USF® Muma Business Review
A publication of the Muma College of Business | University of South Florida
Volume 7, Number 12 | Empirical Findings | 31 DECEMBER 2023

Crowdsourced Online Biometric Studies: Is the juice worth the squeeze?

By
Robert Hammond, University of South Florida
Yuqi Wang, University of South Florida

Otherwise promising, online biometric studies with crowdsourced participants can have vagaries that limit success. This paper shares researcher experience with biometric online data collection and a study that compared the differences of online and lab-based eye tracking (ET) and facial expression (FE) data.

Students and the public were separated into two groups to place orders from a well-known restaurant online menu application. Group 1 accessed the menu web application routed through an online biometric data collection tool while Group 2 interacted with the menu web application through a desktop application that collected the biometric data.

The study found little difference in the user experience metrics, the facial expression data, or the number of gaze points between the two data collection methods. Interestingly and consistent with field experience, a significant difference in the ET data was observed.

The vagaries of the field present significant hazards for researchers and even more so if study participants are compensated. Recommendations for researchers to optimize research effort are provided regarding recruitment, participant compensation, analysis effort and expertise, maximizing the usable data, and employing eye tracking tools.

Crowdsourcing respondents for biometric online studies present opportunities to expand subject pools, reduce research costs, while simultaneously accelerating data collection and obtaining data in more ecologically similar environments.

Keywords: Affective Computing, Emotional AI, Biometrics, Online Data Collection, Crowdsourcing, Eye Tracking, Facial Expression Analysis, Human Insights

Our experience with biometric online data collection was born from necessity. We had just opened a new university human insights lab equipped with multiple biometric sensors and received external funding for a project, when a pandemic closed the campus. We were unable to extend the funding timeline and thus were faced with the choice of either foregoing the research or finding a new path forward. We chose to move forward by collaborating with the lab's technology partner (Pedersen, 2021) to develop the capability to collect facial expression and eye tracking data online.

The data and studies suggest online biometric data collection is a valuable tool for connecting with uncommon population segments (Parvanta et al., 2022) and can be used to collect data quickly. Our experience is also consistent with prior research describing challenges with crowdsourced populations. We have encountered significant loss of usable data with field online data collection (Hammond et al., 2022; Parvanta et al., 2022). Moreover, eye tracking (ET) visualizations from online field data seem to be more diffused and to include more spurious focus points when compared to lab-based equipment. This finding is supported in a white paper published by the technology provider (iMotions Product Specialist Team, 2022). In contrast, and notwithstanding lower usable data yield, facial expression (FE) data collected online in the field appears to be like facial expression (FE) data collected in the lab.

Our field online data collection experience prompted us to consider the following research question.

RQ1: What is the source of the difference between online and lab-based equipment studies for eye tracking and facial expression analysis.

Expanding the research question, we proposed the following hypotheses for examination.

H1: When environmental factors are controlled, participant usable data yield will be similar whether collected using lab-based or online equipment.

H2: Participant behavior will be unaffected whether the data is collected using online or lab-based equipment.

H3: Eye tracking (ET) data collected with online customer premise equipment will be less accurate than lab-based equipment.

H4: Facial expression (FE) data will be similar when collected with lab-based or online equipment.

To evaluate the research question, we conducted an experiment that compared online and lab-based biometric data acquisition systems while isolating the vagaries of the field environment. Online and lab-based equipment were employed in a lab setting to collect ET and FE data while participants navigated a well-known national fast casual restaurant menu. Usable data yield was evaluated, and statistical analysis was performed on participant behaviors, sensor data, and collection metrics.

Support was found for all four hypotheses. The facial expression metrics and participant experience metrics comparison analysis between the online and lab-based equipment were not noteworthy. Consistent with our qualitative field observations, we found a statistically significant difference in the number of eye tracking fixations between the online and lab-based equipment. In addition, we also experienced improved participant yield of usable data with both online and lab-based equipment in the study compared to our field experience.

The following sections of this paper describe the methodology, results of the lab study, discussion of the findings, and suggestions for researchers.

Review of Research

As technology has evolved, access to biometric experimental data has increased. Research has expanded from medical facilities with specialized tools and techniques like functional magnetic resonance imaging (fMRI) to multipurpose computers equipped with biometric sensors in dedicated laboratory facilities (Vujičić et al., 2021). Today, state-of-the-art biometric labs include diverse biometric sensors that enable data collection for eye tracking (ET), facial expression (FE), galvanic skin response (GSR), heart rate (ECG), and even electroencephalogram (EEG) (iMotions, 2022a). While these tool enhancements have provided researchers greater access, little has changed to improve study respondent access and the logistical burden of participant travel or transport. Historically, study participants must travel to dedicated facilities to participate in biometric research. The requirement for study participant travel has the potential to skew results by sample bias or by creating a less ecologically familiar study environment (lab compared to field). Travel can also increase the cost of conducting research due to transportation assistance and elongated study times associated with managing logistics.

Online research studies that leverage panels, groups, and the community are commonly called crowdsourced and have become a common tool for researchers (Goodman & Paolacci, 2017). There has been significant research across academic disciplines (Ozcan et al., 2020) examining the use (Sheehan, 2018), optimization (Robinson et al., 2019), options (Peer et al., 2017), and quality (Eyal et al., 2021). We ourselves also noted challenges with crowdsourcing in a publication that found significant disinterest and deceit among crowdsourced samples. The observed effect in our study was large enough to significantly skew the study results (Hammond et al., 2022).

Another area of innovation for researchers is the accessibility and usability of biometric sensors for research. Early biometric research employed observation and manual coding with techniques like the Facial Action Coding System (FACS) (Zhao et al., 2021) and mapping participant gaze points (Valtakari et al., 2021). Clinical tools have also been employed to provide researchers with objective measurements. These clinical tools are generally time or capital expensive, located in specialized facilities, and require specific skill sets for data analysis (Vujičić et al., 2021). Functional magnetic resonance

Protocol

Study participants were assigned to one of two identical testing stations in the university biometric lab and instructed to place a restaurant order for themselves using a well-known fast casual restaurant online web menu application. Participants completed the entire order sequence except for entering a credit card and they did not receive the selected food. Respondents were a mix of students from the university's respondent pool and the public. The students from the study pool received course extra credit for participating in the study. The public respondents received no compensation.

Group 1 accessed the menu web application through a web browser that routed the session through an online data collection platform (iMotions, 2022b). Group 2 accessed the menu through a computer software application installed on a multi-purpose computer (iMotions, 2022a). The collected online data was then downloaded and analyzed with the same tools and processes used to analyze the lab-based collected data. Except for the introductory directions in the online application informing participants how to position themselves for maximum data quality, the participant experience was the same for each group.

In both the online and lab-based equipment configurations, the facial expression data was collected by recording the web camera and then post processing the data. The post processing process for the FE data employed an integrated third-party platform for facial expression analysis (Affectiva, 2017). The ET data collection process was different for study participant Group 1 and Group 2. The online configuration (Group 1) utilized a vendor integrated algorithm to calculate gaze points from the web camera recording while the lab-based configuration employed an external Tobii X2 30Hz Eye Tracker peripheral. By using a 30 Hz eye tracker, the sample rate was similar for Group 1 and Group 2.

The Group 1 online eye tracking algorithm employed duration dispersion to determine if a gaze point was a fixation or saccade. The online algorithm assigned a gaze point as being a fixation if the gaze point fell within 100ms of time and 0.1 degrees of space. All other gaze points in the Group 1 online algorithm were determined to be saccades. In contrast, the Group 2 lab algorithm determined fixations based on a velocity-based filter. In the lab configuration, gaze points with over 30 degrees / second in velocity were determined to be saccades. Gaze points with velocities below the 30 degrees / second velocity threshold were determined to be fixations in the Group 2 lab configuration. (Holmqvist & Andersson, 2017, Salvucci & Goldberg, 2000).

To compare the participant user experience navigating the menu, we calculated T-tests to compare checkout time and the length time each participant used to place the first item in the shopping cart. Next, we calculated T-tests to compare the number of ET gaze points and ET fixations between the groups. The eye tracking data was compared by identifying areas of interest (AOIs) for each of the menu items then rank ordering the AOIs based on dwell time (DT) and on time to first fixation (TTFF). We then compared the number of AOIs that appeared in same quartiles between Group 1 and Group 2. DT and TTFF variables were selected to compare the ET data due to the variables reporting unique values that would enable rank ordering within the groups. In a similar method two variables were selected to compare the FE data.

The data collection software generated multiple FE variables. The base unit of data was Facial Action Units that could be evaluated separately or combined into various composite measures including core emotions, and valance. Each of the variables was reported as a value between 0 and 100. We chose the Attention and Engagement variables due to the generality of engagement and the difference measure provided in attention. The Engagement metric is the aggregation of several Facial Action Units that include Inner Brow Raise, Outer Brow Raise, Brow Furrow, Cheek Raise, Nose Wrinkle, Lip Corner Depressor, Chin Raise, Lip Press, Mouth Open, Lip Suck, and Smile and the Attention variable is derived by head position (Affectiva, 2017). Given that the amount of time allotted to interact with the menu application was not controlled the data was normalized by using the provided time percent metrics. The default threshold value of 50 was used in the post processing of the FE data.

The study data was analyzed using a combination of the statistics package JASP and Excel. Usable data was assessed by reviewing the video recordings and data visualizations available in the analysis software (iMotions, 2022a).

imaging (fMRI), encephalogram (EEG), electrocardiogram (ECG) are exemplars. More recently technology has evolved to enable biometric research on general purpose PCs in offices and computer labs. The computers run standard operating system software with software applications designed to operate with diverse sensors and connect the sensors to the computing platforms using standard computing peripheral standards like USB and Bluetooth (iMotions 2022a). Elements of clinical assets like EEG and ECG are incorporated in the tools along with mechanized data collection and analysis of FE, ET, and galvanic skin response (GSR) data. The new tools have democratized access to biometric research by lowering the price of acquisition and skill needed to generate meaningful output.

Findings

In total 15 individuals participated in the study. Usable eye tracking (ET) data was collected from all 15 participants and usable facial expression (FE) data was collected from 11. FE data from one of the lab-based sessions was removed due to a recording issue. Two additional participants from the lab-based equipment and one from the online group were removed based on low attention scores (3.41, 47.75, 0.01).

Eye Tracking and User Experience

On average, the lab-based equipment respondents spent 188 seconds (2min and 28s) from the start of the navigation to checkout, while the online equipment respondents spent an average of 203 seconds (3min and 23s). T-test results show no statistically significant difference between the two groups for either checkout time ($p = 0.714 > 0.05$) or the time to put the first item into the checkout cart ($p = 0.993 > 0.05$) suggesting the user behavior was similar in the two groups. Figures 1 and 2 provide details on the group user experience with the menu.

Figure 3 and Figure 4 summarize the group statistics and T-test results of the eye tracking (ET) gaze point and fixation metrics. On average, the lab-based configuration respondents recorded 1421 gaze points and online data collection group participants recorded 1384 gaze points. The difference in gaze points was not statistically significant. Interestingly, there was a statistically significant 48% difference in the number of fixations recorded between the lab-based and online configurations ($p = 0.032$). On average 230 fixations were recorded for the lab-based equipment and 119 for the online configuration.

Next, we examined the eye tracking data. To compare the online and lab-based configurations we gen-

Lab-Based		
Mean	Std Deviation	N
187.57	72.21	7
106.98	55.83	7

Figure 1: Group statistics of time metrics

		Lavene's Test for Equality of Variances		T-test for Equality of Means						
		F	Sig	t	df	Sig (2 tailed)	Mean Difference	Std Error Difference	Lower	Upper
Checkout Time	Equal var assumed	0.205	0.658	0.66	13.00	0.71	14.98	40.00	-71.44	101.39
	Equal var not assumed			0.37	12.06	0.72	14.98	40.39	-72.98	102.93
First Item In Cart	Equal var assumed	0.138	0.716	-0.01	13.00	0.99	-0.26	30.65	-66.47	65.95
	Equal var not assumed			-0.01	12.16	0.99	-0.26	-67.51	-67.51	66.98

Figure 2: T-test results of time metrics

	Lab-Based			Online-Based		
	N	Mean	Std Deviation	N	Mean	Std Deviation
Gaze Count	8	1421.00	698.74	7	1383.71	753.01
Fixations Count	8	230.38	101.50	7	119.00	73.38

Figure 3: Group statistics of ET metrics

		Lavene's Test for Equality of Variances		T-test for Equality of Means						
		F	Sig	t	df	Sig (2 tailed)	Mean Difference	Std Error Difference	Lower	Upper
Gaze Points	Equal var assumed	0.021	0.887	0.10	13.00	0.92	37.29	374.86	-772.55	847.11
	Equal var not assumed			0.10	12.49	0.92	37.29	376.87	-780.86	855.43
Fixations	Equal var assumed	0.526	0.481	2.40	13.00	0.03	111.38	46.28	11.17	211.58
	Equal var not assumed			2.46	12.61	0.03	111.38	45.35	13.09	209.66

Figure 4: T-test results of ET metrics

	ET Areas of Interest			
	Quartile 1	Quartile 2	Quartile 3	Quartile 4
TTFF Lab-Based	24	24	24	25
TTFF Online-Based	22	21	21	23
TTFF AOI in common	14	7	5	5
DT Lab-Based	25	24	24	25
DT Online-Based	21	20	21	21
DT AOI in common	4	5	4	6

Figure 5: Comparison of Time-To-First-Fixation (TTFF) and Dwell Time (DT) Areas of Interest (AOI)

	Lab-Based				Online-Based			
		Mean	Std Deviation	Range	N	Mean	Std Deviation	Range
Engagement	5	28.93	37.59	90	6	7.04	6.58	15.61
Attention	5	98.10	1.77	3.73	6	98.64	1.32	3.49

Figure 6: FE data comparison

erated Areas of Interest (AOIs) for each of the menu items and rank ordered the AOIs based on dwell time (DT) and on time to first fixation (TTFF). We then compared the number of AOIs that appeared in same quartiles from the two configurations. As reported in Figure 5, the maximum number of repeated AOIs in a quartile was 14 in the first quartile of TTFF and the least was in the third quartile of DT.

Facial Expression

Three (3) of the lab-based equipment group FE data were removed from the analysis. One participant encountered a web camera recording problem and two other lab-based equipment participants had low attention values. In addition, one participant from the online equipment group's data was removed due to low attention. There was significant variation in the remaining lab-based participant's Engagement. Participant Engagement in the lab-based equipment group ranged from 1.35 to 91.35. Variation in Engagement was less in the six (6) participants from the online group, ranging from 0.12 to 15.73. Attention was considerably more consistent in the two groups and between the groups with values ranging from 96.27 to 100 in the lab-based group and between 96.07 and 99.36 in the online group. In recognition of the small sample size, observed significant variation in engagement data, uniformity in the attention variables, and common method for FE data, we forwent T-tests and instead report descriptive statistics for the group FE variables in Figure 6.

Discussion

As hypothesized, there was little difference in the user experience metrics (time to first item in cart or shopping time), the facial expression (FE) Attention, the number of eye tracking (ET) gaze points, or the yield of usable data between the lab-based and online configurations for data collection. In contrast, we found significant variation in the respondent group's FE Engagement variable and ET AOIs duplicated by quartile. Consistent with our field experiences we also observed a significant difference (48%) in the number of eye tracking fixations recorded in the two configurations.

Review of the individual navigation paths of the participants revealed no common ET path within the groups. Consistent with this observation, the differences in facial expression (FE) variable Engagement and the eye tracking (ET) area of interest (AOI) variables Dwell Time (DT) and Time-To-First-Fixation (TTFF) are likely explained by individual variations or perhaps the result of the relatively small sample size of the two groups.

The differences in ET fixations are consistent with our field study experience. We believe the difference in ET fixations is the result of variation in the frame-to-frame web camera recording calculated gaze points. This belief is supported by the technology provider and is potentially explained in the different methods employed by the eye tracking algorithms to determine fixations and saccades. The lab configuration eye tracking algorithm is based on velocity to determine saccades and the online configuration algorithm used duration dispersion to identify fixations. The technology provider also shared online algorithm improvements were becoming available. While new online algorithm enhancements will likely improve performance, it is reasonable to expect the active ET sensor will provide a more consistent reading of eye position and thus have the potential of providing more fixations. Moreover, the gap in fixations (accuracy) between the two groups (online and lab-based) may also expand over time as we employed a modest 30Hz eye tracking sensor in our lab configuration. Much higher rate active trackers are now common which would produce even more data samples and thus potentially exacerbate the fixation count gap between web camera recording (online) and active sensor (lab-based).

It is also clear from the study that the vagaries of the field environment have a significant impact on usable data yield. In our lab study, the lab-based and online configurations delivered almost uniformly usable data for ET. This is substantially different from our field experience. In a field study very similar to the study described here, we collected ET data for participants navigating an earlier version of the same menu. In that field study, we received 56% usable data from the participants that completed the study. Moreover, 64% of the participants that initiated the field study abandoned the study before completing (Hammond et al., 2023).

FE data presents a two-sided story. On one side, the attention metric provides added information to assess participant performance while at the same time eliminating the respondent reduces yield. In this small study and despite controlling the environment, 37.5% of the lab-based group participant's FE data were removed and 14% of the online group.

Category	Researcher Consideration
Recruitment	Plan to recruit significantly more participants for an online study. At least four times (4X) the number of participants as for a lab study and may require considerably. Ensure potential participants are aware their face will be recorded.
Compensation	Mitigate fraud exposure by managing access to the survey link. Do not post link to open forum. Negotiate terms with vendors to pay on usable data. Include statements in the consent that all data will be reviewed, and payment will only be made after researchers have determined the submission is usable.
Effort and Expertise	Plan on significant effort trouble shooting issues and cleaning data. Consider augmenting research team with IT expertise.
Maximize Yield	Internalization appears to influence participant yield. Consider samples that are affiliated with the study even if smaller populations. Employ a moderator or leverage a remote computer lab to mitigate environmental effects. Leverage crowdsourcing literature to help mitigate fraud, insufficient and careless effort.
Eye Tracking Data	Online and lab eye tracking are different tools. Crispness of data visualizations is diminished with online. Plan for additional effort interpreting findings. Gaze point may provide added value when interpreting online data.

Figure 7: Researcher considerations for field biometric online data collection

Conclusions, Limitations, and Future Study

This research suggests online-based and lab-based biometric studies are both viable and valuable but are also different tools with different benefits and challenges. The flexibility and reach afforded online by leveraging a web camera for data collection is offset by reduced eye tracking accuracy and usable data yield reduction. Further, bad actors, vagaries of the field environment, dynamic and diverse consumer IT equipment and networks, along with advances in sensors ensures the gap between lab and field data collection will likely remain. That said, online tools provide researchers with a viable option to collect biometric data in a more "ecologically valid" location and when it is logistically or cost prohibitive to source a sample or operate a lab environment.

The vagaries of the field present significant hazards for researchers and may be more pronounced if participants are compensated (Hammond et al., 2022). Figure 7 notes suggestions based on this study and our experience with field online data collection to optimize field biometric online data collection.

While this research produced insights that assist researchers, the study focused on one vendor's technology and commercially available tools. Given the rapid change in computing it is certain technology will continue to evolve both in evaluating online data and in lab equipment. Additional research comparing these evolving approaches will assist researchers in selecting the best tool for their research.

Another area of limitation in this study was the small sample size and sample of convenience. Although unlikely, based on the observed statistics, it is possible a larger or different sample could produce different results.

References

Affectiva. (2017). *Emotion AI 101: All about emotion detection and Affectiva's emotion metrics. Emotion AI 101: All About Emotion Detection and Affectiva's Emotion Metrics.* Retrieved January 17 from https://blog.affectiva.com/emotion-ai-101-all-about-emotion-detection-and-affectivas-emotion-metrics

Eyal, P., David, R., Andrew, G., Zak, E., & Ekaterina, D. (2021). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods*, 1-20.

Goodman, J. K., & Paolacci, G. (2017). Crowdsourcing consumer research. *Journal of Consumer Research, 44*(1), 196-210.

Hammond, R. W., Parvanta, C., & Zemen, R. (2022). Caught in the Act: Detecting Respondent Deceit and Disinterest in On-Line Surveys. A Case Study Using Facial Expression Analysis. *Social Marketing Quarterly, 28*(1), 57-77.

Hammond, R.W., Norton, C., Wang, Y. (2023). Menu Navigation Served 3 Ways: An exploration of consumer eye navigation of digital restaurant menus using eye tracking. Collection of Robert Hammond, Tampa, Florida.

Holmqvist, K., Andersson, R. (2017) *Eye tracking: A comprehensive guide to methods and measures*, Lund, Sweden: Lund Eye-Tracking Research Institute, 219-234.

iMotions. (2022a). *iMotions Lab.* In (Version 9.3)

iMotions. (2022b). *iMotions Online*. https://imotions.com/products/imotions-online/

iMotions Product Specialist Team. (2022). iMotions Webcam Based Eye Tracking - Whitepaper (v2). Retrieved March 15, 2023, from https://go.imotions.com/WebcamETWhitepaper_April22

Ozcan, S., Boye, D., Arsenyan, J., & Trott, P. (2020). A scientometric exploration of crowdsourcing: research clusters and applications. *IEEE Transactions on Engineering Management*.

Parvanta, C., Hammond, R., He, W., Zemen, R., Boddupalli, S., Walker, K., Chen, H., & Harner, R. (2022). Face Value: Remote facial expression analysis adds predictive power to perceived effectiveness for selecting anti-tobacco PSAs. *Journal of Health Communication, 27*(5), 281-291.

Pedersen, N. (2021). *Novel engineering based on a specific client need - a collaboration story*. Imotions. Retrieved January 9, 2022, from https://imotions.com/blog/novel-engineering-based-on-client-need/

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology, 70*, 153-163.

Robinson, J., Rosenzweig, C., Moss, A. J., & Litman, L. (2019). Tapped out or barely tapped? Recommendations for how to harness the vast and largely unused potential of the Mechanical Turk participant pool. *PloS one, 14*(12), e0226394.

Salvucci, D., & Goldberg, J. (2000). Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 symposium on Eye tracking research & applications*, 71-78.

Sheehan, K. (2018). Crowdsourcing research: data collection with Amazon's Mechanical Turk. *Communication Monographs, 85*(1), 140-156.

Valtakari, N. V., Hooge, I. T., Viktorsson, C., Nyström, P., Falck-Ytter, T., & Hessels, R. S. (2021). Eye tracking in human interaction: Possibilities and limitations. *Behavior Research Methods, 53*(4), 1592-1608.

Vujičić, M., Grbac, B., & Mozetič, V. (2021). Contributions of Functional Magnetic Resonance Imaging in Market Research: Obstacles and Perspectives. *Market-Tržište, 33*(2), 185-203.

Zhao, X., Zhu, J., Luo, B., & Gao, Y. (2021). Survey on facial expression recognition: History, applications, and challenges. *IEEE MultiMedia, 28*(4), 38-44.

Review

This article was accepted under the ***constructive peer review*** option. For futher details, see the descriptions at:

http://mumabusinessreview.org/peer-review-options/

Authors

Robert Hammond is the Director of the Center for Marketing and Sales Innovation at the University of South Florida. In addition to his role as the center's director, he oversees the sales program and teaches strategy in the doctorate program. Dr. Hammond is also the managing partner of RH Insights LLC, a business strategy and personal development firm, a corporate board member, guest lecturer, member of the National Academy of Inventors and the Institute of Electrical and Electronics Engineers. He holds a Doctor of Business Administration, Master of Business Administration, Master of Science in Engineering, Bachelor of Science in Electrical Engineering, and five patents.

Yuqi Wang is originally from China. She earned her master's degree in marketing from the University of South Florida. As a graduate student, Yuqi had the exciting opportunity to work as a Graduate Research Assistant in the CMSI Customer Experience Lab, conducting eye-tracking research projects. She is now a Research and Development Analyst at a business consulting company, Revenue Management Solutions. Yuqi's current role specializes in menu eye-tracking research and data analytics, delving into the fascinating world of consumer behavior.