Investigating Word Retrieval Bias During Verbal Semantic Fluency Tasks in Adolescents At-Risk for Mental Health Disorders

Brian Stasak[1,2] and Mark Larsen[2,3]

[1]School of Elect. Eng. & Telecomm., University of New South Wales, Australia, [2]Black Dog Institute, University of New South Wales, Australia, [3]Centre for Big Data Research in Health, University of New South Wales, Australia

b.stasak@unsw.edu.au, mark.larsen@unsw.edu.au

Abstract

This study investigates mental health (e.g., anxiety, depression, distress, suicidal ideation) and word-retrieval biases exhibited by 122 adolescents in 188 verbal categorical semantic fluency audio recordings. Traditional metrics show that adolescents with suicidal ideation produced up to twice as many unrelated keywords than those without. Experimental keyword metrics (e.g., affective ratings, dietary nutrients, RGB colour, sport attributes) were also evaluated. Results show that *high* severity groups produce food category keywords that are low in fibre and vitamin C; and *high* severity groups for anxiety and suicidal ideation generate food category keywords with significantly higher caloric, fat, cholesterol, and protein content.

Index Terms: anxiety, depression, distress, speech, suicidality

1. Introduction

Studies [1, 2] have shown that mental illness is pervasive in school aged populations. This is concerning because early episodes of mental illness often result in life-long wellbeing struggles [3]. According to global studies [4, 5], half of mental health illnesses found in adulthood begin by age 14 and most of these cases go undetected and/or untreated. While early screening in teenagers with mental illness improves clinical outcomes, there is still a desire to better understand the complex heterogenous symptoms associated with mental illness and its effects on verbal behaviour in this vulnerable population.

While hundreds of subjective self-survey mental health severity screening protocols exists for mental illnesses like depression (e.g., BDI-II, PHQ-9, HAMD) [6], experimental studies [7–9] have explored objective speech-based methods to further help identify mental illness symptoms. For instance, during isolated verbal tasks, speech-based mental health studies (e.g., bipolar, suicidal behaviour) [7–10] have reported abnormal neurocognitive and physiological speech-language behaviours (e.g., disfluencies, flat prosody, poor voice quality).

While abnormal emotional bias in adults with clinical depression has been previously studied, highlighting a propensity for increased negative valence fixation [11], risk word usage [12], and personal pronouns [13], there has been little investigation into other indices that could reveal verbal word retrieval bias in individuals with mood disorders. One isolated method that could be useful in quickly ascertaining bias is the semantic fluency task, which is a verbal test that gauges how well a person can quickly name items in specific category. Each utterance provides a degree of biased idiolect and can help to reveal mood-related disturbances [12, 13].

Unlike previous verbal semantic fluency mental illness studies [9, 14] which have focused on calculating traditional semantic fluency metrics, such as the number correct, the study herein also investigates category-specific keyword indices, such as affective human ratings, food nutrients, RGB colours, and sport traits. When compared to healthy controls, conversational speech transcripts studies [14, 15] have found lower valence ranges in adults with mood disorders. Therefore, it may be hypothesized that during the semantic fluency emotion category task, adolescents with higher mental health disorder severity scores will utter keywords with lower valence values (i.e., more negative than positive terms).

Previous diet-related mental health studies [16-20] indicate that adolescents diagnosed with mental health disorders often have unhealthy, inadequate diets when compared to healthy controls. For instance, adolescents with clinical depression or suicidal behaviour were shown to consume less fibre [18], less vitamin C [19], and more fat [20]. While deeper insights into nutrimental contents of food keywords uttered by participants have not yet been investigated using semantic fluency tests, it is likely that participants will verbally project their personal eating habits, reflecting the previously mentioned associations.

Studies [21, 22] that have investigated the relationship between colour and specific emotions/moods, have found that colour brightness is associated with positive-emotional words and colour darkness with negative-emotional words. For example, keywords associated with mood disturbances, such as '*anger*', '*failure*', and '*sadness*', are frequently associated with certain culturally designated darker colours, such as red and blue [22]. It is anticipated that during colour semantic fluency tasks individuals with higher mental health disorder severities may more frequently utter darker coloured keywords.

For sport semantic fluency tasks, considerations into competition attributes could also provide signs to mood disorders. It is hypothesized that adolescents with higher symptom severities may prefer to name sports that are individually played or have a smaller number of players; and that avoid direct physical contact with others. A recent study [23] found that sports team athletes were less likely to have anxiety or depression than solo athletes.

This speech-based mental health study examines 122 adolescents' verbal semantic fluency task audio recordings in natural real-world environments to explore differences in responses from participants in *low* (none-to-minimal) versus *high* (moderate-to-severe) severity groups. Using a statistical analysis per severity group, four semantic fluency categories and new index metrics are examined. Based on recorded speech-transcripts, results test the validity of using deeper lexical indices to help identify youths who are more at risk for anxiety, distress, depression, and suicidal ideation disorders.

2. Data

Experiments in this study used speech recordings collected for the Future Proofing Study in Australia [24]. This data included mental health self-report surveys, demographic metadata, and elicited audio recordings collected in Australia using an interactive smartphone app [24]. This data collection study was approved by the University of New South Wales Human Research Ethics Committee. Participants were Year 8 high school students (i.e., 12 to 14 years old) and voluntary enrolment included individual and parental consents. Four semantic fluency task categories, shown in Table 1, were randomly allocated to participants. Recordings were manually vetted for task compliance because some participants failed to comply or there was the presence of excessive background noise. A subset of the original data, herein called BDI-SF, was used for experimentation. In total, there were 122 unique participants and 188 digital audio recordings.

BDI-SF audio was recorded using a mono-channel 44.1 kHz 16-bit sampling rate. Prior to each recording, participants were given read task instructions via a smartphone app. Each recording contained a semantic fluency task (see Table 1), whereby a participant was asked to verbalize in English as many keywords related to a category in a 30-second period. Some participants chose to stop their recordings earlier than the 30-second period. The average recording length was 22 seconds.

Based on mental health self-report surveys (Spence Children's Anxiety Scale Short-Form, Depression Patient Health Questionnaire for Adolescents, Distress Questionnaire-5, Suicidal Ideation Attributes Scale), severity scores were obtained for most participants. Thresholds were used to determine *low* (none-to-minimal) versus *high* (moderate-to-high) severity groups: anxiety (≥10); depression (≥10); distress (≥10); and suicidal ideation (≥1). These cut offs were used in previously published mental health studies [10, 24, 25] and increased sensitivity and sample size per severity group.

Table 1. *Number of BDI-SF dataset participants per low (none-to-minimum) and high (moderate-to-severe) mental health self-survey severity group. More than half of the participants were in the high severity groups for two or more self-surveys.*

	Anxiety		**Depression**		**Distress**		**Suicidal Ideation**	
Task Category	*Low*	*High*	*Low*	*High*	*Low*	*High*	*Low*	*High*
Colours	30	22	31	21	16	36	35	10
Emotions	28	17	27	18	6	39	27	17
Foods	26	16	29	13	17	25	33	9
Sports	30	19	31	18	16	33	33	12
Average Survey Score	*6.6*	*14.2*	*5.2*	*15.7*	*6.8*	*14.4*	*0.0*	*14.5*
Survey Range	*0 to 24*		*0 to 27*		*0 to 25*		*0 to 50*	

3. Experimental Methods

BDI-SF audio recordings were manually transcribed for validation purposes due to variable background noise factors (i.e., automatic speech recognition could be used, but with likely less word recognition accuracy than a human listener). Per recorded transcript, traditional semantic fluency metrics included: the total recording duration in seconds (TRDS); the unique total correct keywords (TCK) related to a given category; percentage of unrelated words (PUW) which was calculated by subtracting TCK from the total number of words then dividing by total number of words; and percentage of word repeats (PWR) which was calculated by counting all word repeats then dividing it by the total number of words uttered.

For the colours category task, the RGB colour code index [26] was used which contained over 200 colour keywords including colour modifiers (i.e., '*burnt sienna*', '*dark blue*', '*neon yellow*'). The RGB has a standardized scale from 0 to 255 for each colour component: (R)ed, (G)reen, and (B)lue. For each of the three colour components, lighter colours are represented by a higher index value, whereas darker colours are represented by a lower index value. Per colour semantic fluency transcript, the RGB component values for each colour keyword uttered were computed. The semantic fluency colour category task transcripts contained 90 unique colour keywords. Fig. 1 contains BDI-SF dataset RGB (R)ed values and corresponding paired colour keyword transcript responses from a *low* and a *high* depression severity participant. These two participants uttered the same number and many of the same colour keywords. However, the *high* severity participant produced a word repeat for the colour '*green*' and generated an average RGB index (R)ed value of 168, whereas the *low* severity participant had a RGB index (R)ed value of 133.

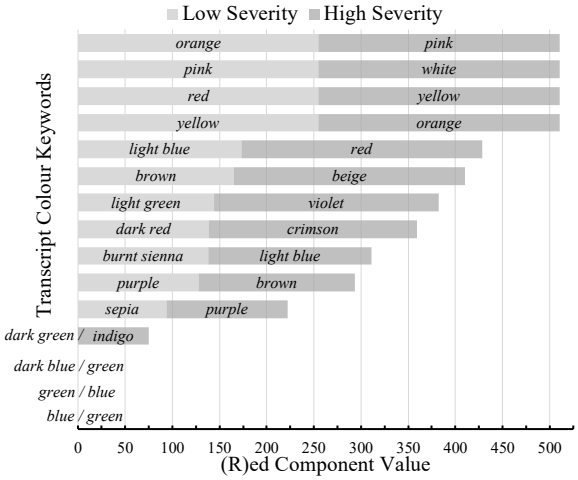

Figure 1: *BDI-SF dataset colour task comparing keyword transcript paired responses from two participants with different depression self-survey severities; (R)ed scale 0-255 per colour.*

For the emotions category task, recording transcripts were evaluated using affective valence keyword human-ratings [27]. The affective valence lookup index included over 20k English words, and each word rating was normalized from 0 to 1 (i.e., a score closer to 1.00 has positive valence, whereas a score closer to 0.00 has more negative valence). In total, semantic fluency emotion category task transcripts contained 147 unique valence-related keywords (e.g., happy, sad, angry, bored).

For the foods category task, a food attribute index was compiled based on an online open-source dietary food guide [28], which contains over 100k food products. Food keywords were based on a serving size of 100 grams and thirteen nutrients were measured: total calories, fat, saturated fat, protein, carbohydrates, sugars, fibre, cholesterol, sodium, iron, potassium, vitamins A and C. In total, participants uttered 208 unique food keywords, including connective terms (e.g., *'bacon and egg sandwich'*, *'spaghetti Bolognese*').

For the sports category task, four proposed metrics were defined in this study. The sport competition metric included player type: solo-based (-1), mixed (0), and team-based (1). The opponent interaction metric was based on physical opponent contact type: indirect (-1), mixed (0); and direct (1). The location type metric was determined by the location where the sport is typically played: indoor (-1), mixed (0), and outdoor (1). For example, rugby is a team-based (1) direct competition (1) in an outdoor (1) setting, whereas fencing is a solo-based (-1) direct competition (1) in an indoor setting (-1). Per sport, the number of on-field teammates was also calculated (i.e., ranging from 1 to 22). In total, participants generated 110 unique sporting activities during the semantic fluency sports task.

For each index measure described above, an automated python script was created to generate participants' task keyword transcript averages. For experiments herein, there was no instance where a semantic fluency transcript contained a correct keyword that was missing from a topic specific index. A composite score involving multiple semantic fluency task indices was inexecutable because participants did not complete all the same tasks (i.e., randomized). Per participant index results, statistical significance between *low* and *high* severity group results were evaluated using two-sample heteroscedastic unequal variance *t*-tests with two reported confidence levels ($p \leq 0.05$, $p \leq 0.10$) [29] and Hedges' effect size ($g \geq 0.40$) [30, 31]. The *t*-test tests whether the difference between response values for two groups is statistically significant or not. The effect size is a quantitative reflection of the magnitude of phenomena concerning how strong a difference there is between two groups. Hedges' effect size provides a better estimate for smaller sample sizes when compared to other methods (e.g., Cohen's *d*, Glass' Δ) [31].

4. Results

Per *low* and *high* severity groups, traditional semantic fluency differences (i.e., low severity minus high severity averages) are shown in Table 2, wherein a positive value indicates that the *low* severity group had a larger average value than the *high* severity group (i.e., a negative value indicates the *high* severity group had a larger value than the *low* severity group). While the total recording duration (TRDS) between the *low* and *high* severity groups were significantly different for many tasks, it should be noted that this measurement does not account for the total duration of speech versus non-speech. For distress and depression, the *high* severity group demonstrated shorter average recording lengths (18.7sec) per emotion category tasks than the *low* severity group (22.8sec). However, the suicidal ideation self-survey emotion category results were an exception because the *high* severity group exhibited longer average recording lengths (23.2sec) when compared to the *low* severity group (20.1sec). Despite the lengthier recordings in the emotions category *high* suicidal ideation severity group, this extra time did not result in a higher percent average of correct words produced (i.e., this duration increase was due to increased unrelated words). The emotions category TRDS results also show that the *low* severity groups for distress, depression, and suicidal ideation had significantly longer recordings (~3 to ~6 seconds) and were less likely to terminate their recordings early when compared to the *high* severity groups. The *high* severity groups possibly chose to end the emotion category recordings early due to emotional avoidance and/or triggers.

The percentage of unrelated words (PUW) was a strong indicator for the *high* severity depression and suicidal ideation groups during the foods task, whereby the *high* severity group results produced significantly greater proportions of unrelated words (40%, 31%) when compared to the low severity groups (20%, 17%). Generally, the PUW was larger for the *high* severity groups. For different disorders, the *low* and *high* severity groups produced overall higher total correct keywords (TCK) averages for the colors (12.8 to 14.6), foods (12.8 to 13.8) and sports (11.9 to 13.0) tasks, whereas the TCK was lower for the emotions (9.04 to 10.24) task. Thus, hinting that some categories are easier and/or more readily familiar. During the foods task, there was a consistent significant increase in the percentage of word repeats (PWR) for the *high* severity groups for anxiety, distress, depression, and suicidal ideation when compared to the *low* severity groups.

Table 2. *Average difference between low and high severity groups for each semantic fluency category using traditional measurements (recording duration in seconds (TRDS), percentage of unrelated words (PUW), total correct keywords (TCK), percentage of word repeats (PWR)). Shaded results indicate significant group differences based on t-tests ($p \leq 0.05$, $p \leq 0.10$) and medium-to-large Hedges' effect size.*

Self-Survey	Category	TRDS	PUW	TCK	PWR
Anxiety	Colours	-0.36	-6%	-1.40	-2%
	Emotions	-0.17	-9%	-1.04	-7%
	Foods	-1.54	-5%	0.32	-5%
	Sports	-1.12	0%	2.30	0%
Distress	Colours	0.43	-1%	-0.45	-2%
	Emotions	3.73	0%	1.55	1%
	Foods	-0.63	-1%	0.15	-4%
	Sports	0.13	-1%	-1.56	1%
Depression	Colours	-0.88	1%	-2.22	1%
	Emotions	3.09	-16%	0.44	-3%
	Foods	-1.50	-10%	1.44	-7%
	Sports	-0.52	-3%	-0.38	4%
Suicidal Ideation	Colours	-0.56	-5%	-5.10	0%
	Emotions	5.94	-6%	1.06	-1%
	Foods	-1.20	-16%	0.33	-9%
	Sports	1.52	-1%	0.19	5%

Per mood disorder self-survey, the colours semantic fluency results indicated that most average RGB individual component index values were similar for *low* and *high* severity groups. For example, individual component ranges recorded included: (R)ed (142 to 157), (G)reen (123 to 133), and (B)lue (105 to 111). There were two exceptions whereby significant (R)ed average differences for the *low* and *high* anxiety/depression severity groups were recorded. Surprisingly, for the anxiety/depression conditions, results indicated that the *high* severity groups produced (R)ed colour keywords that were brighter in shade (157, 157) than the *low* severity groups (146, 123). Further analysis revealed that the increased (R)ed value found in the *high* severity anxiety/depression groups was attributed to a greater number of red-hued colors uttered per recording (i.e., an average of one more red keyword than *low* severity group). Previous research [21] indicated that the colour red is most associated with negative emotion, which may explain this phenomenon exhibited by the *high* severity groups.

Table 3. *Food semantic fluency category results per low (none-to-minimum) and high (moderate-to-severe) groups. Per food keyword, dietary contents were measured using grams, except for sodium, calcium, and potassium which used milligrams. Shaded results indicate significant group differences based on t-tests ($p \leq 0.05$, $p \leq 0.10$) and effect size.*

		Average													
Self-Survey	**Severity**	**Cals.**	**Fat**	**S. Fat**	**Prot.**	**Carb.**	**Sugar**	**Fibre**	**Chol.**	**Sodi.**	**Calc.**	**Iron**	**Pota.**	**Vit. A**	**Vit. C**
Anxiety	***Low***	186.1	8.2	2.9	7.0	21.9	7.0	2.1	28.6	224.7	47.2	1.4	252.4	47.5	8.5
	High	233.8	11.8	4.8	9.6	23.7	8.4	1.4	39.2	279.3	47.1	1.1	256.8	47.7	7.8
Distress	***Low***	199.0	9.3	3.5	7.7	22.0	7.7	2.4	32.0	248.4	52.3	1.4	251.3	45.6	9.2
	High	192.2	8.6	3.2	7.3	22.4	7.0	1.6	29.7	226.2	43.8	1.2	256.0	49.0	7.6
Depression	***Low***	194.9	8.8	3.2	7.6	22.0	6.5	2.1	29.2	254.1	49.1	1.3	254.6	42.9	8.5
	High	195.9	9.0	3.6	7.1	22.8	8.9	1.5	33.7	192.8	42.9	1.2	252.9	58.2	7.7
Suicidal Ideation	***Low***	189.0	8.5	3.1	7.3	21.6	7.1	2.1	28.1	231.4	47.2	1.3	263.0	46.6	9.0
	High	220.0	10.5	4.1	8.2	24.8	8.0	1.5	39.9	248.8	47.3	1.1	221.2	51.2	5.5

During the semantic fluency emotions task, a keyword analysis based on affect valence ratings demonstrated significant differences for distress *low* (0.46) versus *high* (0.40) severity groups, whereby the *high* severity group uttered keywords with lower valence. Unlike previous speech studies [14, 15] concerning depression, adolescents in the *high* severity group did not generate words with significantly lower valence than those in *low* severity group. For anxiety, depression, and suicidal ideation, the *low* and *high* severity groups produced similar average valence scores per recording (0.40 to 0.43). While these valence results do not align with previous free conversation speech studies [14, 15], which indicated that adults with mood disorders focus more negative terms, it should be noted the study herein involved adolescents and non-conversational speech samples.

In Table 3, semantic fluency food category nutrimental attribute keyword results demonstrated many key differences between *low* and *high* severity groups. For example, the distress *high* severity group produced significantly lower fibre (1.6g) and calcium (43.8mg) when compared to the *low* severity distress group (2.4g, 52.3mg). For depression, the *high* severity group produced the highest keyword averages for sugar (8.9g) and lowest for sodium (192.8mg) when compared to other disorders. Interestingly, high sugar consumption is linked to higher depression prevalence [20]. Moreover, the *high* severity suicidal ideation group had seven nutrimental attributes that were significantly different from the *low* severity group. Participants in the suicidal ideation *high* severity group produced keywords with significantly reduced vitamin C and potassium values when compared to the *low* severity group (i.e., these values were also much lower when compared to other disorder *high* severity values).

For all four mental illness self-surveys, the *high* severity groups, produced food keywords lower in fibre, iron, and vitamin C. Food keywords low in sodium and high in sugar were a more unique indicator for depression, whereas food keywords low in calcium were more a unique indicator for distress. Interestingly, for the four mental illness self-surveys, Table 3 results show that individuals with *high* severities uttered keywords significantly lower fibre averages than the low severity groups. Additional food keyword metrics, such as food subtypes (e.g., dairy, meat, ultra-processed), metabolomics (e.g., B-vitamins, antioxidants) and/or costs might also be useful during food semantic fluency tasks to determining elevated mood disorder severity associations.

During the sports semantic fluency task, sport metric results indicated that the depression and suicidal ideation *high* severity groups demonstrated an increase in sport keywords that were team-based (0.26, 0.36) with direct physical opponent interaction (0.70, 0.73) than the *low* severity groups who showed more balanced solo/team-based (0.11, 0.07) and opponent interaction (0.57, 0.54). For the sport location type metric, *low* and *high* severity group ranges were similar across the different mood disorders (0.42 to 0.49), whereby sports in outdoor environments were more frequently verbalized. Further analysis of the sports keyword analysis demonstrated that the number of on-field teammates was significantly higher for depression and suicidal ideation *high* severity groups (6.51, 6.85) when compared to the *low* severity groups (5.71, 5.60).

5. Conclusion

This study demonstrates that deeper analysis of the keywords spoken during semantic fluency categorical tasks, especially concerning habitual topics, can provide insights into individuals' mental lexicons. With regards to verbal semantic fluency tasks, it is shown that in addition to traditional semantic fluency correctness type metrics, new category-specific quantitative keyword indices, such as food nutritional and color shade metrics, can help identify mental lexical biases associated with increased mood disorder symptom severities.

6. Future Work

Further investigation into semantic fluency topic categories (e.g., beverages, fears, television programs) and topic-specific indices should be explored to discover additional biased lexical behaviors associated with specific mood disorders. Experimentation including non-English language and/or bilingual speakers could help to uncover whether word retrieval bias is universally exhibited across different languages/cultures. Other illnesses (e.g., autoimmune, neurological) that impact language skills could be evaluated using semantic fluency tests with topic-specific index metrics. This may help to find unique categorical word retrieval behaviors in certain illnesses; and further, possibly provide new metrics for monitoring patients' performance improvement during therapies. Due to the isolated, word-level production during recorded verbal semantic fluency tasks, automatic speech recognition with topic-specific keyword index text-processing analytics is a practical future consideration.

7. Acknowledgements

The Future Proofing Study funded by an NHMRC Project Grant (GNT1138405). We thank the Black Dog Institute (NSW, Australia), A/Prof. Aliza Werner-Seidler and Prof. Helen Christensen for their research dataset project support.

8. References

[1] The Centers for Disease Control and Prevention (CDC), "Youth Risk Behavior Survey Data Summary and Trends Report: 2011-2021", Official Report, 2021.

[2] Patel, V., Flisher, A.J., Hetrick, S., and McGorry, P., "Mental health of young people: a global public-health challenge", *The Lancet*, vol. 369 (9569), pp. 1302–1313, 2017.

[3] Copeland, W.E., Wolke, D., Shanahan, L., and Costello, J., "Adult functional outcomes of common childhood psychiatric problems: a prospective, longitudinal study", *JAMA Psychiatry*, vol. 72 (9), pp. 892–899, 2015.

[4] World Health Organization (WHO), "Adolescent & young adult health", 2023, downloaded: https://www.who.int/news-room/fact-sheets/detail/adolescents-health-risks-and-solutions

[5] Liu, L., Villavicencio, F., Yeung, D., Perin, J., and Lopez, G., "National, regional, and global causes of mortality in 5–19-year-olds from 2000 to 2019: a systematic analysis", *The Lancet Global Health*, vol. 10 (3), pp. e337–e347.

[6] Nezu, A.M., Nezu, C.M., Friedman, J., and Lee, M., "Assessment of depression", In: C.L. Hammen, I.H. Gotlib, Handbook of Depression (2nd Eds.), The Guilford Press, New York - USA, pp. 45–68.

[7] Sanchez-Moreno, J., Martinez-Aran, A., Tabares-Seisdedos, R., Torrent, C., Vieta, E., and Ayuso-Mateos, J. L., Functioning and disability in bipolar disorder: An extensive review. *Psychotherapy and Psychosomatics*, vol. 78 (5), pp. 285–297, 2009.

[8] Klumpp, H. and Deldin, P., Review of brain functioning in depression for semantic processing and verbal fluency, *Intern. J. of Psychophysiology*, vol. 75 (2), pp. 77–85, 2010.

[9] Gawda, B., and Szepietowska, E., "Trait anxiety modulates brain activity during performance of verbal fluency tasks", *Frontiers in Behavioral Neuroscience*, vol. 10 (10), pp. 1–15, 2016.

[10] Stasak, B., Joachim, D., and Epps, J., "Breaking age barriers with automatic voice-based depression detection", In: *Proc. IEEE Pervasive Computing*, vol. 21 (2), pp. 10–19, 2022.

[11] Joorman, J. and Gotlib, I.H., "Emotion recognition in depression: relation to cognitive inhibition, *Cog. & Emotion*, vol. 24 (2), pp. 281–298, 2010.

[12] Weintraub, M., Posta, F., Ichinose, M.C., Arevian, A.C., and Miklowitz, D.J., "Word usage in spontaneous speech as a predictor of depressive symptoms among youth at high risk for mood disorders", *J. Affect Disord.*, vol. 323, pp. 675–678, 2023.

[13] Zimmermann, J., Brockmeyer, T., Humm, M., Schauenberg, H., and Wolf, M., "First-person pronoun use in spoken language as a predictor of future depressive symptoms: preliminary evidence from a clinical sample of depressed patients", *Clinical Psychology & Psychotherapy*, vol. 24 (2), pp. 384–391, 2016.

[14] Gumus, M., DeSouza, D.D., Xu, M., Fidalgo, C., Simpson, W., and Robin, J., "Evaluating the utility of daily speech assessments for monitoring depression symptoms, *Digital Health*, vol 9, pp. 1–11, 2023.

[15] Yang, C., Zhang, X., Chen, Y., Li, Y., Yu, S., Zhao, B., Wang, T., and Luo, L., "Emotion-dependent language featuring depression", *J. of Behavior Therapy and Experimental Psych.*, vol. 81, 2023.

[16] Jacka, F.N., Rothon, C., Taylor, S., Berk, M., and Stansfeld, S.A., "Diet quality and mental health problems in adolescents from East London: a prospective study", *Social Psychiatry and Psychiatric Epidemiology*, vol. 48 (8), pp. 1297–1306, 2012.

[17] Khalid, S., Williams, C.M., and Reynolds, S.A., "Is there an association between diet and depression in children and adolescents? A systematic review", *British J. of Nutrition*, vol. 116 (12), pp. 2097–2108, 2017.

[18] Saghafian, F., Hajishafiee, M., Rouhani, P., and Saneei, P., "Dietary fiber intake, depression, and anxiety: a systematic review and meta-analysis of epidemiological studies", *Nutritional Neuroscience*, vol. 26 (2), pp. 108–126, 2021.

[19] Ding, J. and Zhang, Y., "Associations of dietary vitamin C and E intake with depression: a meta-analysis of observational studies", *Frontiers Nutrition*, vol. 9, pp. 1–16, 2022.

[20] Knüppel, A., Shipley, M.J., Llewellyn, C.H. et al., "Sugar intake from sweet food and beverages, common mental disorder and depression: prospective findings from the Whitehall II study", *Scientific Reports*, vol 7, pp. 1–10, 2017.

[21] Sutton, T.M, and Altarriba J., "Color associations to emotion and emotion-laden words: A collection of norms for stimulus construction and selection", *Behav. Res. Methods*, vol. 48 (2), pp. 686–728, 2016.

[22] Meier, B., Robinson, M.D., and Clore, G.L., "Why good guys wear white: automatic inferences about stimulus valence based on brightness", *Psych. Science*, vol. 15 (2), pp. 82–87, 2004.

[23] Pluhar, E., McCracken, C., Griffith, K.L., Christino, M.A., Sugimoto, D., and Meehan III, W.P., "Team sport athletes may be less likely to suffer anxiety or depression than individual sport athletes", *J. Sports Sci. Med*., vol 18 (3), pp. 490–496, 2019.

[24] Werner-Seidler, A., Huckvale, K., Larsen, M.E. et al., "A trial protocol for the effectiveness of digital interventions for preventing depression in adolescents: the future proofing study", *Trials*, vol. 21 (1), pp.1–21, 2020.

[25] O'Dea, B., Han, J., Batterham, P.J., Achilles, M.R., Calear, A.L., Werner-Seilder, A., Parker, B., Shand, F., and Christensen, H., "A randomized controlled trial of a relationship-focussed mobile phone application for improving adolescents' mental health", *J. of Child Psych. and Psych*., vol. 61 (8), pp. 899–913, 2020.

[26] Süsstrunk, S., Buckley, R., and Swen, S., "Standard RGB color spaces", In: *Proc. IS&T/SID 7th Color Imaging Conf*., vol. 7, pp. 127–134, 1999.

[27] Mohammad, S.M., "Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 English words", In: *Proc. 56th Annual Meeting of the Assoc. for Comp. Ling*., Melbourne, VIC - Australia, pp. 174–184, 2018.

[28] United State Department of Agriculture (USDA), *Agricultural Research Service*, downloaded Sept. 2023: https://fdc.nal.usda.gov

[29] Kim, T.K., "T test as a parametric statistic", *Korean J. Anesth*., vol. 68 (6), pp. 540–546, 2015.

[30] Hedges, L.V., and Olkin, I., Statistical Methods for Meta-Analysis, Academic Press, Orlando, FL - USA, 1985.

[31] Kelly, K. and Preacher, K., "On effect size", *Psych. Methods*, vol. 17 (2), pp. 137–152, 2012.