

Natural language processing of youth speech predicts psychopathology across adolescence

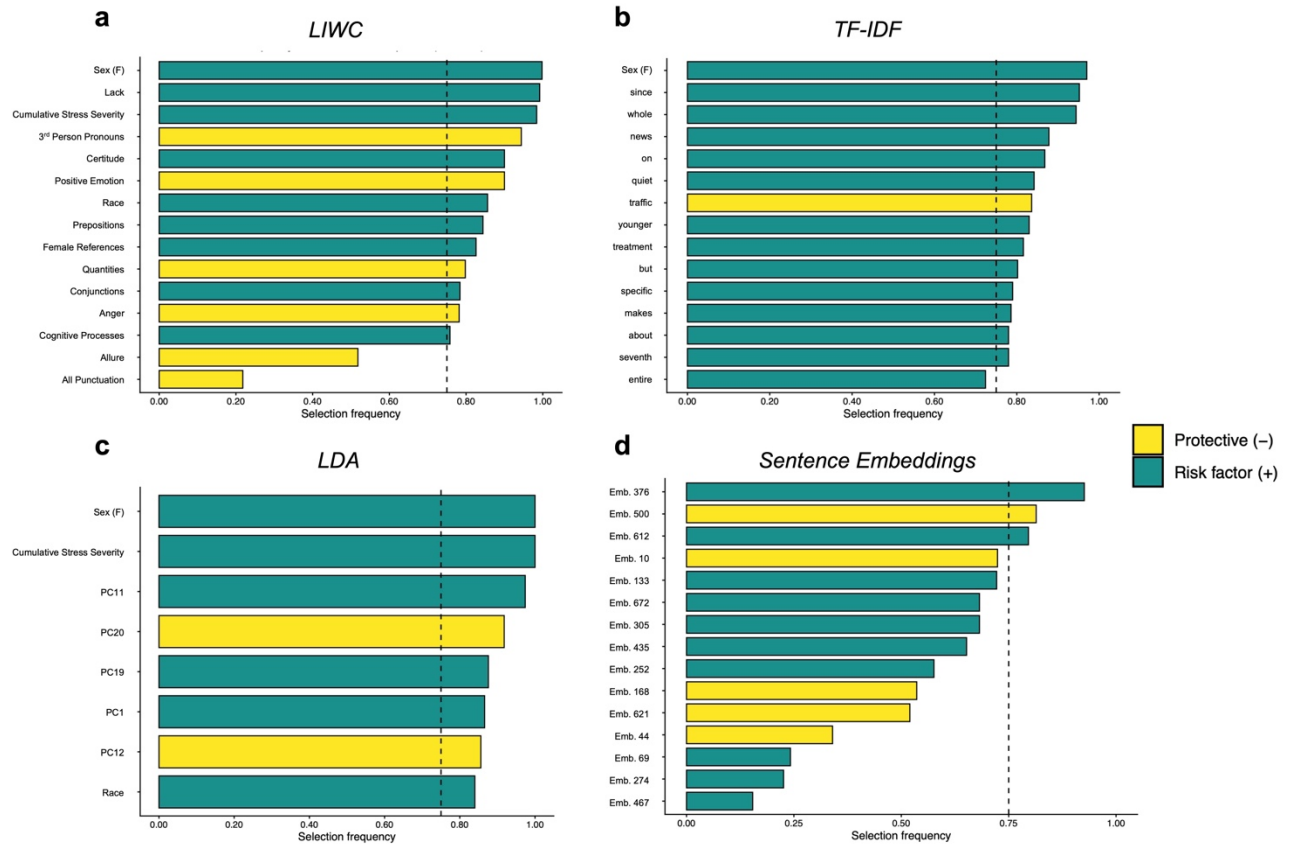
In the format provided by the
authors and unedited

Supplemental Information

Contents

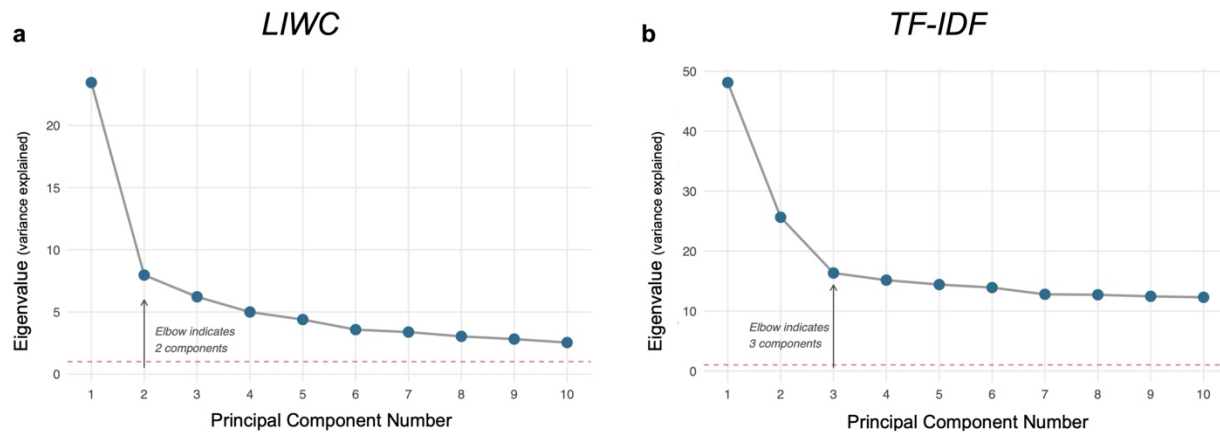
1. Figure 1: Bootstrap Stability of Elastic Net Feature Selection
2. Figure 2: LIWC and TF-IDF Scree Plots
3. Table 1: Diagnostic operating characteristics of baseline and language-augmented logistic regression models
4. Figure 3: ROC Curves for Diagnostic Prediction: Baseline vs Language-Augmented Models
5. Table 2: Top 100 and bottom 100 sentences from dot product projection of elastic net model weights onto individual sentence embeddings
6. TESI Interview questions
7. Standardized Prompt for Secure GPT-4o transcript cleaning
8. Table 3: Coding of demographic variables
9. Analysis of sex as a moderator

1. Figure S1: Bootstrap Stability of Elastic Net Feature Selection



Note. Bootstrap stability of elastic net feature selection across NLP methods ($B = 500$ refits, $N = 204$). Bars show the selection frequency (proportion of bootstrap refits with non-zero coefficients) for the top 15 predictors from each full-sample elastic net model; within each panel, predictors are ordered by selection frequency. Colors indicate the direction of association in the full-sample model (risk factor: positive coefficient; protective: negative coefficient). The dashed vertical line marks a 75% selection-frequency threshold used to denote highly stable features. Feature selection was highly stable for LIWC, TF-IDF, and LDA (LIWC: 13/15 predictors; TF-IDF: 14/15 predictors; LDA: 15/15 predictors) with selection frequencies $\geq 75\%$ across refits. In contrast, for RoBERTa sentence embeddings, the top-ranked embedding dimensions showed more moderate selection frequencies, consistent with a high-dimensional, collinear embedding space, predictive signal is distributed across many correlated dimensions such that the specific embedding coefficients selected can vary between refits, even as overall model performance remains robust.

2. Figure S2: LIWC and TF-IDF Scree Plots



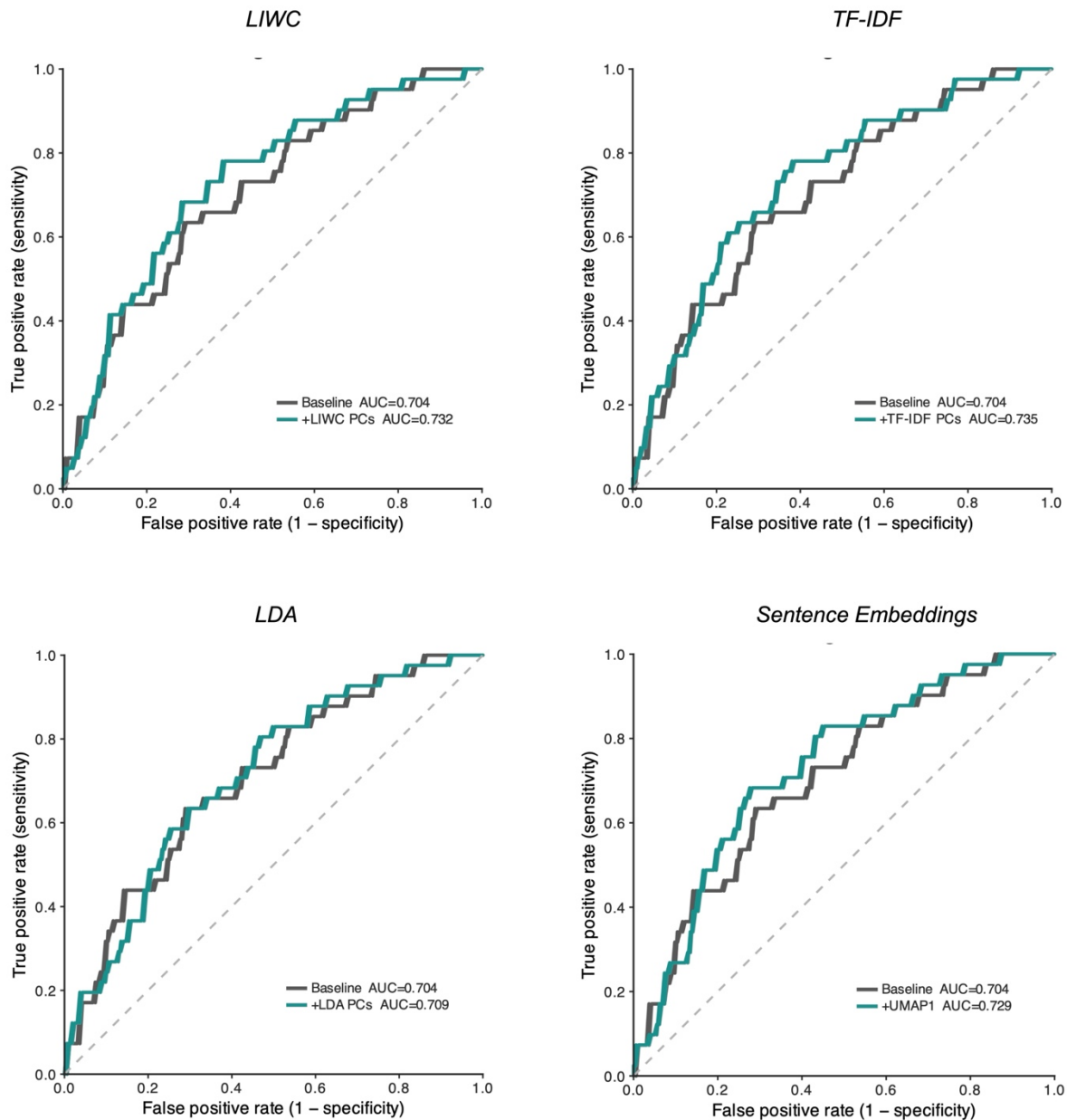
Note. Scree plots show eigenvalues across principal components for the (a) LIWC and (b) TF-IDF feature sets; the elbow criterion supported retaining two components for LIWC and three components for TF-IDF, consistent with the PCA solutions used in the main analyses. No scree plot is shown for the LDA pipeline because topics were reduced *a priori* by applying PCA to the 500-topic participant-topic matrix and retaining 20 orthogonal components to yield a stable, interpretable thematic space for prediction (see Methods). No scree plot is shown for sentence embeddings because PCA was not applied to the RoBERTa embeddings; instead, contextualized embeddings were used directly in predictive models, and UMAP was applied for low-dimensional visualization and semantic-structure analyses. N = 204.

3. Table S1: Diagnostic Operating Characteristics of Baseline and Language-Augmented Logistic Regression Models

	AUC	Bootstrap 95% CI	Youden Threshold	Sensitivity	Specificity	PPV	NPV
Baseline Model (Demographics + SumSev)	0.704	[0.616, 0.785]	0.240	0.634	0.712	0.356	0.885
LIWC Model (Demographics + SumSev + LIWC PCs)	0.732	[0.648, 0.812]	0.238	0.683	0.718	0.378	0.900
TF-IDF Model (Demographics + SumSev + TF-IDF PCs)	0.735	[0.649, 0.814]	0.182	0.780	0.620	0.340	0.918
LDA Model (Demographics + SumSev + LDA PCs)	0.709	[0.623, 0.792]	0.238	0.634	0.706	0.351	0.885
RoBERTa Model (Demographics + SumSev + RoBERTa PCs)	0.729	[0.641, 0.805]	0.226	0.683	0.724	0.384	0.901

Note. Diagnostic operating characteristics of baseline and language-augmented logistic regression models. Area under the ROC curve (AUC) is reported with nonparametric bootstrap 95% confidence intervals (B=2,000 resamples; n=204 participants). Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) are computed at the probability threshold maximizing Youden's J statistic. Models compare a baseline covariate model (demographics + expert-rated cumulative stress severity) against models augmented with linguistic dimensions derived from LIWC, TF-IDF, LDA, and RoBERTa features. Prevalence of mood/anxiety disorders in the sample = 20.1%. SumSev = expert-rated cumulative stress severity.

4. Figure S3: ROC Curves for Diagnostic Prediction: Baseline vs Language-Augmented Models



Note. Receiver operating characteristic (ROC) curves comparing a baseline logistic regression model (demographics + expert-rated cumulative stress severity, SumSev) to models augmented with language-derived predictors for mood/anxiety diagnosis (prevalence = 20.1%). Across methods, language features yielded modest improvements in overall discrimination relative to baseline (AUCs shown in-panel; operating characteristics and bootstrap 95% CIs ($C=2,000$ resamples) reported in Table S1). Importantly, consistent with the inferential aim of these analyses, language-derived dimensions significantly predicted diagnostic outcomes in three of the four NLP approaches (LIWC PCs, TF-IDF PCs, and RoBERTa/UMAP1), whereas SumSev did not significantly predict diagnosis in these models; only the LDA-derived topic components did not show a significant association. Dashed diagonal indicates chance performance. $N = 204$.

5. Table S2: Top 100 and bottom 100 sentences from dot product projection of elastic net model weights onto individual sentence embeddings

A significant challenge in using transformer-based embeddings for clinical prediction is their "black box" nature, which often obscures the specific features driving model performance. To overcome this, we employed a novel interpretive technique to directly link the model's prediction back to the raw text; specifically, we projected the 768-dimensional coefficient vector learned by the elastic net model onto each individual sentence embedding via the dot product. This yielded a score for every sentence, representing its direction in "risk space" and its contribution to the overall prediction. The following are the top 100 (0.3%) sentences associated with the greatest risk, and the bottom 100 (0.3%) sentences associated with the lowest risk. PHI has been redacted and replaced with descriptors, in brackets.

Sentence		Predicted YSR Score
#	Highest Predicted Risk	
1	I chose not to.	308.56
2	Um, they were sort of punching each other and holding each other and choking each other.	303.24
3	And then I stopped and everything just like.	301.07
4	Both stories conflict each other.	288.16
5	And there's a girl that was paired up with my sister, but for some reason she became like sort of obsessed with me so she'd follow me.	275.07
6	That's why it was just total silence.	275.01
7	And the other guy just went up to him and just did it.	267.02
8	And then all of a sudden I saw this guy that looked exactly like my dad.	265.89
9	Why didn't you tell me?	265.26
10	No, except for one time it did.	260.69
11	And then she just is like, I don't know.	259.69
12	And then my cousin went outside and was like screaming.	257.75
13	He threatened to tell my sister on more than one occasion.	254.76
14	There was one girl who was having a problem with something that was going on so she called certain people in there for different reasons.	250.67
15	And then she like raises and raises and raises her voice.	250.13
16	It happened one time and another time.	248.50
17	Only on a different occasion.	247.77
18	It just depended on what was going on.	247.52
19	They were just completely ignored.	247.45
20	So she threatened me and actually hurt me.	246.41
21	So I was wondering what the heck was happening.	245.24
22	No, they just kind of stared at each other.	245.09
23	It would only happen every once in a while.	244.98
24	Okay, and so you didn't go with them.	244.77
25	What got what was going on?	244.42
26	Okay, and when was that?	243.48

27	They were kind of just like trying, they were kind of ignoring it.	243.43
28	Like why was I literally on fire.	243.33
	Um, I was like best friend with this girl and she sort of like turned and she	243.28
29	just got into a fight with me and she kind of told everyone other things that I was not saying.	
30	He was just pushing me.	242.62
31	Then he like pushed me aside and was like, whoopah!	242.07
32	They just kind of like did things.	242.03
33	And when I was seven, the one bully only bullied me.	241.67
34	Because I just moved somewhere else.	241.46
35	She was actually just sitting there, just yelling.	240.95
36	And the entire school went against me.	237.94
37	Yeah, I was still like three and I was just like, I couldn't do anything because he's a teacher and everything.	237.66
	So then, like, I felt that something was going to happen, and then	237.52
38	something happened.	
39	Then like a whole bunch of black smoke just going up.	237.22
40	And I'm like, oh, ow.	237.08
41	and then just pushed me away.	236.39
42	A random person just went up to him and shot him.	236.15
43	That's what happened.	235.55
44	They just kind of argued.	235.54
45	And then she told the story and that's how that happened.	235.29
46	And when I was eight, the one bully only bullied me.	235.24
47	They kind of just disagree on things.	234.32
48	No, nothing other than panic.	234.28
49	And then everyone started to freak out.	234.23
50	I was just building in my room.	233.77
51	What happened with that?	233.05
52	And then what happened?	232.75
53	And pushing each other and then they walked away.	232.43
54	And I was like, okay, okay, okay, okay, don't shoot me, don't shoot me.	231.60
55	At first it began to start, it started as one to two, and then it began increasing.	231.22
56	Okay, and what happened with them?	229.30
57	That was what caused me to throw up.	229.29
58	Okay, and what happened?	229.10
59	She was walking behind me with my brother.	229.09
60	What had happened?	228.91
61	Because everyone started freaking out.	228.73
62	They were shouting at each other and shoving each other.	228.35
63	They just threatened to hit each other.	228.18
64	It was just a thing that happened and then it was over.	227.87
65	It was like, I don't know.	227.82
66	He just kind of started bullying me.	227.10
67	They would just yell at each other.	227.09

68	Just sat there watching people, you know, hit each other.	227.02
69	And it just annoyed me.	226.95
70	Or was there a certain time where it stood out to you?	226.85
71	Um, one time my counselor in my school where there was a conflict between me and two other classmates that spanned three years and the teachers just had enough.	226.73
72	He was just, he was just weird.	226.58
73	Okay, and what did you see happen?	226.35
74	So I told my mom, and then she told my teacher, and then she told my principal, and then she brought, and then the principal brought me, and he made me write on this little letter that would basically tell what happened.	225.91
75	They sort of just did things that kind of acted like they were going to do things to me but didn't.	225.10
76	I've only told like two of my friends.	224.91
77	He punched me, it was then immediately stopped.	224.87
78	No, except for one time.	224.84
79	And she was like, I don't know.	224.50
80	No, he just ignored me.	224.31
81	She knew what had to be done and everything.	223.74
82	And then I found out.	223.53
83	And then towards the end of the year, he just stepped up to my brother and bent his finger backwards.	223.42
84	I mean, one and a half second.	223.37
85	Okay, when was that?	223.20
86	I was like, I don't know.	223.16
87	I'm like, I don't know.	222.84
88	My brother was okay with it.	222.76
89	Except for two of my friends.	222.75
90	It just kind of happens.	222.36
91	I was just sitting there.	222.14
92	They just kind of yell at each other.	221.49
93	They just kind of dismissed me.	221.27
94	I exactly knew that my sister was swinging around.	221.18
95	No one else knew.	220.78
96	And then other people would join in and start calling me that same name.	220.69
97	They were just kind of arguing over things.	220.38
98	And then he stopped when I was eight.	219.91
99	There was only this one time where this girl it's just her personality, and now I understand that.	219.91
100	I didn't see it happening, but I saw the aftermath.	219.90

Sentence		Predicted YSR Score
#	Lowest Predicted Risk	
1	Kaiser Permanente Hospital.	-198.05
2	[School name]	-193.02
3	B, C, and A?	-192.30
4	I like the Bugatti Veyron.	-179.33
5	As we start to start off a new year, a new school year, we start out well.	-173.94
6	and the Homework Club, and the Breakfast Book Club.	-173.81
7	privileges?	-173.74
8	Therapist, psychologist, psychiatrist?	-171.35
9	arm...	-170.70
10	Grades?	-170.50
11	Antidepressants, stimulants, lithium?	-170.26
12	Buck Club and Performing Arts Club.	-164.99
13	10?	-163.07
14	Um, Foursquare, soccer, um...	-161.98
15	[School name]	-161.39
16	Boys and Girls Club.	-159.38
17	Teenagers?	-157.76
18	[Date]?	-157.43
19	[School name]	-155.70
20	[School name]	-154.06
21	Um, St.	-150.31
22	arms...	-149.90
23	Older?	-148.36
24	Before it's bed, we have some party.	-147.57
25	Let's do some ER experiences.	-145.84
26	[Date].	-145.40
27	What's a turtle bed?	-145.09
28	Ballet.	-142.50
29	Shortstop and pitcher.	-141.94
30	Poly-Mona-Sorting.	-138.65
31	Grandma [Name] or Grammy.	-137.66
32	Well, at the YMCA, I do swimming, so...	-136.32
33	[Name]	-135.75
34	Let's say...	-135.32
35	Like finder keepers, loser sweepers.	-133.28
36	[Name]	-133.18
37	Not gotten...	-132.23
38	[Name] or [Name] gotten fairly...	-131.47
39	[Friends' names]?	-130.86
40	Soccer?	-130.03
41	Um, did you guys have to really work hard to put food on the table or get new clothes?	-129.19
42	Hm?	-129.12

43	[Name]'s, so...	-128.40
44	Closest?	-128.09
45	We even made a quilt for him.	-126.36
46	It was really easy for the beginning of the year.	-125.80
47	Tests are hard.	-125.71
48	Like a therapist, or a psychiatrist, or just a regular doctor?	-125.14
49	27?	-123.76
50	[Name]	-123.59
51	[Date]	-123.46
52	The hard math homework.	-122.96
53	Emotions?	-122.18
54	Gymnastics, dance.	-121.45
55	We have a garden on the side of the driveway.	-121.01
56	On a what retreat?	-120.98
57	[City].	-120.57
58	[Name]	-119.24
59	Mm-hmm?	-118.38
60	[Date].	-118.35
61	[School name]	-117.89
62	But in school, just the curriculum isn't very rigorous.	-117.75
63	Play video games or read a book.	-117.69
64	Well, my dad sometimes he works he works for [Company name] and sometimes they have um They do send him out to they send a group of people out to [City] to take some classes Normally, they're gone for about a week, but we they do have their cell phones with them and stuff So we we do maintain in connection.	-117.59
65	Like, Again, like...	-116.84
66	No Have you ever been in the hospital for your emotions?	-116.65
67	Scary movies?	-116.35
68	Annoying?	-116.34
69	Like A's, or even A.	-116.11
70	[Date].	-115.82
71	[Name]	-115.42
72	9?	-115.25
73	I think I've been in the hospital to take out my...	-114.44
74	I like the Lamborghini Aventador.	-114.42
75	Taken any medications for your emotions, like an antidepressant or a stimulant?	-113.69
76	I'm not very good at it.	-113.45
77	A 0.	-113.39
78	Been really, really sick?	-113.20
79	Gymnastics.	-113.09
80	Like are they...	-112.85
81	So [Name].	-111.91
82	[Dormitory name]?	-111.83
83	[Name]	-111.73

84	[City].	-111.46
85	Trying, or reading, or playing games.	-111.32
86	Uh, Give kids hearing aids for free.	-110.87
87	eight...	-110.44
88	[Friends' names]?	-110.40
89	I go to [School name].	-110.36
90	So, He, uh...	-110.16
91	Minutes.	-110.13
92	Kids' Claritin or regular Claritin if it's really bad.	-109.67
93	[School name] Art.	-108.60
94	It's a Doctor Who Bevy, so we talk about Doctor Who.	-108.49
95	[Name]	-108.31
96	Breast implant.	-107.39
97	Auntie [Name].	-107.32
98	What's a ballpark?	-106.95
99	Well, at school or like...	-106.90
100	Hi, Mom.	-106.78

6. Traumatic Events Screening Inventory for Children (TESI) Questions and Structure

Background Section

1. Do you have a nickname or another name that you prefer, or should we just call you ____ when you're here with us?
2. What gender do you identify as?
3. What city do you live in?
4. Do you go to school?
5. Where do you go to school?
6. What grade are you in?
7. Do you like school?
8. What do you like about school?
9. Do you have a favorite subject?
10. What don't you like about school?
11. Do you have a least favorite subject?
12. Do you play any sports or belong to any clubs?
13. How do you get along with the people at school?
14. Have you had any problems with the people at school, like getting teased, picked on, or getting into fights?
15. Who are your closest friends and first names are fine?
16. Are you social right now?
17. What are your hobbies? What do you like to do with your friends?
18. Who lives with you at home?
19. Do you have any siblings that don't live with you at home?
20. Do you have any contact with your biological ____? What kind of contact do you have?
21. How do you get along with your mom?
22. If you were in trouble, would you go to your mom for help or advice?
23. How do you get along with your dad?
24. If you were in trouble, would you go to your dad for help or advice?
25. How do you get along with your siblings?
26. Are there any other family members or adults in your life that you're close to and you'd go to for help or advice if you needed it?
27. Do you currently take any medications regularly or on a daily basis?
28. Since your last visit, have you taken any medications regularly or on a daily basis, but have stopped as of today?
29. Since your last visit, have you ever seen a counselor, psychologist, psychiatrist or other professional for your emotions?
30. Since your last visit, have you been in the hospital for your emotions?

TESI-C

31. Experienced Traumatic Accident

- a. Have you ever been in a really bad accident where you or someone else could have been (or actually was) badly injured or killed?
 - b. Did any of these events cause problems for you at home with your family, with your friends, or at school?
 - c. How bad were these events on a scale of 0-3?
 - d. How scared, confused, or helpless did you feel on the scale of 0-3?
32. Witnessed Traumatic Accident
- a. Have you ever seen a really bad accident (that didn't happen to you) where someone could have been (or actually was) badly injured or killed?
 - b. Did any of these events cause problems for you at home with your family, with your friends, or at school?
 - c. How bad were these events on a scale of 0-3?
 - d. How scared, confused, or helpless did you feel on the scale of 0-3?
33. Natural Disaster
- a. Have you ever been in a natural disaster, such as a tornado, hurricane, fire, floods, or earthquake where someone could have been, or actually was, badly injured or killed?
 - b. What about a natural disaster where your family had to leave their home?
 - c. Did any of these events cause problems for you at home with your family, with your friends, or at school?
 - d. How bad were these events on a scale of 0-3?
 - e. How scared, confused, or helpless did you feel on the scale of 0-3?
34. Experience Illness/Injury
- a. Have you ever been so sick that you or the doctor thought you might die?
 - b. Have you ever had to stay overnight in a hospital or have an operation where you were put to sleep the entire time?
 - c. Did any of these events cause problems for you at home with your family, with your friends, or at school?
 - d. How bad were these events on a scale of 0-3?
 - e. How scared, confused, or helpless did you feel on the scale of 0-3?
35. Witness Illness/Injury
- a. Has someone close to you ever been so badly injured or so sick that he or she almost died?
 - b. Has someone close to you ever had to stay overnight in a hospital or have an operation where they were put to sleep the entire time?
 - c. Did any of these events cause problems for you at home with your family, with your friends, or at school?
 - d. How bad were these events on a scale of 0-3?
 - e. How scared, confused, or helpless did you feel on the scale of 0-3?
36. Traumatic Death of a Loved One
- a. Has someone close to you ever passed away?
 - b. Did any of these events cause problems for you at home with your family, with your friends, or at school?
 - c. How bad were these events on a scale of 0-3?
 - d. How scared, confused, or helpless did you feel on the scale of 0-3?
37. Separation from Loved One (Work, Travel)

- a. Has there ever been a time when you did not have contact with one of your parents because they were away for a long time, such as, being away for work for several days or longer?
 - b. Did any of these events cause problems for you at home with your family, with your friends, or at school?
 - c. How bad were these events on a scale of 0-3?
 - d. How scared, confused, or helpless did you feel on the scale of 0-3?
38. Moved Between Families/Others Moving In or Out of Home
- a. Has your family ever moved?
 - b. Has anyone moved into or out of your home?
 - c. Did any of these events cause problems for you at home with your family, with your friends, or at school?
 - d. How bad were these events on a scale of 0-3?
 - e. How scared, confused, or helpless did you feel on the scale of 0-3?
39. Parental Divorce/Separation
- a. Have your parents ever separated or divorced?
 - b. Did any of these events cause problems for you at home with your family, with your friends, or at school?
 - c. How bad were these events on a scale of 0-3?
 - d. How scared, confused, or helpless did you feel on the scale of 0-3?
40. Separation from Loved One (Rehab, Foster Center, Detention Center)
- a. Have you ever been separated from someone you depend on for love or security for more than a few days, for example, you or they went to live in a foster home, a group home, or with an adoptive family?
 - b. You or they went to a detention center, residential treatment center, hospital or prison?
 - c. You or they were in a war or disaster?
 - d. Did any of these events cause problems for you at home with your family, with your friends, or at school?
 - e. How bad were these events on a scale of 0-3?
 - f. How scared, confused, or helpless did you feel on the scale of 0-3?
41. Witness to Loved One Mental Illness/Substance Abuse
- a. Has someone you depend on for love or security ever had a serious mental or emotional problem such as, depression, seeing or hearing things that aren't real or panic attacks?
 - b. Did they ever get very drunk or use drugs a lot?
 - c. Did any of these events cause problems for you at home with your family, with your friends, or at school?
 - d. How bad were these events on a scale of 0-3?
 - e. How scared, confused, or helpless did you feel on the scale of 0-3?
42. Witness to Self-Harm/Suicide
- a. Has someone close to you ever tried to hurt him or herself on purpose?
 - b. Has anyone you love or care a lot about ever tried to kill him or herself?
 - c. Did any of these events cause problems for you at home with your family, with your friends, or at school?
 - d. How bad were these events on a scale of 0-3?

- e. How scared, confused, or helpless did you feel on the scale of 0-3?
- 43. Physical Abuse/Attack
 - a. Has anyone ever physically attacked you or have you ever been so badly hurt by someone that you had to go to the doctor or hospital?
 - b. What about being attacked with a dangerous weapon?
 - c. Did any of these events cause problems for you at home with your family, with your friends, or at school?
 - d. How bad were these events on a scale of 0-3?
 - e. How scared, confused, or helpless did you feel on the scale of 0-3?
- 44. Threat of Physical Abuse/Attack
 - a. Has someone ever threatened to hurt you really badly or said they would kill you, even if they didn't actually do it?
 - b. Has someone acted like they were going to hurt you really badly or kill you, even if they didn't actually do it?
 - c. Did any of these events cause problems for you at home with your family, with your friends, or at school?
 - d. Did any of these events cause problems for you at home with your family, with your friends, or at school?
 - e. How bad were these events on a scale of 0-3?
 - f. How scared, confused, or helpless did you feel on the scale of 0-3?
- 45. Victim of Mugging/Intimidation/Bullying
 - a. Has anyone ever robbed you? Or attacked you to steal from you?
 - b. Has anyone ever bullied you? Or made you feel scared or intimidated?
 - c. Have you ever seen a family member or someone you care about get robbed or bullied?
 - d. Did any of these events cause problems for you at home with your family, with your friends, or at school?
 - e. How bad were these events on a scale of 0-3?
 - f. How scared, confused, or helpless did you feel on the scale of 0-3?
- 46. Kidnapping
 - a. Has anyone ever kidnapped you or taken you away from your home when they weren't supposed to?
 - b. Has someone close to you ever been kidnapped?
 - c. Did any of these events cause problems for you at home with your family, with your friends, or at school?
 - d. How bad were these events on a scale of 0-3?
 - e. How scared, confused, or helpless did you feel on the scale of 0-3?
- 47. Animal Attack
 - a. Have you ever been attacked by a dog or another animal?
 - b. Did any of these events cause problems for you at home with your family, with your friends, or at school?
 - c. How bad were these events on a scale of 0-3?
 - d. How scared, confused, or helpless did you feel on the scale of 0-3?
- 48. Domestic Violence (Non-Physical)
 - a. Has there ever been a time when your parents were fighting a lot or did not get along?

- b. What about any other family members?
 - c. Did any of these events cause problems for you at home with your family, with your friends, or at school?
 - d. How bad were these events on a scale of 0-3?
 - e. How scared, confused, or helpless did you feel on the scale of 0-3?
49. Domestic Violence Threat
- a. Have there been any other times when you saw or heard people in your family act like they were going to kill or hurt each other really badly, even if they didn't actually do it?
 - b. Did any of these events cause problems for you at home with your family, with your friends, or at school?
 - c. How bad were these events on a scale of 0-3?
 - d. How scared, confused, or helpless did you feel on the scale of 0-3?
50. Domestic Violence (Physical)
- a. Have you ever seen or heard people in your family physically fighting?
 - b. If yes, Did they use any dangerous weapons like a knife or gun?
 - c. Did any of these events cause problems for you at home with your family, with your friends, or at school?
 - d. How bad were these events on a scale of 0-3?
 - e. How scared, confused, or helpless did you feel on the scale of 0-3?
51. Family Legal Trouble/Arrest
- a. Has anyone in your family ever had any sort of trouble with the law, such as having to go to court?
 - b. Have you ever seen or known that a family member who was arrested or taken away by the police or authorities?
 - c. Did any of these events cause problems for you at home with your family, with your friends, or at school?
 - d. How bad were these events on a scale of 0-3?
 - e. How scared, confused, or helpless did you feel on the scale of 0-3?
52. Witness to Stranger Arguments
- a. Have you ever seen or heard people outside of your family really yelling or screaming at each other a lot at school or in your neighborhood?
 - b. Did any of these events cause problems for you at home with your family, with your friends, or at school?
 - c. How bad were these events on a scale of 0-3?
 - d. How scared, confused, or helpless did you feel on the scale of 0-3?
53. Witness to Stranger Physical Fights
- a. Have you ever seen or heard people outside of your family physically fighting at school or in your neighborhood?
 - b. Did any of these events cause problems for you at home with your family, with your friends, or at school?
 - c. How bad were these events on a scale of 0-3?
 - d. How scared, confused, or helpless did you feel on the scale of 0-3?
54. Witness to War/Terrorism

- a. Have you ever seen or heard people attacking each other for real on television like a war on the news or a building blowing up or a terrorist attack? (note: a real event, not watching a film about a war)
 - b. Did any of these events cause problems for you at home with your family, with your friends, or at school?
 - c. How bad were these events on a scale of 0-3?
 - d. How scared, confused, or helpless did you feel on the scale of 0-3?
55. Victim of Sex or Sexual Abuse
- a. Has someone ever made you see or do something sexual that was unwanted?
 - b. What about make you see or touch them in a way that made you feel uncomfortable?
 - c. What about make you watch them touch themselves in a way that made you feel uncomfortable?
 - d. Did any of these events cause problems for you at home with your family, with your friends, or at school?
 - e. How bad were these events on a scale of 0-3?
 - f. How scared, confused, or helpless did you feel on the scale of 0-3?
56. Witness to Sexual Abuse
- a. Have you seen or heard someone else doing something sexual that made you feel uncomfortable or was unwanted?
 - b. Did any of these events cause problems for you at home with your family, with your friends, or at school?
 - c. How bad were these events on a scale of 0-3?
 - d. How scared, confused, or helpless did you feel on the scale of 0-3?
57. Emotional Abuse
- a. Have you ever repeatedly been told by adults who were supposed to love and care about you that you were no good or them yelling at you in a scary way?
 - b. What about them wishing they didn't have to take care of you, wanting to send you away, or wishing you had never been born?
 - c. Did any of these events cause problems for you at home with your family, with your friends, or at school?
 - d. How bad were these events on a scale of 0-3?
 - e. How scared, confused, or helpless did you feel on the scale of 0-3?
58. Neglect
- a. Was there ever a time in your life when you did not have the right care, like not having enough to eat or drink?
 - b. What about being left alone when you were too young or unable to care for yourself?
 - c. What about being left alone with someone who was very drunk or using drugs?
 - d. Did any of these events cause problems for you at home with your family, with your friends, or at school?
 - e. How bad were these events on a scale of 0-3?
 - f. How scared, confused, or helpless did you feel on the scale of 0-3?

59. Financial Troubles

- a. Has there ever been a time when you or your family had money troubles?
- b. Did any of these events cause problems for you at home with your family, with your friends, or at school?
- c. How bad were these events on a scale of 0-3?
- d. How scared, confused, or helpless did you feel on the scale of 0-3?

60. Other

- a. Is there anything we haven't talked about that seems to fit in with the questions that I have been asking you? Something that was pretty scary or bad?
- b. Did any of these events cause problems for you at home with your family, with your friends, or at school?
- c. How bad were these events on a scale of 0-3?
- d. How scared, confused, or helpless did you feel on the scale of 0-3?

7. Standardized Prompt for Secure GPT-4o Data Cleaning

The below prompt was used to correct occasional speaker label errors generated by PyAnnote. This post hoc text-only relabeling step was performed using a secure, intuitively compliant GPT-4o deployment (<https://securegpt.stanfordhealthcare.org/>) configured for PHI with no external data retention. The model inferred speaker roles from lexical cues and turn-taking patterns and did not alter transcript content. Corrected speaker labels were used to remove interviewer speech. Given the input token limit, transcripts were input in 150-line chunks.

// New chat for each subject //

You are provided with a diarized transcript of a clinical interview below. Each line contains the following:

- 1. start time,*
- 2. stop time,*
- 3. speaker label (e.g., SPEAKER_00, SPEAKER_01), and*
- 4. transcribed text.*

The transcription was generated with WhisperX and is generally accurate, but the speaker assignments are often incorrect. Your task is to revise the speaker labels based on contextual language cues, so that the resulting transcript reflects accurate speaker turns between the interviewer and participant.

Instructions:

- Use the content and flow of the conversation to infer correct speaker changes (e.g., question vs answer, first-person vs second-person pronouns, discourse markers).*
- DO NOT ALTER THE TRANSCRIPTION TEXT OR TIMESTAMPS—only correct speaker labels.*
- Assume that there are only two speakers throughout: Interviewer and Participant. Assign these labels consistently.*
- Ensure that every line of the input appears exactly once in the output, with only the speaker label corrected. Do not remove or duplicate lines.*
- Return the corrected transcript in the same structure as the input file, replacing misassigned speaker labels with the correct ones.*
- Only output the corrected transcript—do not add explanations, summaries, or extra text. Also return the complete transcript; do not stop early.*

Let me know if this makes sense, and I will paste in the transcript.

...

[TRANSCRIPT - 150 lines at a time]

...

8. Table S3: Coding of Demographic Variables

Variable	Coding Scheme/Description
Sex	1 = Male; 2 = Female
Race	1 = White; 2 = Black; 3 = Hispanic; 4 = Asian; 5 = Biracial; 6 = Other
Income-to-Needs Ratio	To operationalize SES, we calculated the Income-to-Needs ratio for each child's family. The parent who accompanied the child to the laboratory session reported total family income over the previous 12 months and the number of people in their family. We collected income data in bins, with parents reporting income on a 10-point scale as follows: $\leq \$5000$ ($N = 1$); $\$5001-\$10,000$ ($N = 2$); $\$10,001-\$15,000$ ($N = 1$); $\$15,001-\$25,000$ ($N = 6$); $\$25,001-\$35,000$ ($N = 2$); $\$35,001-\$50,000$ ($N = 10$); $\$50,001-\$75,000$ ($N = 13$); $\$75,000-\$100,000$ ($N = 17$); $\$100,001-\$150,000$ ($N = 40$); $\geq \$150,000$ ($N = 54$). To calculate the INR, we divided the midpoint of the endorsed income bin by the low-income limit for Santa Clara county (80% of the median income) determined by the Department of Housing and Urban Development based on the number of people in the household (https://www.huduser.gov/portal/datasets/il/il2017/2017summary.odn ; \$84,750 for a family of four within the study period).
Parent Education	1 = less than high school diploma; 2 = GED/high school diploma; 3 = some college; 4 = two-year college degree; 5 = four-year college degree; 6 = master's degree; 7 = professional degree; 8 = doctorate

9. Analysis of sex as a moderator

Given that females reported significantly higher internalizing symptoms and had higher rates of mood or anxiety diagnoses in our sample, we conducted a supplementary analysis to test whether sex moderated the effect of our primary language predictors on symptom severity. To do this, we added a sex-by-language interaction term to the final linear regression models for each of the four main analytic approaches (LIWC-PCA, TF-IDF-PCA, LDA-PCA, and RoBERTa-UMAP). Across all models, none of the interaction terms were statistically significant (all $ps > .105$), indicating that the predictive utility of the linguistic features did not differ between males and females.