Are You Sleeping? Accounting for Naptime in Long-Form Recordings of Bilingual Preschoolers

Sarah Surrain[1,2], Kelly Vaughn[2], Yarián García Alayón[2], Kimberli Izaguirre[2], & Kevin Rosales[2]

[1]University of Nevada, Las Vegas; [2]University of Texas Health Science Center at Houston

Why exclude sleep in long-form recordings?

- Removing periods of sleep should…
  - Increase the signal-to-noise ratio
  - Reduce measurement error
  - Yield stronger correlations with child language
- But listening to full recordings is rarely feasible, and parent reports of naps can be incomplete
- Identifying sleep in recordings of preschool-aged (3-5 year-old) children can be especially challenging
  - Large variation in napping rates and routines
  - Napping can happen in a variety of settings
  - More independent - periods of silence not always sleep
- **RQ: Do different approaches to excluding sleep yield different associations with child expressive vocabulary skills?**

132 Spanish-speaking families in Houston, Texas

- **Families**
  - Recruited from a larger IES-funded intervention study
  - Children in home: M = 2.5
  - Adults in home: M = 2.6
- **Parents**
  - 95% mothers
  - Median parent education high school degree
  - 85.5% educated in Latin America
  - 85% use mostly or only Spanish with child
- **Children**
  - 58% girls
  - Age 3.5 to 5 (M = 4.7 years)
- **Measure of child language:**
  - EOWPVT-SBE Conceptual Standard Score (*M*=105, *SD*=16)
  - Spanish & English tested on different days, credit for all items correct in at least 1 language (missing = 9)
- **Long-form recording procedure:**
  - Recorded for at least 8 hours on a non-school day
  - Completed a 1-page log
- Only baseline data used for this analysis

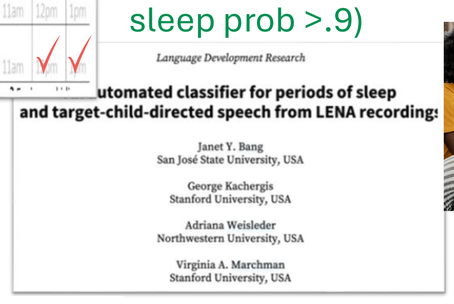
Comparing 3 efficient & 1 inefficient approach

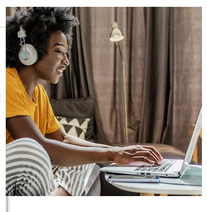
1. Include the full recording

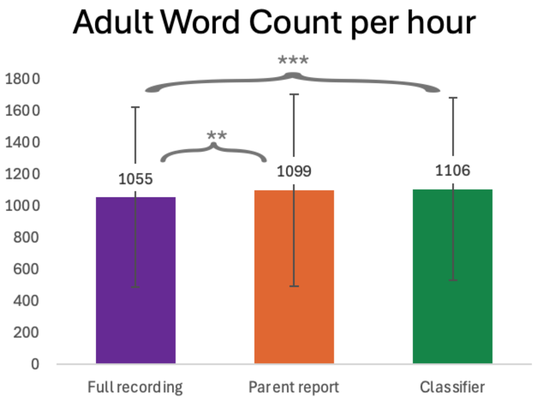
2. Rely on parent reports

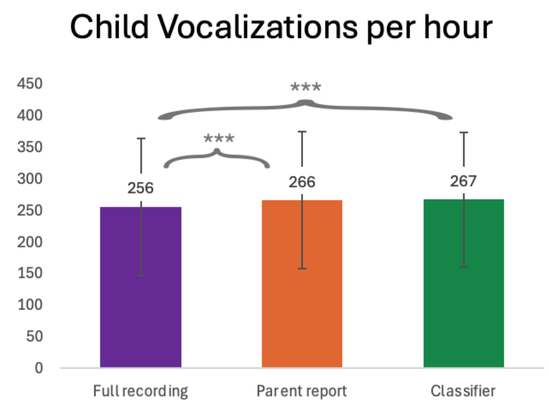
3. Use an automated sleep classifier tool (remove 5 min. segments with sleep prob >.9)

4.Machine+ Human (Human listeners verify all sleep boundaries)



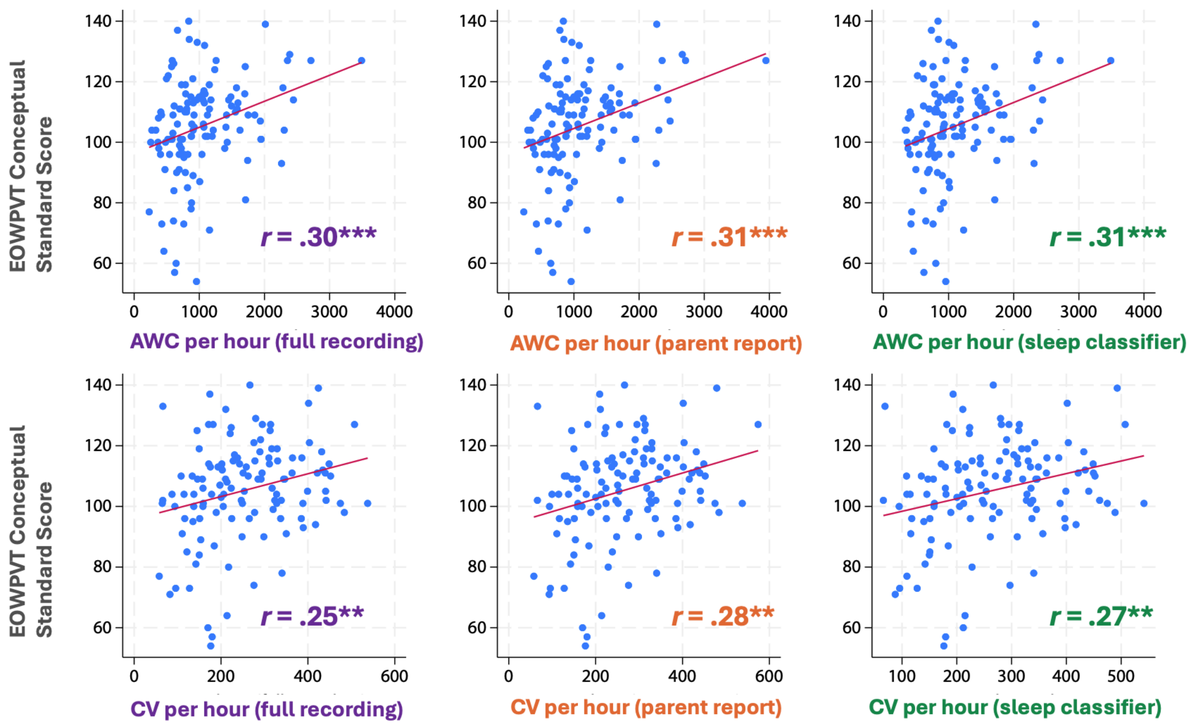






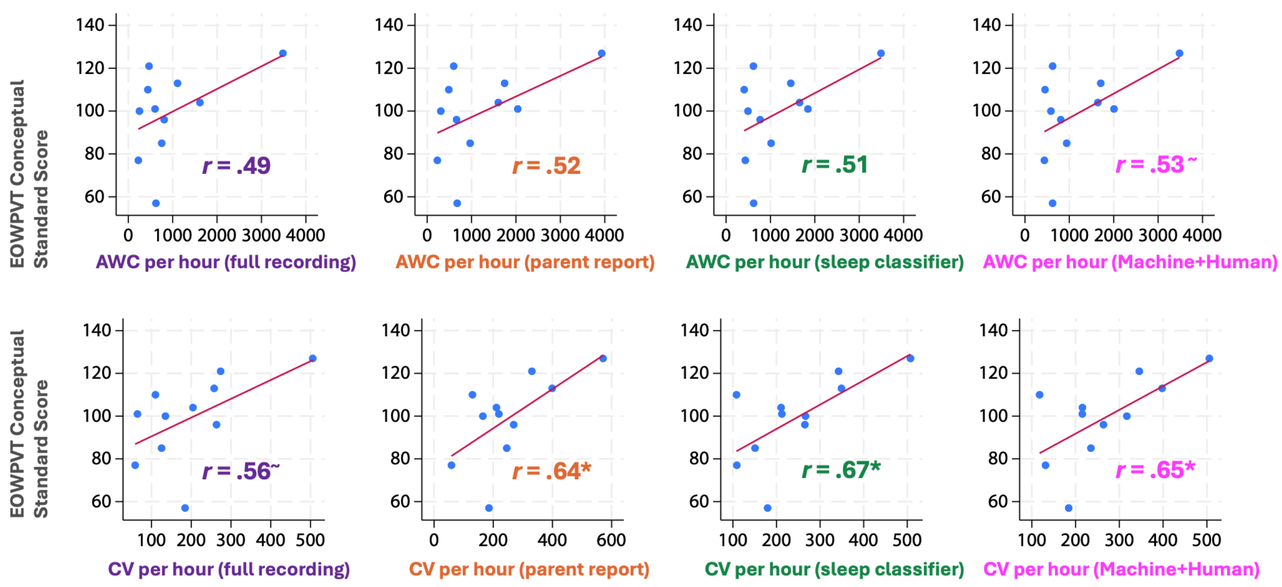
All estimates correlated > .96

3 Efficient Approaches: All Recordings (*n*=123)



- All yielded significant, positive correlations with child language
- No significant differences among among approaches (Steiger's *Z*'s > |.29|, *p*'s > .21)
- Regression models using **parent report** & **sleep classifier** estimates of AWC and CV explained more variability in child language and both indicators retained significance

4 Approaches: Recordings Containing Sleep (*n*=11)



- **Machine+human** approach used with 11 recordings containing parent-reported sleep and permission to listen
- **Parent report**, **sleep classifier** & **machine+human** yielded slightly stronger correlations than **full recordings**
- Under-powered to test for significant differences

Discussion & Next Steps

- In long-form recordings of 3-5 year-old bilingual children
  - Measures of adult and child talk were robustly associated with child language skills
  - Removing sleep led to slightly stronger correlations and explained more variability in language skills, regardless of approach
- Implies that we can reduce burden on coders and parents
  - Time-intensive human coding may not be needed if only hourly averages are needed
  - If parent reports are not feasible or available, the automated sleep classifier can be used instead (Bang et al., 2023)
- Limitations
  - Findings may be specific to preschool-aged children who no longer nap regularly (<15% reported daytime nap)
  - Only a small subset of recordings were human-coded
- Future directions: Identifying which languages are being spoken and extracting features of language environment that support the home language in bilingual children

Sarah.Surrain@unlv.edu