research assistantship: An analysis of the benefits. *Teaching of Psychology*, *29*, 15–19. doi:10.1207/S15328023TOP2901_04

Marek, P., Christopher, A. N., & Walker, B. J. (2004). Learning by doing: Research methods with a theme. *Teaching of Psychology*, *31*, 128–131.

Perlman, B., & McCann, L. I. (2005). Undergraduate research experiences in psychology: A national study of courses and curricula. *Teaching of Psychology*, *32*, 5–14. doi:10.1207/s15328023top3201_2

Pontin, J. (2007, March 25).Artificial intelligence: With help from the humans. *The New York Times*. Retrieved from http://www.nytimes.com/2007/03/25/business/yourmoney/25Stream.html

Sprouse, J. (2011). A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory. *Behavior Research Methods*, *43*, 155–167. doi:10.3758/s13428-010-0039-7

Titel, E., & Noble, J. (2011). *HTML, XHTML, & CSS for dummies* (7th ed.). Hoboken, NJ: Wiley.

Wayment, H. A., & Dickson, K. L. (2008). Increasing student participation in undergraduate research benefits students, faculty, and department. *Teaching of Psychology*, *35*, 194–197. doi:10.1080/00986280802189213