Scribbling intervention for depression, anxiety and stress

Author One
Institute One
Address One
State, Country
author@one.com

Author Two
Institute Two
Address Two
State, Country
author@two.com

ABSTRACT

Depression, anxiety and stress are disorders with significant impact on quality of life and preventable death. The pervasiveness of these problems and their diversity requires that interventions be matched to the personality of the subject and to their context. There is a need for tools that have broad effects and are easy to use, accessible, and affordable. In this study we evaluate a scribbling intervention motivated by theories in art therapy. Many art therapy based interventions are not evidence based as quantitative studies are rarely performed. We use a crowdsourcing platform to recruit subjects and provide evidence for the efficacy of an art based intervention. Our results show that a short time spent scribbling can have a significant effect reducing depression, anxiety and stress, at least in the short term. It could be used as a soothing technique for people who suffer from these disorders.

CCS Concepts

•**Applied computing** → **Psychology;** •**Information systems** → *Crowdsourcing;*

Keywords

Art Therapy; Depression; Anxiety; Stress

1. INTRODUCTION

Depression, anxiety, and stress are common disorders which are responsible for poor quality of life and preventable death. It is estimated that 18.1% of adults in the US experience an episode of anxiety every year and 6.7% of the adult population experience severe depression every year [8]. At the same time, 75% of adults in the US report at least one symptom of stress every year [1]. In 2004, the World Health Organization (WHO) listed unipolar depression as the 3^{rd} most important cause of disease burden in the world [11]. There is a need for tools to assist people suffering from these disorders. The magnitude of this problem dictates that solutions must be scalable, affordable and accessible to be able to make a significant impact.

Many art therapy interventions are accessible, affordable, and cross cultural [4]. Scribbling [12], is an art therapy intervention which is supported by neurological theories. However, there is very little quantitative research to support the effectiveness of art therapy interventions [4]. One possible explanations for the lack of such quantitative studies is the perceived high cost and complexity of conducting them [4].

In this study we show that a simple and short computerized scribble intervention has a positive impact on the experience of depression, anxiety, and stress. The self administered intervention consists of two minutes of free scribbling followed by 30 seconds of reflecting on the image, writing a short title and a description of the drawing. Therefore, it can be performed by almost anyone, anywhere. In order to study the efficiency of this intervention we created an online scribbling tool. We used Amazon's mechanical Turk (mTurk),[1] a crowd-sourcing platform, to recruit participants for the study. This technique allowed us to recruit 284 participants in under 24 hours. We used the Depression, Anxiety, and Stress Scales (DASS) [10] to measure subjects before and after the intervention and found significant improvement in all scales: 0.8 points in reduction for depression, 0.9 points reduction for anxiety, 2.1 points reduction for stress and 3.8 points reduction overall. In all these scales subjects experienced on average a reduction of $10\% - 20\%$ in their reported symptoms. Moreover, most of the subjects ($> 70\%$) who reported some change in their experience, reported a positive change (p-value $\ll 0.01$ using the exact binomial test).

We tested two variants of our intervention: subjects were assigned at random to use either their dominant or their non-dominant hand to perform the scribbling task. Both variants generated significant improvements in their depression, anxiety and stress experiences. The results indicate that using the non-dominent hand generates stronger improvement compared to using the dominent hand. However, the differences between the hands measured in this study are not large enough to be statistically significant.

This study has two main contributions. First, we show that the computerized scribbling intervention has a significant effect on depression, anxiety, and stress levels in a sample of US adults. The effect is significant both in magnitude and in statistical terms. The second contribution is in the methods we developed to conduct this study which can be used to test other art therapy interventions.

Pervasive Health '16 May 16–19, 2016, Cancun, Mexico

DOI: 10.475/123_4

[1] mturk.com

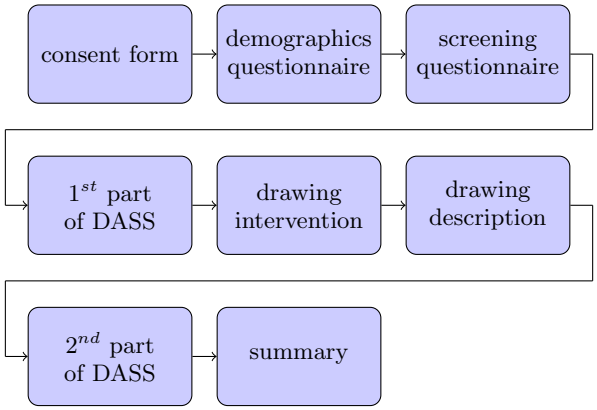

Figure 1: A flow chart of the study procedure. In Table 2 we show the number of participants that completed each stage.

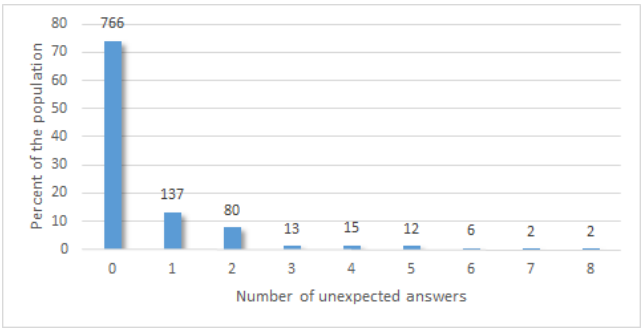

Figure 2: A histogram of the number of unexpected answers to the short screening questionnaire.

2. METHOD

The study was conducted during December 2015. Figure 1 shows a schematic view of the data collection process. All the participants in this study signed an informed consent form and the study was approved by the institutional review board of our institute.

2.1 Participants

We have used Amazon's Mechanical Turk (mTurk) to recruit subjects for this study. The subjects were required to be at least 18 years old, US residents, with at least 100 HITs (Human Intelligent Tasks) completed with an acceptance rate of at least 90%. Furthermore, we asked the subjects to use a large screen touch device to complete the task. Subjects who completed all the stages of the study were given a code to redeem $2.50 in their Amazon account.

The short questionnaire consisting of seven questions about gender, year born, first spoken language, education, marital status, employment and handedness was used to identify the demographics profile of the participants. A summary of the data is provided in Table 1.

According to our records, 1170 subjects signed the consent form but only 284 subjects completed the entire task. We conjecture that these were not necessarily unique subjects because subjects may first try to perform the task on a device that is not touch enabled and switch to a touch enabled device once they observed that they could not complete the task on a non-touch device. The number of participants completing each step of the task dropped steadily (see funnel data in Table 2). The screening form was designed to screen out bots and low performing users. It consists of eight yes/no questions and the score is the number of answers which differs from the answers given by the majority of the population. We accepted in to the study subjects with a score of zero or one. As seen in Figure˜2, 87% of the subjects (903 out 1033) were accepted. Since for only two out of the eight question the expected answer was "yes", a score of two was achievable by answering "no" on all questions. Indeed, 49 out of the 80 subjects who had a score of two answered "no" on all questions.

The biggest drop in the number of subjects happened during the drawing task. Since this is the only part of the process for which a touch screen is mandatory, we conjecture that many tried to complete the task without a touch screen and so were not able to complete this phase. This is supported by a few complaints that we got from participants.

Completing the task took 780 seconds on average (median 614) for the subject who completed the entire task . The time begins the moment they signed the consent and continued until they had been notified that the experiment had ended and had gotten a code to receive their payment through Amazon. Given that each subject received $2.50, the per hour rate was $11.54 on average (median $14.66). Therefore, the pay rate is higher than the minimum hourly wage in the US ($7.25).

The final datasets consisted of 284 subjects (157 females), ages 18 to 66. Table 1 shows the demographic characteristics of the participants in this study. The sample is diverse in many aspects, including gender, education, marital status and age. However, there are some biases in this sample: The participants are more educated than the general population in the US. According to the United States Census Bureau in 2014, 19% of the population had a bachelor's degree and additional 19% had some college credit while 30% were high school graduate. Even when restricting to 25 years and older, only 20% have bachelor's degree compared to $\sim 40\%$ among the participants in this study. .

2.2 Materials

Several tools were used in this study to collect information about the participants.

2.2.1 Filtering tool

Some workers on mTurk try to use bots to perform tasks in order to earn the money while others try to perform tasks as fast as possible which yields low quality results [17]. Therefore, it is common to use some filtering tools to identify these workers and reject their work [15]. One method that has been used in psychological studies is the Infrequency-Psychopathology Scale [2, 16] to identify these workers. This scale is composed of 27 yes/no questions that the vast majority of the population answers in the same way. In order to reduce the load on the subjects, we used data from a previous study to select eight questions that effectively help identify bots:

- "Sometimes when I am not feeling well I am irritable"
- "I get angry sometimes"
- "Someone has been trying to poison me"

Table 1: Demographics: the following table shows the main demographic characteristics of the participants in the study. The "All participants" column shows the numbers for all the particiapnts who completed the demographics questionnaire while the column "Participant who scribbled" shows the numbers for the participants who completed all the stages of the study.

	All participants	Participants who scribbled
Number of participants	1050	284
Gender: Female / Male	493 / 557	157 / 127
Handedness: Left / Right	118 / 932	37 / 247
Education: college credit / Associate / Bachelor's / Master's	239 /116 / 421 / 117	73/ 33 / 117 / 23
Martial status: Single / Married / Other	509 / 461 / 80	136 / 129 / 159
Age: min / max / avg / median	18 / 76 / 33 / 31	19 / 66 / 33 / 31

Table 2: Number of subjects completing successfully each phase of the task

Phase	# subject completing
Consent	1170
Demographics	1050
Screening	903
First part of DASS	880
Drawing Task	300
Second DASS	284
Entire Task	284

- "Someone has been trying to rob me"
- "Everything tastes the same"
- "Sometimes I enjoy hurting persons I love"
- "Someone has control over my mind"
- "I hate my whole family"

Subjects were expected to answer "yes" on the first two questions and "no" on the rest. Users whose answeres disagreed with more than one of these expected questions were eliminated from the study.

2.2.2 *Depression, Anxiety, and Stress Scales (DASS)*

The DASS questionnaire [10] is made of 42 questions that ask about the experience of the subject during the last week. Each question is answered by a 4 level scale. The questions are divided into two sets of 21 questions such that in each set has 7 questions about depression, 7 questions about anxiety and 7 questions about stress. We used one set of these questions before the intervention and another set after the intervention so that we were able to measure changes. To prevent biases, we assigned at random the set of 21 questions to be completed before and after the intervention. Furthermore, to make the reported scores compatible with the scores of the DASS [10] and similar scales (for example, DASS-21 [6]), the scores were multiplied by 2 to compensate for the fact that only 21 questions out of the 42 questions were used pre and post intervention.

The DASS scale, even when administered online was found to have a good test-retest reliability [3, 18]. Therefore, we expect changes between pre-intervention and post-intervention measurment to reflect well the effect of the intervention.

2.2.3 *Drawing tool*

The drawing tool (see Figure 3) was implemented using Java-Script and allowed participants to draw using their fingers on the touch screen of the computer. The tool contains only simple features: selecting one color at a time among 13 available colors (red, orange, blue, turquoise, green, dark-green, hot-pink, magenta, purple, brown, white, yellow, and black) and clearing the screen.

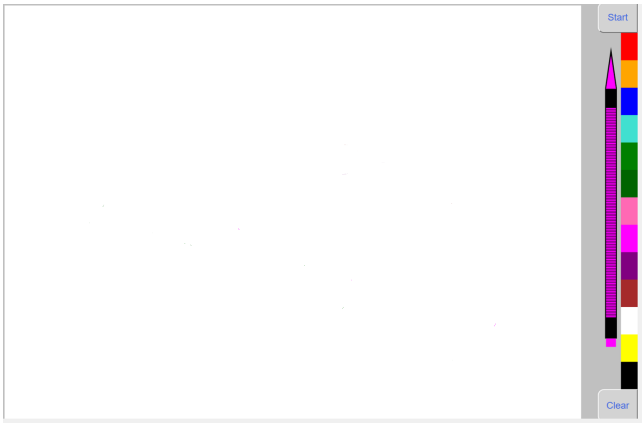

Figure 3: Screenshot of the drawing tool used in the study.

2.3 Procedure

Subjects who elected to participated in our study were asked to visit a web-site that was built for this study and hosted on Microsoft Azure services. A flow chart of the process the subjects completed is provided in Figure 1. First, the subjects were asked to review the consent form and accept its terms. Next, subjects were asked to complete the short demographics questionnaire described in Section **??**. Upon completing the demographic questionnaire subjects were asked to complete the short screening questionnaire described in Section 2.2.1. Subjects that answered at least 7 out of 8 questions of the screening questions correctly were asked to complete the first set of 21 DASS questions (see Section 2.2.2). At this point the scribbling intervention was administered as described in Section 2.2.3. Users were asked to try out the tool to learn how to use it and press the start button once ready. Upon pressing the start button, subjects were asked to scribble for 2 minutes. Each subject was assigned, at random, to draw with their dominant or non-dominant hand. They were presented with the following instructions: "Use your dominant/non-dominant hand to scribble during the next 2 minutes. Your drawing does not have to have any particular meaning and you can use any combination of colors and shapes. Note however, that if

you stop for more than 5 seconds, the timer on the top-right corner will reset."[2]

After 2 minutes, the subjects were presented with the following instructions: "Please use the following 30 seconds to look at the image from different points of view. If you find objects or emotions in it, you can use this time to add more details.". Next, they were presented with a new screen in which they were asked to give a title to their drawing and provide a more detailed description: "What did you find in the scribble (types of emotions, people, shapes, ...)? please write a paragraph about it".

Once the intervention was done, the subjects were asked to complete the second half of the DASS questions. Next, subjects were presented with a summary page in which they were given a code to redeem $2.5 for completing the task. Users were also given the option to leave a comment in this page.

All the data was transferred to Azure servers over an encrypted channel (using https). The data was stored on password protected Azure tables and blobs. To preserve the privacy of the users, we assigned each user a random globally unique identifier (Guid). This identifier was not linked to their Amazon worker ID which was not collected in our systems.

2.4 Statistical Analysis

We used a few statistical methods to analyze the results.

2.4.1 Exact Binomial Test

We use the exact binomial test when the null hypothesis is that the intervention has no effect or maybe a negative effect. In this case we have W success in N trails and the null hypothesis states that the probability of success is $p \leq 0.5$. If $W \leq {}^{N}/_{2}$ we accept the null hypothesis, otherwise, the probability of the observation under the null hypothesis is $2^{-N} \sum_{k=W}^{N} \binom{N}{k}$ which is the tail of the binomial distribution.

We used this test to compare, for example, pre-intervention and post-intervention results. In this case we count as "success" every subject that scored lower in the post-intervention scale relative to the pre-intervention measurment. The number of trials (N) was cosidered as the number of subjects who had some difference in the pre-intervention and post-intervention scale.

2.4.2 Fischer's Exact Test

We use Fischer's exact test [5, 14] when comparing two arms of the experiment. Assume that arm 1 was pulled N_1 times and W_1 successes were observed, similarly, arm 2 was pulled N_2 times and W_2 successes were observed. The null hypothesis is that arm 2 is equal or worse than arm 1. Therefore, if ${}^{W_1}/_{N_1} \geq {}^{W_2}/_{N_2}$ we accept the null hypothesis. Otherwise, the difference in the rate of successes is due to the random assignment of subjects to receive the treatment of arm 1 or arm 2. Hence, the probability of seeing a big difference in the rates is

$$\sum_{k=0}^{W_1} \frac{\binom{N_1}{k}\binom{N_2}{W_1+W_2-k}}{\binom{N_1+N_2}{W_1+W_2}}$$

[2]Resetting the timer upon inactivity is the method we used to make sure that the subject was using the 2 minutes to scribble as opposed to just sitting for 2 minutes.

Table 3: DASS average (standard-deviation) reports before and after the scribble intervention

Scale	Before	After	avg-change
Depression	8.12 (9.2)	7.30 (9.4)	-0.82 (3.5)
Anxiety	5.42 (6.4)	4.53 (6.4)	-0.90 (4.5)
Stress	10.90 (8.1)	8.79 (8.2)	-2.11 (5.2)
Total	24.46 (20.5)	20.62 (21.2)	-3.83 (9.4)

Table 4: The rate of improvement vs. deterioration in each scale. For the sample here we excluded the subjects who reported 0 in the relevant pre-intervention scale and subjects who reported no change. The p-Value is computed using exact binomial test.

	# improvements / N	p-Value
Depression	${}^{100}/_{144} = 0.70$	1.7×10^{-6}
Anxiety	${}^{105}/_{148} = 0.71$	1.8×10^{-7}
Stress	${}^{151}/_{201} = 0.75$	2.8×10^{-13}
Total	${}^{161}/_{225} = 0.72$	4×10^{-11}

3. RESULTS

284 subjects completed all the stage of the study. In Section 3.1 we analyze the changes in DASS before and after the interventions. In Section 3.2 we present a sentiment analysis of the free text comments and descriptions left by the participants.

3.1 Changes in DASS

Each subject was asked to complete one part of the DASS questionnaire before the scribble intervention one the other part after the intervention. The order of the two parts was selected at random to compensate for potential biases. We observed improvement in the reported levels in every scale as presented in Table 3. To put these numbers in context, consider the results of Pengel et al. [13] that showed an intervention that included exercise sessions and advice for patients suffering from low back pain did not improve the DASS depression scale by more than 0.7 points.

Figure 5 shows a scatter plot of the DASS scales before and after the intervention. It can be seen that more subjects show improvement, and this is also evident from Table 4. This table shows the percentage of subjects who experienced improvement as a fraction of all the subjects that experienced a change in the any of the scales. In all scales, 70-75% of the subjects who experienced a change had a positive change. These results are very significant (p-Value $< 1.7 \times 10^{-6}$, exact binomial test).

The regression lines in Figure 5 suggest that there might be a trend for greater improvement for subjects with more sever conditions. To verify that we extract the subjects who had the higher 33% of the scores in each scales and computed the rate of improvement and the associated p-Value. While it still holds that more subjects experienced improvement compared to deterioration, the reduced sample size does not allow us to show statistical significance in cases of depression or stress, but there is statistical significance in the improvement in anxiety (p-Value = 0.032, exact binomial test).

The intervention had two variants. Subjects were assigned at random to scribble using either their dominant hand or non-dominant hand. Our main hypothesis was that the

Table 5: The impact of the hand used on the effect of the intervention. The p-value was computed using Fischer's exact test.

	improvement rate		p-Value
	dominant hand	non-dominant hand	
Depression	$^{46}/_{71} \cong 0.65$	$^{54}/_{81} \cong 0.67$	0.47
Anxiety	$^{44}/_{78} \cong 0.56$	$^{61}/_{87} \cong 0.70$	0.048
Stress	$^{73}/_{105} \cong 0.70$	$^{78}/_{100} = 0.78$	0.11
Total	$^{74}/_{107} \cong 0.69$	$^{87}/_{118} \cong 0.74$	0.27

scribble intervention will improve the experience of depression, anxiety, and stress. Moreover, we hypothesized that using the non-dominant hand for drawing will have a larger effect than using the dominant hand and indeed this is the case as presented in Table 5. In all scales, the subjects who used their non-dominant hand were more likely to experience improvement than subjects who were using their dominant hand. When using Fischer's exact test to compute the statistical significance of this result we find the difference not to be statistically significant (the p-value for anxiety is 0.048 which is only trending to significance because Boneferroni correction requires that the critical p-Value for this experiment be 0.0125). We note that we have no way to verify that all subjects who were instructed to use their non-dominant hand were following the instructions. Therefore, it is possible that the difference is larger than reported here. It is interesting to note that the results in Table 5 are consistent with the regression line in Figure 5 which indicate that the largest impact is on anxiety followed by stress while the impact on depression is smaller.

3.2 Sentiment analysis

In two places, the subjects were asked to answer questions using free form text: once when describing their scribble and the second time when providing an optional comment at the end of the task. These bodies of text provide an opportunity to measure the sentiment of the subjects. To do so, we have measured the use of positive and negative terms in the texts to see if emotions were elicited and what kinds of emotions they were. We have used lexicons of negative and positive words constructed by Hu and Liu [7] to count the number of negative and positive words used. Lexicon based sentiment analysis has many limitations (see Section 1.2.2 in [9] for more details). For example, one of the subjects described their painting in the following way "I found irritation and anger. I made squiggles, lines, and dots of every color. It was fun.". This description contains two negative words (irritation, anger) and one positive word (fun) and therefore the lexicon based analysis will conclude that it has a negative sentiment, but one may argue that the main emotional reaction of the subject is positive. Another subject used the following description "I found that this is a fond memory that I had of my childhood. When I had no worries or responsibilities unjust played outside with my friends. I miss the freedom of being young.". In this case the phrase "no worries" contains the negative word "worries" while the phrase "no worries" has a positive sentiment that the lexicon based method misses.

However, since more sophisticated sentiment analyzers are domain specific (see Section 3.4 in [9]), the lexicon based method provides us with a good proxy. Therefore, in our analysis we have compared each text to the positive and negative lexicons and counted the number of unique words from each of the lexicons were used. The goal was to find whether sentiments were solicited, what types of sentiments there were and how were the parameters different when people were using their dominant or non-dominant hand to draw.

3.2.1 Scribble description

As a part of the scribble intervention, subjects were asked to give a title and a description of their drawing. We have combined the title and the description into a single bag of words and compare it to the lexicon. Table 6 shows a summary of the results of this experiment. Subjects used, on average, 2 positive words in their description and only one negative word. Moreover, 70% of the subjects used more positive words than negative words (p-value 9.2×10^{-8}, exact binomial test). When looking at the subjects who used their dominant hand, we notice a slight increase in the number of negative words used, as well as increase in the overall use of sentiment words, either positive or negative. Nonetheless, regardless of the hand used, there was a clear preference for the positive sentiment compared to the negative one.

The top positive vocabulary used in describing the scribble were "happy" (57 subjects), "like" (46 subjects), "happiness" (23 subjects), "love" (19 subjects), and "fun" (17 subjects). The top negative vocabulary used in the descriptions were: "chaos" (12 subjects), "confusion" (8 subjects), "crazy" (6 subjects), "anger" (6 subjects), "sad" (6 subjects), and "chaotic" (6 subjects)

3.2.2 Comments

After completing the entire task, the subjects had the option of leaving a comment. As seen in Table 6, about a half of the subjects (118) left comments. Of these, in 83 showed a difference in the number of positive and negative references. In 79 out the 83, the number of positive references was greater than the number of negative ones (p-value=2×10^{-19}, exact binomial test). Most comments referred to the task. As examples: "this was a great HIT, really fun and interesting.", "Thanks for letting me be creative", or "I felt a little sorted after the scribbling. Thank you for the opportunity.". Some subjects had technical difficulties with the drawing tool: "I had some issues getting the page oriented correctly on my iPad, it was not easy to see the whole thing. Maybe my iPad had too low of a resolution. Thanks for the survey though!".

The drawing tool itself was very limited which was frustrating to some: "I'm an artist and I appreciate art therapy, but this particular means felt pretty crappy. The rudimentariness of the tool was unsatisfying - the fact that there was no way e.g. to change brushes/stroke thickness, or to blend colors. I felt the one type of stroke that was available was neither relaxing (as a thicker/softer one might be, reminis-

Table 6: Sentiment in scribble description. This table shows the usage of positive and negative words in the description of the drawing and the comments left by the users. The win rate is the ratio of the subjects who used more positive terms (subjects that used the same number of positive and negative terms were excluded). The p-Value applies the exact binomial test.

	Drawing Description				Comment			
	Average # words		win rate	p-Value	Average # words		win rate	p-Value
	positive	negative			positive	negative		
All participants	1.95	1.01	$^{134}/_{195} = 0.69$	$\ll 0.01$	0.528	0.09	$^{79}/_{83} = 0.95$	$\ll 0.01$
Dominant hand	1.95	0.94	$^{65}/_{94} = 0.70$	$\ll 0.01$	0.52	0.12	$^{37}/_{40} = 0.925$	$\ll 0.01$
Non-Dominant	1.95	1.08	$^{69}/_{101} = 0.68$	$\ll 0.01$	0.54	0.06	$^{42}/_{43} = 0.98$	$\ll 0.01$

cent of painting) nor expressive (as a thinner and/or sharper one might be). It just looked like stringy crap." while others liked the drawing tool "Love this exercise and the UI of the art palette was really easy to use. Love it."

The top positive vocabulary used in the comments were: "thanks" (27 subjects), "fun" (25 subjects), "interesting" (14 subjects), "enjoyed" (8 subjects) and "good" (5 subjects). In the negative vocabulary used, the word "sorry" was used by 2 subjects and 20 additional words were used only once by any subject.

4. DISCUSSION

Our study shows that a simple scribbling exercise has a significant effect on levels of depression, anxiety and stress as measured by the DASS scales. Majority of the subject experienced improvement after the intervention (p-Value $\ll$ 0.01, exact binomial test). We note however that these are short term effect and can be considered as soothing. The DASS questionnaire asks about the experience in their last week, and the short scribbling intervention did not change the experience the subjects had but maybe rather their perception of their experience. It would be interesting to measure the long term effect of using this intervention as a part of treatment. However, even if the effect is only short term, this intervention has many merits. It is affordable, accessible, crosses cultural barriers and has good results. While we measured the intervention using computers and web-sites, this intervention could be administered using paper and colors which extends its reach even further.

Overall, the participants' reaction to drawing as a soothing mechanism was ver positive. Most subjects who left comments after the exercise used more positive language than negative (95%, p-value$\ll$ 0.01). The negative comments were mostly about technical aspects of the online drawing tool we used. The drawing tool was minimalistic with only 13 different colors to choose from and the drawing tool did not work as expected on some devices. These issues should be fixed for future studies.

We find these results very encouraging. However many questions remain. For example, we would like to verify that the positive feedback is not just because subjects assumed that was what we are trying to achieve and tried to "make us happy". It would also be interesting to study the long term effect of this intervention and its effects on other cognitive, mental and behavioral metrics. Moreover, the sample that was used in this study, while diverse, is more educated and contains more female than the general population. Therefore, it remains to be found whether this intervention works well for the entire population or for a specific subset of the population.

5. REFERENCES

[1] N. B. Anderson, C. D. Belar, S. J. Breckler, K. C. Nordal, D. W. Ballard, L. F. Bufka, L. Bossolo, S. Bethune, A. Brownawell, and K. Wiggins, editors. *Stress in America - Paying with our Health*. American Psychological Association, 2015.

[2] P. A. Arbisi and Y. S. Ben-Porath. An mmpi-2 infrequent response scale for use with psychopathological populations: The infrequency-psychopathology scale, f (p). *Psychological Assessment*, 7(4):424, 1995.

[3] T. A. Brown, B. F. Chorpita, W. Korotitsch, and D. H. Barlow. Psychometric properties of the depression anxiety stress scales (dass) in clinical samples. *Behaviour research and therapy*, 35(1):79–89, 1997.

[4] D. Edwards. *Art therapy*. Sage, 2014.

[5] R. A. Fisher. The logic of inductive inference. *Journal of the Royal Statistical Society*, pages 39–82, 1935.

[6] J. D. Henry and J. R. Crawford. The short-form version of the depression anxiety stress scales (dass-21): Construct validity and normative data in a large non-clinical sample. *British Journal of Clinical Psychology*, 44(2):227–239, 2005.

[7] M. Hu and B. Liu. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 168–177. ACM, 2004.

[8] R. C. Kessler, W. T. Chiu, O. Demler, and E. E. Walters. Prevalence, severity, and comorbidity of 12-month dsm-iv disorders in the national comorbidity survey replication. *Archives of general psychiatry*, 62(6):617–627, 2005.

[9] B. Liu. Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1):1–167, 2012.

[10] S. Lovibond and P. F. Lovibond. *Manual for the depression anxiety stress scales*. Psychology Foundation of Australia, 1996.

[11] C. Mathers, D. M. Fat, and J. T. Boerma. *The global burden of disease: 2004 update*. World Health Organization, 2008.

[12] C. M. McNamee. Using both sides of the brain: Experiences that integrate art and talk therapy through scribble drawings. *Art Therapy*, 21(3):136–142, 2004.

[13] L. H. Pengel, K. M. Refshauge, C. G. Maher, M. K. Nicholas, R. D. Herbert, and P. McNair. Physiotherapist-directed exercise, advice, or both for

Figure 4: Examples drawings. Some drawings are abstract while other have a theme. Several the drawings contain text.

subacute low back pain: a randomized trial. *Annals of Internal Medicine*, 146(11):787–796, 2007.

[14] M. Raymond and F. Rousset. An exact test for population differentiation. *Evolution*, pages 1280–1283, 1995.

[15] N. B. Shah, D. Zhou, and Y. Peres. Approval voting and incentives in crowdsourcing. *arXiv preprint arXiv:1502.05696*, 2015.

[16] D. N. Shapiro, J. Chandler, and P. A. Mueller. Using mechanical turk to study clinical populations. *Clinical Psychological Science*, page 2167702612469015, 2013.

[17] P. Welinder and P. Perona. Online crowdsourcing: rating annotators and obtaining cost-effective labels. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, pages 25–32. IEEE, 2010.

[18] K. R. Zlomke. Psychometric properties of internet administered versions of penn state worry questionnaire (pswq) and depression, anxiety, and stress scale (dass). *Computers in Human Behavior*, 25(4):841–843, 2009.

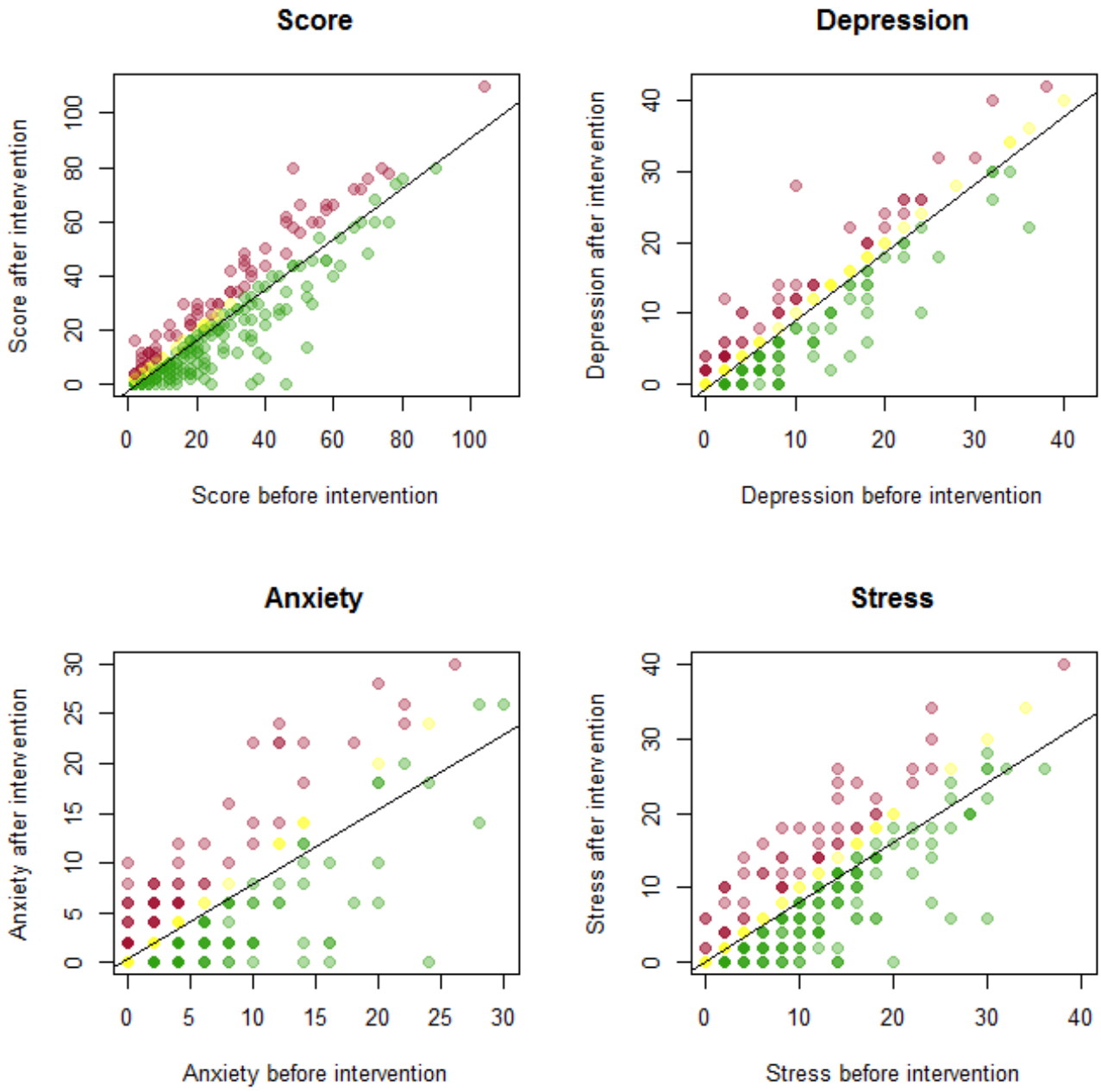

Figure 5: The effect of the intervention: the X axis shows the scale before the intervention and the Y axis shows the scale after the intervention. Green points present a subject for which there was an improvement, red points represent subjects for whom there was deterioration and yellow point show subjects for which there was no effect. The black line is a linear regression line showing the trend.