

BACKGROUND & PURPOSE

- The Frankl Behavior Rating Scale (FBRS) is one of the most used behavior scales in pediatric dentistry.^{4,6}
 - Helpful for treatment planning.
 - The FBRS has general guidelines to assess patient behavior but can be subjective when not applied as intended.
- Natural language processing (NLP) is an area of research and application that explores how computers can be used to understand and manipulate natural language text to do useful things.^{1,2,5}
- The purpose of our study aims to evaluate the current method of assigning behavior scores at Nationwide Children's Hospital (NCH) and to utilize NLP to assign behavior scores more objectively and accurately to our pediatric dental patients.**

METHODS

(1) Video Survey

- Pediatric dental faculty & residents were shown 10 recorded dental visits and requested to provide:
 - Overall FBRS score
 - Behavior comment
- The survey was administered through REDCap.
- 19 participants completed the survey: 8 faculty & 11 residents.

(2) Manual Annotation

- Manual annotations were completed using MedTator, a serverless annotation software.
- Three, independent annotators reviewed 200 behavior notes, where restorative treatment was completed, from January 2018 – July 2023.
 - Each annotator reviewed the data set two separate times:
 - To identify the overall FBRS score.
 - To identify key words and phrases.
 - temperament, action, behavior management technique, cooperation/overall takeaway
 - +/, +, N/A, -, -/-

(3) Adjudication and Analysis of Protocol

- Overall FBRS Score Analysis
- Entity Tagging Analysis

(1) Video Survey

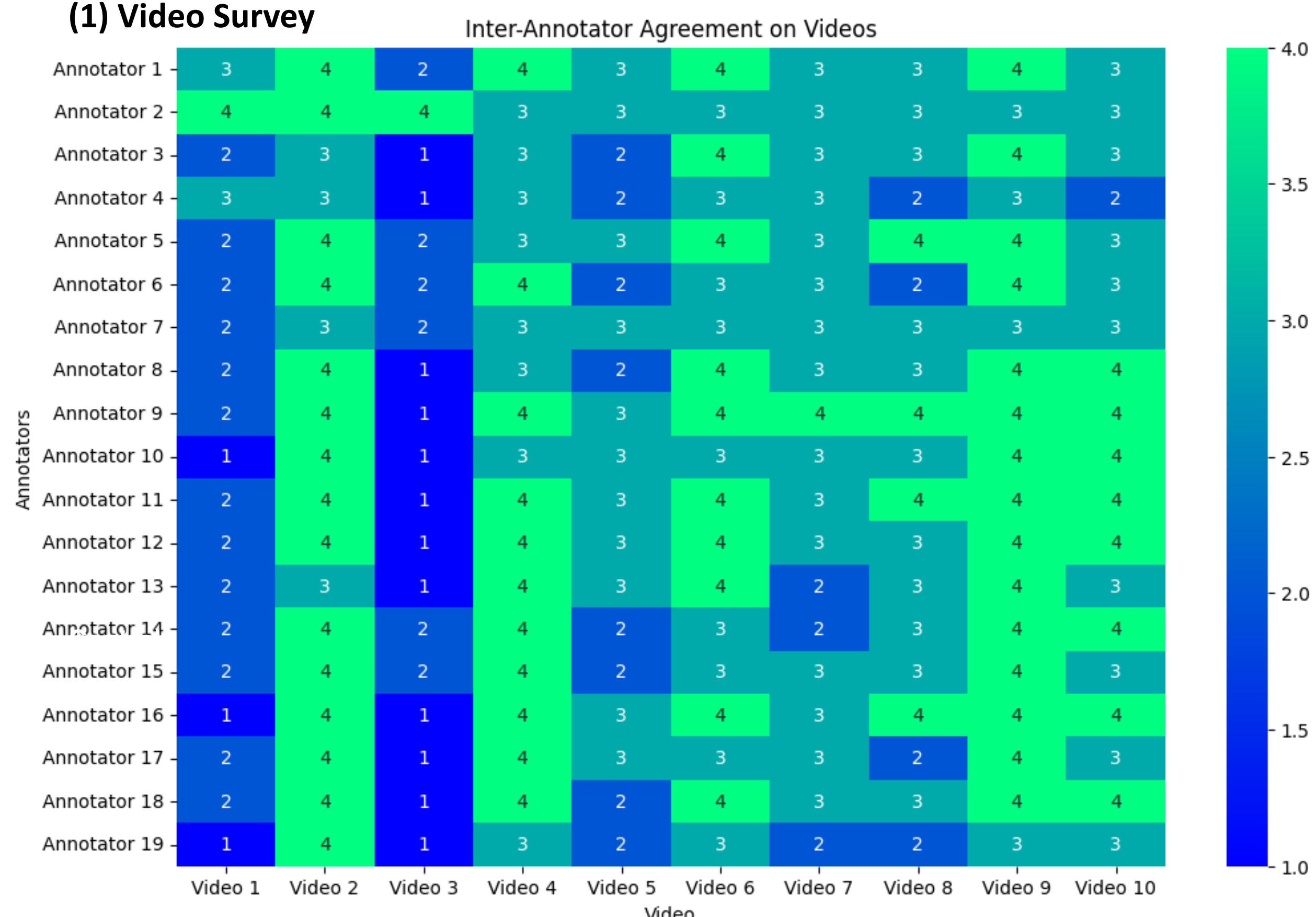


Figure One: Video Survey Results → Krippendorff's Alpha = 0.620 (0.232, 0.806)

(3) Adjudication & Analysis of Protocol

1. Overall FBRS Score Analysis

- Agreement with True Score vs IAA:
 - Cohen's Kappa, Quadratic w/ 95% CI
 - True Score vs...
 - Annotator #1/True: **0.560** (0.445, 0.663)
 - Annotator #2/True: **0.611** (0.483, 0.719)
 - Annotator #3/True: **0.592** (0.453, 0.726)
 - IAA:
 - Annotator#1/#2: **0.773** (0.667, 0.857)
 - Annotator#1/#3: **0.820** (0.756, 0.873)
 - Annotator #2/#3: **0.738** (0.621, 0.825)
- Krippendorff's alpha w/ 95% CI – all 3 annotators: **0.801** (0.747, 0.845)

2. Entity Tagging Analysis

GLS Regression Results

| Dep. Variable: | Overall | R-squared: | 0.644 |
|-------------------|------------------|---------------------|----------|
| Model: | GLS | Adj. R-squared: | 0.629 |
| Method: | Least Squares | F-statistic: | 42.54 |
| Date: | Tue, 09 Apr 2024 | Prob (F-statistic): | 2.65e-38 |
| Time: | 14:59:14 | Log-likelihood: | -211.92 |
| No. Observations: | 197 | AIC: | 441.8 |
| Df Residuals: | 188 | BIC: | 471.4 |
| Df Model: | 8 | | |
| Covariance Type: | nonrobust | | |

| | coef | std err | t | P> t | [0.025 | 0.975] |
|-----------|---------|---------|--------|-------|--------|--------|
| Intercept | 1.1805 | 0.113 | 10.440 | 0.000 | 0.957 | 1.404 |
| Temp_Neg | -0.3351 | 0.070 | -4.808 | 0.000 | -0.473 | -0.198 |
| Temp_Pos | 0.2847 | 0.088 | 3.242 | 0.001 | 0.111 | 0.458 |
| Act_Neg | -0.5112 | 0.053 | -9.708 | 0.000 | -0.615 | -0.407 |
| Act_Pos | -0.0277 | 0.109 | -0.253 | 0.800 | -0.243 | 0.188 |
| Coop_Neg | -0.4232 | 0.130 | -3.255 | 0.001 | -0.680 | -0.167 |
| Coop_Pos | 0.2479 | 0.065 | 3.798 | 0.000 | 0.119 | 0.377 |
| BMT_Neg | -0.3759 | 0.103 | -3.634 | 0.000 | -0.580 | -0.172 |
| BMT_Pos | -0.1461 | 0.060 | -2.421 | 0.016 | -0.265 | -0.027 |

Table Three: Generalized Least Squares Regression

(2) Manual Annotation

1 **C0 cooperative** and **T0 excited** to be here but VERY **A0 Active**, **A2 moves** around a lot, **A1 strong gag reflex**. Does **C1 really well** with continous **B0 distraction** and **B1 firm rule setting**.

| Tag | ID | Spans | Text | Attributes |
|-------------|----|-------|-------------------|------------------------------------|
| Temperament | T0 | AB | excited | Timing: NA, Score: + |
| Action | A0 | AB | Active | Timing: During Procedure, Score: - |
| Action | A1 | AB | strong gag reflex | Timing: During Procedure, Score: - |
| Action | A2 | B | moves | Timing: NA, Score: - |

Figure Two: Example of Entity Tagging on MedTator

Krippendorff's Alpha:

Standard metric for multiple annotators annotating multiple items.

$$\alpha = 1 - \frac{\text{Disagreement Observed } (D_o)}{\text{Disagreement Expected } (D_e)}$$

Alpha = 1, indicates perfect reliability

Alpha = 0, indicates no reliability (random)

Cohen's Kappa, Quadratic:

Measures IAA among two annotators

$$\text{Equation: } \kappa = \frac{p_o - p_e}{1 - p_e}$$

Where p_o is the observed probability of agreement

And p_e is the expected agreement when assigning labels randomly

F1 Score:

Harmonic mean of precision and recall

Confusion matrix between annotators:

$$\text{Equation: } F_1 = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} = \frac{2TP}{2TP + FP + FN}$$

F1=1 means perfect matching

F1=0 means not matching

| Cohen's Kappa | Interpretation |
|---------------|------------------------|
| 0 | No agreement |
| 0.10 - 0.20 | Slight agreement |
| 0.21 - 0.40 | Fair agreement |
| 0.41 - 0.60 | Moderate agreement |
| 0.61 - 0.80 | Substantial agreement |
| 0.81 - 0.99 | Near perfect agreement |
| 1 | Perfect agreement |

Table One: Cohen's Kappa Interpretation of Data³

Table Two: Entity Tagging Analysis → Cohen's Kappa & F1 Score

| Tag | Annotator1 | Annotator2 | 5-point Scale | | 3-point Scale | |
|------------------|--------------|--------------|---------------|--------|---------------|--------|
| | | | Kappa | F1 | Kappa | F1 |
| Temperament | Annotator #1 | Annotator #2 | 0.7082 | 0.6216 | 0.9439 | 0.9730 |
| | Annotator #1 | Annotator #3 | 0.6624 | 0.6613 | 0.7819 | 0.9355 |
| | Annotator #2 | Annotator #3 | 0.8170 | 0.6324 | 0.8517 | 0.9265 |
| Action | Annotator #1 | Annotator #2 | 0.6853 | 0.7000 | 0.8805 | 0.9833 |
| | Annotator #1 | Annotator #3 | 0.8825 | 0.8704 | 1.0000 | 1.0000 |
| | Annotator #2 | Annotator #3 | 0.7603 | 0.7611 | 0.8825 | 0.9735 |
| Cooperation | Annotator #1 | Annotator #2 | 0.8358 | 0.8406 | 1.0000 | 1.0000 |
| | Annotator #1 | Annotator #3 | 0.6531 | 0.6863 | 0.8517 | 0.9608 |
| | Annotator #2 | Annotator #3 | 0.5622 | 0.7681 | 0.6477 | 0.9420 |
| Overall Takeaway | Annotator #1 | Annotator #2 | 0.0676 | 0.8148 | 0.0000 | 0.9907 |
| | Annotator #1 | Annotator #3 | 0.2188 | 0.5732 | 1.0000 | 1.0000 |
| | Annotator #2 | Annotator #3 | 0.2744 | 0.5094 | 0.1733 | 0.9717 |

DISCUSSION & CONCLUSION

• Video Survey

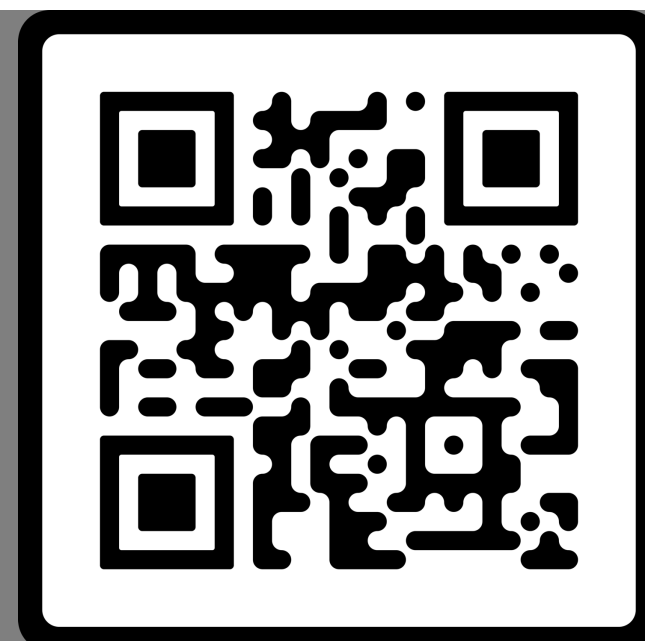
- Krippendorff's Alpha score of ~0.6 suggests that dental faculty & resident's provide moderate reliability to identify the true FBRS score when provided video footage of the encounter.
- Moreover, the lack of statistical significance (large confidence interval) suggests that providers largely do not agree with one another when evaluating dental behavior.

• MedTator Analysis

- Overall FBRS Score:
 - A Cohen's Kappa average score of ~0.588 suggests moderate agreement for annotators to correctly identify the true FBRS score when provided the behavior comment.
 - However, the IAA suggests strong reliability and substantial agreement among annotators.
 - Cohen's Kappa, Quadratic: 0.777
 - Krippendorff's alpha: 0.801
- Entity Tagging:
 - When asked to provide entity tags and their associated scores, annotators displayed substantial agreement on a 5-point scale and near perfect agreement on a 3-point scale.
 - The adjudicated entity tags (Exception "Act_Pos") are statistically significant on a $p < 0.05$ and majority of tags are statistically significant on a $p < 0.001$.
 - Confidence that the variable is correlated with the response variable.

Limited quantifiable data exists to support the continued use of the FBRS. The issue may not be the fault of the scale itself but rather how providers are currently applying it. This study describes an alternative method of using NLP to supplement the FBRS to create a more objective, accurate, and consistent scale. More research needs to be completed to substantiate this claim.

Please Scan To
Access References



SCAN ME