WILEY | Hindawi

Research Article

How Differing Audiences Were Associated with User Emotional Expression on a Well-Being App

Maya Topitzer [ID],[1] **Yueming Kou** [ID],[1] **Robert Kasumba** [ID],[2] **and Philip Kreniske** [ID][3]

[1]*Columbia University Mailman School of Public Health, USA*
[2]*International Center for Child Health and Development, Washington University in St. Louis, USA*
[3]*HIV Center for Clinical and Behavioral Studies, New York State Psychiatric Institute and Columbia University, USA*

Correspondence should be addressed to Philip Kreniske; pk2361@cumc.columbia.edu

In the last five years, there has been an explosion of mobile apps that aim to impact emotional well-being, yet limited research has examined the ways that users interact, and specifically write to develop a therapeutic alliance within these apps. Writing is a developmental practice in which a narrator transforms amorphous thoughts and emotions into expressions, and according to narrative theory, the linguistic characteristics of writing can be understood as a physical manifestation of a narrator's affect. Informed by literacy theorists who have argued convincingly that narrators address different audiences in different ways, we used IBM Watson's Natural Language Processing software (IBM Watson NLP) to examine how users expression of emotion on a well-being app differed depending on the audience. Our findings demonstrate that audience was strongly associated with the way users expressed emotions in writing. When writing to an explicit audience users wrote longer narratives, with less sadness, less anger, less disgust, less fear, and more joy, these findings have direct relevance for researchers and well-being app design.

1. Introduction

In the last five years, there has been an explosion of apps designed to support emotional well-being, yet limited research has examined the ways that users interact, and specifically write to develop a therapeutic alliance, within these apps [1, 2]. Informed by literacy and psychological research that has illuminated the ways that writers address different audiences in different ways [3–9], the current analysis examined how users expressed emotion in a well-being app when they were writing to an explicit as compared to an implied audience.

In 2014 alone, there were over 100,000 apps available to provide lay-people with health information [10]. By 2017, this number increased to approximately 325,000 [11]. The mobile app of study was designed to enhance mental well-being and like many mental health apps relied on a chatbot to deliver therapeutic content. One way the chatbot interacted was to send messages that encouraged users to share their thoughts and feelings in writing. Chatbot's have been

defined as digital tools "that use machine learning and artificial intelligence methods to mimic humanlike behaviours and provide a task-oriented framework with evolving dialogue able to participate in conversation" [12]. Self-disclosure to a chatbot may have many of the same positive psychological outcomes as self-disclosure to a person [13], and people appear to interact with bots in much the same way that they interact with other humans [9].

Most studies of commercially available mental health apps focus on whether the app increases well-being or decreases mental health symptoms [14, 15]. However, researchers have yet to identify the emotional and cognitive processes by which mental health apps may influence well-being. According to narrative and psychological theory, the linguistic characteristics of a narrator's writing are a physical manifestation of their affect [16–19]. Thus, writing is a developmental practice in which a narrator transforms amorphous thoughts and emotions into expressions. In fact, a recent study showed that the expression of high rates of negative affect in writing was associated with higher

neuroticism and depression and lower overall health, and conversely, positive affect was associated with higher extraversion, agreeableness, overall health, and lower self-reported neuroticism and depression [20].

However, narrative expression is not fixed, but relational and narrators express thoughts and emotions differently when addressing distinct audiences [16, 21]. Researchers have found that when a specific and active audience is addressed, it potentially influences writers' compositions more than when writers are directed to write for a general audience that offers no feedback (Black, 1989; Cohen & Riel, 1989; Freedman, 1994; Purcell-Gates, Duke & Martineau, 2007). Researchers have found that when a specific and active audience is addressed, it potentially influences writers' compositions more than when writers are directed to write for a general audience that offers no feedback [5, 22–24]. In one such study, first year college students used more cognitive (e.g., think, know, and realize) and intensifying expressions when randomly assigned to a blogging activity where peers could interact and comment on the blog posts as compared to a control group of students who wrote to an imagined peer [7]. The current study is aimed at identifying how users' expression of emotion on a well-being app differed depending on the audience.

2. Method

2.1. Participants and Procedure.
The app of study, LiveBetter, is a mobile application designed to promote emotional well-being. It is free and publicly available for Apple and Android devices. An update was pushed to all app users with the option to opt in or out of research. Data was pulled for users who did not opt out. This initial data was deidentified (email addresses and birthdates were removed) and pulled from the server on July 16, 2019. The data was then cleaned to only include users who had completed at least one in-app activity and who had written at least three words in response to this activity. This dataset was provided to the authors for analysis. Columbia University's Institutional Review Board determined this was *Not Human Subjects Research*. The final sample included 1269 app users from five continents who wrote at least three words on the mobile application between 2017 and 2019. All written content and chatbot interactions were in English.

2.2. Presence and Strength of Emotion.
Over the course of studying human emotion, philosophers and psychologists have proposed between three and eleven primary emotions [25]. All these proposals include fear, anger, and sadness and most include joy as well. IBM Watson is the software selected to code for the presence and strength of emotions in this study identifies sadness, fear, anger, joy, and disgust. Disgust was a useful addition as it can be understood as an intersection of contempt and remorse, on the pathway between loathing and boredom. Therefore, fear, anger, sadness, joy, and disgust were the five emotions examined in this study.

2.3. IBM Watson and Natural Language Understanding.
IBM Watson has been used to analyze sentiment and emotion in a wide range of writing contexts and has been found to have high reliability, validity, and efficiency [26, 27]. The dataset was narrowed to include only activities with user generated text content. These activities were eventually ordered within each user by timestamp to include each participant's first written interaction with the mobile app. This text was then analyzed using IBM Watson Natural Language Understanding (NLU). Natural Language Processing (NLP) refers to the automatic process of extracting information from spoken or written text. Natural Language Understanding was used in this context to determine the strength of emotion (sadness, fear, anger, joy, and disgust) present in the participant response to the activity.

One example of this process would be the example user input, "I think that there are flaws with my body and how others perceive me though I know that's my own thought projection." This input resulted in the following output for emotion strength: sadness: 0.418, joy: 0.236, fear: 0.306, disgust: 0.09, and anger: 0.088. Each emotion was coded for strength on a scale of 0-1 with 1.0 being the strongest possible emotion. To complete analysis, Watson NLU requires at least three words per response which it then automatically compares to a preexisting library of language that has been coded for presence and strength of emotion.

The dataset was then imported into an SQL database which was used to organize user message length by character count. We then coded the activities within the app based on intended audience and identified two types of audience, *writing to an implied audience* and *writing to an explicit audience*. One example of an activity prompt that would be coded as "implied audience" was "What feeling did you get from your happy memory?" An example of an activity prompt that was coded as "explicit audience" was "What would you tell a friend if they had your worry?" Potential "explicit audiences" included in writing prompts could be friends, partners, family members, or work colleagues.

2.4. Statistical Methods.
Prevalence of sociodemographic characteristics, device type (Apple or Android), length of writing, and the strength of emotion were evaluated using basic descriptive statistics. For the categorical variables, we showed the frequency table that has counts and percentages. For the continuous variables, we presented the following: number of observations, mean, standard deviation, max, min, the first quartile (Q1), and the third quartile (Q3).

We used chi-square tests to explore if the sociodemographic characteristics were dependent on the type of audience, either writing to an implied or explicit audience. We used a series of Wilcoxon-Mann–Whitney tests to examine if writing length and strength of emotion differed by explicit or implied audience.

Six separate multivariate linear regression models were built to find how writing length and emotion scores are related to audience adjusted by gender and year. Since neither length of writing nor emotion scores are normally

TABLE 1: User and activity characteristics.

Characteristics of participants		Observations (%)					
Number of unique persons		1269 (100)					
Gender	Female	776 (61.15)					
	Male	209 (16.47)					
	Unknown	284 (22.38)					
Year	2017	63 (4.96)					
	2018	611 (48.15)					
	2019	595 (46.89)					
Device	Android	312 (24.63)					
	iPhone	955 (75.37)					
Time zone	Africa	22 (1.74)					
	America	870 (68.67)					
	Asia	90 (7.10)					
	Australia	75 (5.92)					
	Europe	203 (16.02)					
	Pacific	7 (0.55)					
Activity type	Writing to explicit audience	94 (7.41)					
	Writing to implied audience	1175 (92.59)					
	N	Mean	Std	Max	Min	Q1	Q3
Length	1269	18.04	22.30	335.00	3.00	6.00	21.00
Sadness	1227	0.25	0.23	0.97	0.00	0.07	0.40
Anger	1227	0.11	0.14	0.95	0.00	0.03	0.14
Fear	1227	0.15	0.19	1.00	0.00	0.03	0.18
Disgust	1227	0.08	0.12	0.93	0.00	0.01	0.09
Joy	1227	0.42	0.32	0.99	0.00	0.12	0.71

All analyses include only users first completed activity from users who wrote at least 3 words, wrote in English.

distributed, we took the natural logarithm for these outcome variables.

Shapiro-Wilk test was used to test normality and found sufficient evidence that writing length and strength of emotions for each audience type were not normally distributed. Median values and the 25th and 75th percentiles of each dependent variable were presented by audience. A series of Mann–Whitney U tests were performed to examine if writing length and strength of emotions differed by audience. Hodges-Lehmann estimate and 95% confidence interval of the difference between the median values of the two audience types were presented. Effect sizes were calculated by dividing the absolute standardized test statistic z by the square root of the number of pairs: $r = |z|(n)^{-1/2}$.

3. Results

Most participants used Apple devices (75.37%) and were female (61.15%). Most were located in North America (68.67%), followed by Europe (16.02%), Asia (7.1%), Australia (5.92%), and Africa (1.74%) as noted in Table 1. Geography was determined by time zone which was captured from app user metadata. In terms of audience, most of the writing coded as "writing to implied audience" (92.59%).

Out of the 1269 app users, the average length of writing was 18.04 words, the average sadness score was 0.25, the average anger score was 0.11, the average fear score was 0.15, the average disgust score was 0.08, and the average joy score was 0.42. There was no significant dependency between sociodemographic characteristics and audience type, apart from time (Table 2). Writing length and emotions were significantly associated with audience type (Table 3). Fitting linear regression models, we found when writing to an explicit audience as compared to writing to an implied audience, users wrote longer narratives (CI 1.32-1.99, $p < 0.0001$) and wrote with less sadness (CI 0.33-0.58, $p < 0.0001$), less fear (CI 0.23, 0.43, $p < 0.0001$), less anger (CI 0.34-0.62, $p < 0.0001$), more joy (CI 1.83-3.53, $p < 0.0001$), and less disgust (CI 0.24-0.45, $p < 0.0001$) (Table 4).

The two-sided p values of the Mann–Whitney U tests were all less than 0.0001 which indicated the writing length and emotions were significantly differed by audience type (Table 3). The effect sizes were all less than 0.3 which would be considered as small effects according to Cohen's classification of effect sizes. Hodges-Lehmann estimates indicated that the median difference (95% CI) of length between writing to other and writing to self was 6.00 (3.00, 9.00), the median difference (95% CI) of sadness was -0.08 (-0.12,

Table 2: First activity descriptive of categorical variables by audience.

		To implied audience	To explicit audience	p value
Number of people		1175	94	
Gender	Female	719 (61.19%)	57 (60.64%)	
	Male	192 (16.34%)	17 (18.09%)	0.8956
	Unknown	264 (22.47%)	20 (21.28%)	
Android/iPhone	Android	294 (25.06%)	18 (19.15%)	0.2003
	iPhone	879 (74.94%)	76 (80.85%)	
Year	2017	59 (5.02%)	4 (4.26%)	
	2018	550 (46.81%)	61 (64.89%)	0.0031
	2019	566 (48.17%)	29 (30.85%)	
Time zone	Africa	20 (1.71%)	2 (2.13%)	
	America	808 (68.88%)	62 (65.96%)	
	Asia	83 (7.08%)	7 (7.45%)	
	Australia	70 (5.97%)	5 (5.32%)	0.9251
	Europe	185 (15.77%)	18 (19.15%)	
	Pacific	7 (0.60%)	0 (0%)	

Table 3: First activity descriptive of continuous variables by audience.

	To self		To other		p value	Median difference (95% CI)	Effect size
	Median	Q1, Q3	Median	Q1, Q3			
Length	11.00	6.00, 20.00	19.00	9.00, 36.00	<.0001	6.00 (3.00, 9.00)	0.131
Sadness	0.19	0.08, 0.41	0.07	0.03, 0.21	<.0001	-0.08 (-0.12, -0.05)	0.154
Anger	0.07	0.03, 0.14	0.03	0.01, 0.08	<.0001	-0.03 (-0.04, -0.01)	0.132
Fear	0.09	0.03, 0.19	0.03	0.01, 0.07	<.0001	-0.04 (-0.06, -0.03)	0.182
Disgust	0.04	0.02, 0.09	0.02	0.01, 0.04	<.0001	-0.02 (-0.03, -0.01)	0.159
Joy	0.35	0.11, 0.68	0.76	0.59, 0.87	<.0001	0.31 (0.23, 0.40)	0.231

Table 4: Parameter estimates of linear regression models.

	Covariate	β	95% CI of β	p value
Length	Writing to other person	1.619	(1.317, 1.989)	<.0001
	Year	0.946	(0.863, 1.038)	0.2420
	Female	1.123	(0.984, 1.282)	0.0859
Sadness	Writing to other person	0.430	(0.323, 0.573)	<.0001
	Year	1.087	(0.955, 1.238)	0.2045
	Female	1.188	(0.985, 1.433)	0.0709
Anger	Writing to other person	0.452	(0.336, 0.610)	<.0001
	Year	1.130	(0.987, 1.293)	0.0759
	Female	1.149	(0.945, 1.396)	0.1627
Fear	Writing to other person	0.308	(0.223, 0.427)	<.0001
	Year	1.146	(0.989, 1.327)	0.0694
	Female	1.206	(0.975, 1.492)	0.0835
Disgust	Writing to other person	0.326	(0.237, 0.449)	<.0001
	Year	1.107	(0.958, 1.278)	0.1684
	Female	1.016	(0.825, 1.252)	0.8783
Joy	Writing to other person	2.540	(1.829, 3.528)	<.0001
	Year	0.797	(0.687, 0.924)	0.0027
	Female	0.913	(0.737, 1.131)	0.4034

-0.05), the median difference (95% CI) of anger was -0.03 (-0.04, -0.01), the median difference (95% CI) of fear was -0.04 (-0.06, -0.03), the median difference (95% CI) of disgust was -0.02 (-0.03, -0.01), and the median difference (95% CI) of joy was 0.31 (0.23, 0.40).

There was no significant dependency between sociodemographic characteristics and audience type, apart from time (Table 2).

Fitting linear regression models, we found when writing to an explicit audience as compared to writing to an implied audience, users wrote longer narratives (CI 1.32-1.99, $p <$ 0.0001) and wrote with less sadness (CI 0.33-0.58, $p <$ 0.0001), less anger (CI 0.34-0.62, $p < 0.0001$), more joy (CI 1.83-3.53, $p < 0.0001$), less disgust (CI 0.24-0.45, $p < 0.0001$), and less fear (CI 0.23, 0.43, $p < 0.0001$) (Table 4).

4. Discussion

Our findings demonstrate that audience was strongly associated with the way users expressed emotions when writing in a well-being app. When writing to an explicit audience users wrote longer narratives, with less sadness, less anger, less disgust, less fear, and more joy, this has direct relevance for researchers and well-being app designers as eliciting greater positive emotion may contribute to improved well-being and physical health.

The current work is the first to our knowledge to demonstrate how the audience impacts a narrator's expression of affect in the context of a well-being app. However, more generally, the importance of audience has been long documented in narrative and psychological research [5, 7, 8, 21]. Further, narrative and psychological theory posit that affective expression is in fact a physical representation of emotional experience [16–19]. Recent empirical evidence supports this theoretical position and demonstrates that affective expression in writing corresponds with emotional experiences such that increased use of positive emotions was correlated with higher well-being and better physical health [20]. Therefore, creators of chatbot-based well-being interventions may consider prompting users to write to an explicit audience to elicit more content and stronger positive emotion.

Future research must continue to examine how writing in well-being apps impacts psychological well-being overtime. For instance, might writing to explicit others and thus using more positive emotions (e.g., joy) increase well-being overtime, and could writing about sadness, fear and disgust be linked with decreased well-being? Conversely, overtime might addressing and expressing negative emotions through writing help people work through psychosocial challenges and thus experience increased well-being in the long run.

4.1. Limitations. One limitation of the current study was that many more people wrote to an implied audience as opposed to writing to an explicit audience. Therefore, it is possible that there may be a lack of generalizability. Another limitation was that we only analyzed the first writing activity, and it is possible that emotions would differ overtime by audience. Future research should examine the differences in writing based on audience and how writing on this platform may change overtime.

Another limitation of the current study was that most of the users were in North American or European time zones and using iPhone, and thus, the generalizability of our findings may be limited. Three major benefits of mobile appbased interventions are their scalability, affordability, and potential to overcome stigma [28, 29]. Further studies with broader global populations and more detailed demographics including age, race, and socioeconomic measures could properly assess the generalizability of these findings and the appropriateness of this intervention across cultures and genders.

5. Conclusion

As well-being apps become increasingly popular [11], it is necessary to ensure that app developers consider how specific writing prompts within the app elicit distinct psychological processes. The stark differences depending on audience identified in our analysis suggest that audience impacts the ways that narrator's express emotion. Therefore, researchers and app designers must carefully consider the impact of audience when designing well-being apps. Accordingly, developers who aim to foster writing with more positive affect should consider using explicit as opposed to implicit audiences in their writing prompts. App-based well-being interventions should give special consideration to the wording of questions and specifically to the use of explicit or implied audience of written prompts.

Data Availability

We can provide deidentified data upon request.

Disclosure

Any opinions, findings, and conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Institute of Health.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] J. Kim, S. J. Park, and L. P. Robert, "Conversational Agents for Health and Wellbeing: Review and Future Agendas presented at the Identifying Challenges and Opportunities in Human AI Collaboration," in *Healthcare at the 22th ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW 2019)*, Austin, Texas, 2019.

[2] G. Neff and P. Nagy, "Automation, algorithms, and politics| talking to bots: symbiotic agency and the case of Tay," *International Journal of Communication*, vol. 10, article 17, 2016.

[3] M. M. Bakhtin, "The Bildungsroman and its significance in the history of realism," *Speech Genres and Other Late Essays*, vol. 10, article 21, 1986.

[4] C. Bazerman, *Intertextualities: Volosinov, Bakhtin, literacy theory, and literacy studies*, A. F. Ball and S. W. Freedman, Eds., Cambridge University Press, Cambridge, United Kingdom, 2004.

[5] M. Cohen and M. Riel, "The effect of distant audiences on students' writing," *American Educational Research Journal*, vol. 26, no. 2, pp. 143–159, 1989.

[6] L. Ede, "Writing as a social process: a theoretical foundation for writing centers?," *The Writing Center Journal*, vol. 9, no. 2, pp. 3–13, 1989.

[7] P. Kreniske, "How First-year Students Expressed Their Transition to College Experiences Differently Depending on the Affordances of Two Writing Contexts," *Computers and Composition*, vol. 45, no. 2, pp. 1–20, 2017.

[8] C. Daiute, R. Todorova, and T. Kovacs-Cerovic, "Narrating participation and power relations in a social inclusion program," *Language & Communication*, vol. 45, pp. 46–58, 2015.

[9] M. Duradoni, S. Collodi, S. C. Perfumi, and A. Guazzini, "Reviewing stranger on the internet: the role of identifiability through "reputation" in online decision making," *Future Internet*, vol. 13, no. 5, article 110, 2021.

[10] D. Lupton and A. Jutel, "It's like having a physician in your pocket!A critical analysis of self-diagnosis smartphone apps," *Social Science & Medicine*, vol. 133, pp. 128–135, 2015.

[11] R. S. Larson, "A path to better-quality mHealth apps," *JMIR mHealth and uHealth*, vol. 6, no. 7, article e10414, 2018.

[12] A. N. Vaidyam, H. Wisniewski, J. D. Halamka, M. S. Kashavan, and J. B. Torous, "Chatbots and conversational agents in mental health: a review of the psychiatric landscape," *The Canadian Journal of Psychiatry*, vol. 64, no. 7, pp. 456–464, 2019.

[13] A. Ho, J. Hancock, and A. S. Miner, "Psychological, relational, and emotional effects of self-disclosure after conversations with a chatbot," *Journal of Communication*, vol. 68, no. 4, pp. 712–733, 2018.

[14] M. Eisenstadt, S. Liverpool, E. Infanti, R. M. Ciuvat, and C. Carlsson, "Mobile apps that promote emotion regulation, positive mental health, and well-being in the general population: systematic review and meta-analysis," *JMIR Mental Health*, vol. 8, no. 11, article e31170, 2021.

[15] Y. Lin, C. Tudor-Sfetea, S. Siddiqui, Y. Sherwani, M. Ahmed, and A. B. Eisingerich, "Effective behavioral changes through a digital mHealth app: exploring the impact of hedonic well-being, psychological empowerment and inspiration," *JMIR mHealth and uHealth*, vol. 6, no. 6, article e10024, 2018.

[16] C. Daiute and P. B. Gómez, "Narrating Change in and against Time in Colombia," in *Representations of War, Migration, and Refugeehood*, pp. 154–176, Routledge, London, United Kingdom, 2014.

[17] D. E. Polkinghorne, *Narrative Knowing and the Human Sciences*, State University of New York Press, Albany, NY, 1988.

[18] Merleau-Ponty, *Phenomenology of Perception (Donald A. Landes, Trans.)*, Routledge, New York, NY, 1945/2014.

[19] J. Wertsch, *Voices of the Mind*, Harvester Wheatsheaf, London, 1991.

[20] V. Vine, R. L. Boyd, and J. W. Pennebaker, "Natural emotion vocabularies as windows on distress and well-being," *Nature Communications*, vol. 11, no. 1, pp. 1–9, 2020.

[21] L. Lucić, "Use of evaluative devices by youth for sense-making of culturally diverse interpersonal interactions," *International Journal of Intercultural Relations*, 2013.

[22] K. Black, "Audience analysis and persuasive writing at the college level," *Research in the Teaching of English*, vol. 23, no. 3, pp. 231–253, 1989.

[23] S. W. Freedman, *Exchanging writing, exchanging cultures: Lessons in school reform from the United States and Great Britain*, Harvard University Press, Cambridge, Ma, 1994.

[24] V. Purcell-Gates, N. K. Duke, and J. A. Martineau, "Learning to read and write genre-specific text: Roles of authentic experience and explicit teaching," *Reading Research Quarterly*, vol. 42, no. 1, pp. 8–45, 2007.

[25] R. Plutchik, "The nature of emotions," *American Scientist*, vol. 89, no. 4, pp. 344–350, 2001.

[26] L. W. Lee, A. Dabirian, I. P. McCarthy, and J. Kietzmann, "Making sense of text: artificial intelligence-enabled content analysis," *European Journal of Marketing.*, vol. 54, no. 3, pp. 615–644, 2020.

[27] S. Provoost, J. Ruwaard, W. V. Breda, H. Riper, and T. Bosse, "Validating automated sentiment analysis of online cognitive behavioral therapy patient texts: an exploratory study," *Frontiers in Psychology*, vol. 10, article 1065, 2019.

[28] M. Berger, T. H. Wagner, and L. C. Baker, "Internet use and stigmatized illness," *Social Science & Medicine*, vol. 61, no. 8, pp. 1821–1827, 2005.

[29] D. M. Hilty, S. Chan, T. Hwang, A. Wong, and A. M. Bauer, "Advances in mobile mental health: opportunities and implications for the spectrum of e-mental health services," *Focus*, vol. 16, no. 3, pp. 314–327, 2017.