1. Training on random sample of all Mumsnet UK parenting and feminist forum users in our sample
⇨ Trying to capture as many kinds of parents and feminists as possible with forum data

2. Testing on random posts from users who posted in BOTH Mumsnet forums (parent & feminist)
⇨ Controlling for demographics, personality and other stable characteristics

3. Testing on random posts from users who posted in Reddit r/feminist AND r/parent
⇨ Ensuring that it's not a "Mumsnet thing" and testing on different set of demographics (here: US rather than UK, not restricted to middle class)

ASIA Validation Pathway

1. Ethical considerations
2. Selection of the training dataset
3. Quantifying stylistic features from text
4. Training the model
5. Cross-validating the model on within-data
6. Cross-platform validation
7. **Experimental validation**
8. Concurrent validation

Experimental validation

Why do we need to validate the model experimentally?

Assumptions:

- Forums = active social identities
- People who write in both forums hold both identities
- Different topics do not cause the differences
- Different audiences do not cause differences

Experimental validation

**1) Audience:** via information sheet – researchers are the only audience

2) Participants: self-report that they hold BOTH identities of interest (yes/no)

3) Manipulation of social identity:

- Randomly assign to Group A or Group B

- 3 things task (Haslam et al., 1997)

4) Topic: Ask participants to write 5-10 sentences about a topic/several topics (e.g., relevant to each identity, "neutral" topic)

5) Analysis: Use forum-trained & -validated model to calculate AUCs etc.
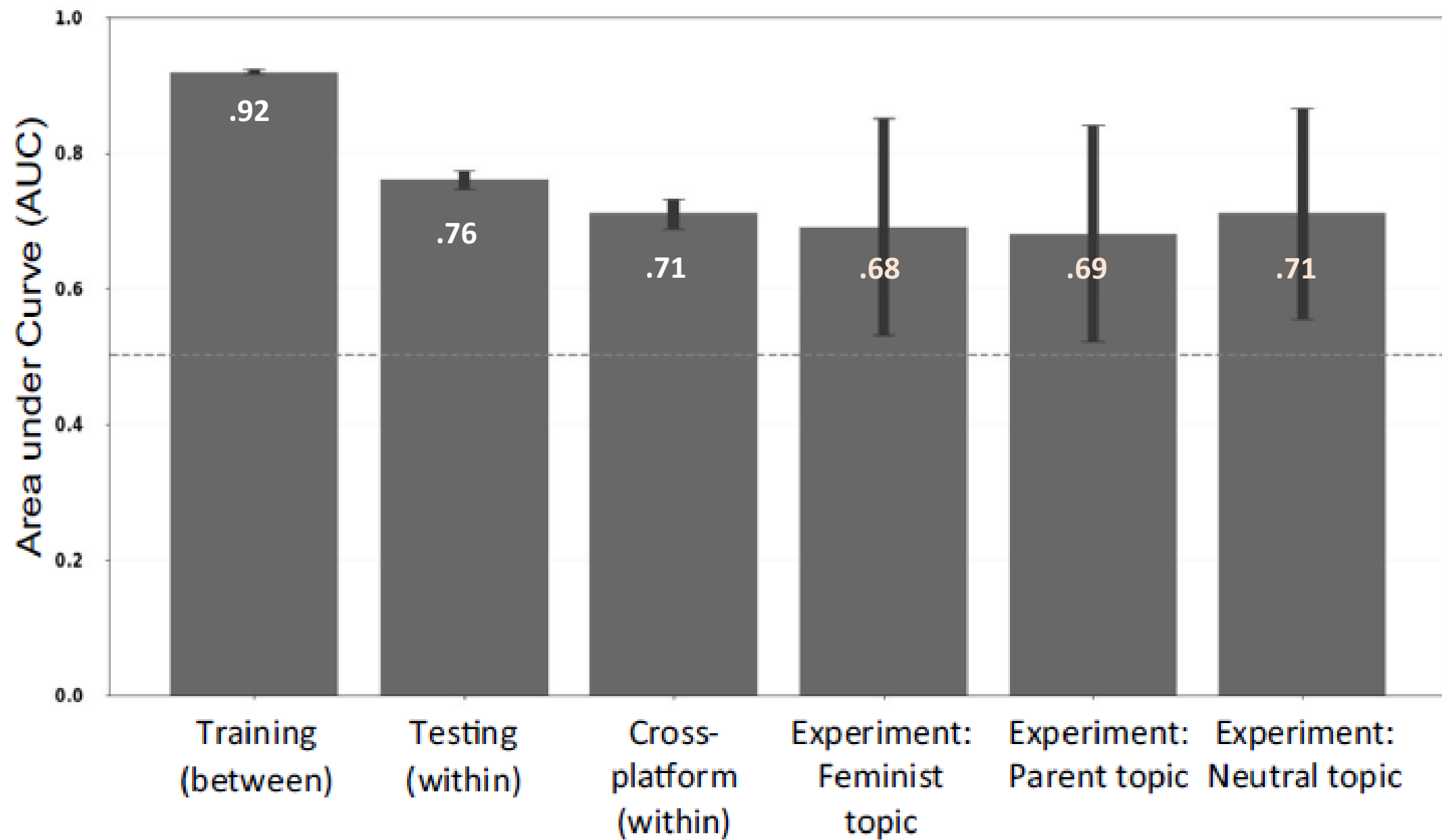
ASIA Validation Pathway

1. Ethical considerations
2. Selection of the training dataset
3. Quantifying stylistic features from text
4. Training the model
5. Cross-validating the model on within-data
6. Cross-platform validation
7. Experimental validation
8. **Concurrent validation**

Concurrent validity

Can we use the classifier to predict outcomes in a way that would be expected from theory/other research?

Example: Parent identity & postnatal depression

- PND affects 10-20% of parents

- Usually diagnosed 3mth after birth

- Cross-sectional: negative effect of parent identity on PND (Seymour-Smith et al., 2017)

Hypothesis: First-time mums with PND should have significantly lower active parent identity than those without PND (when they write in a context that activates a parent identity!)

Concurrent validity
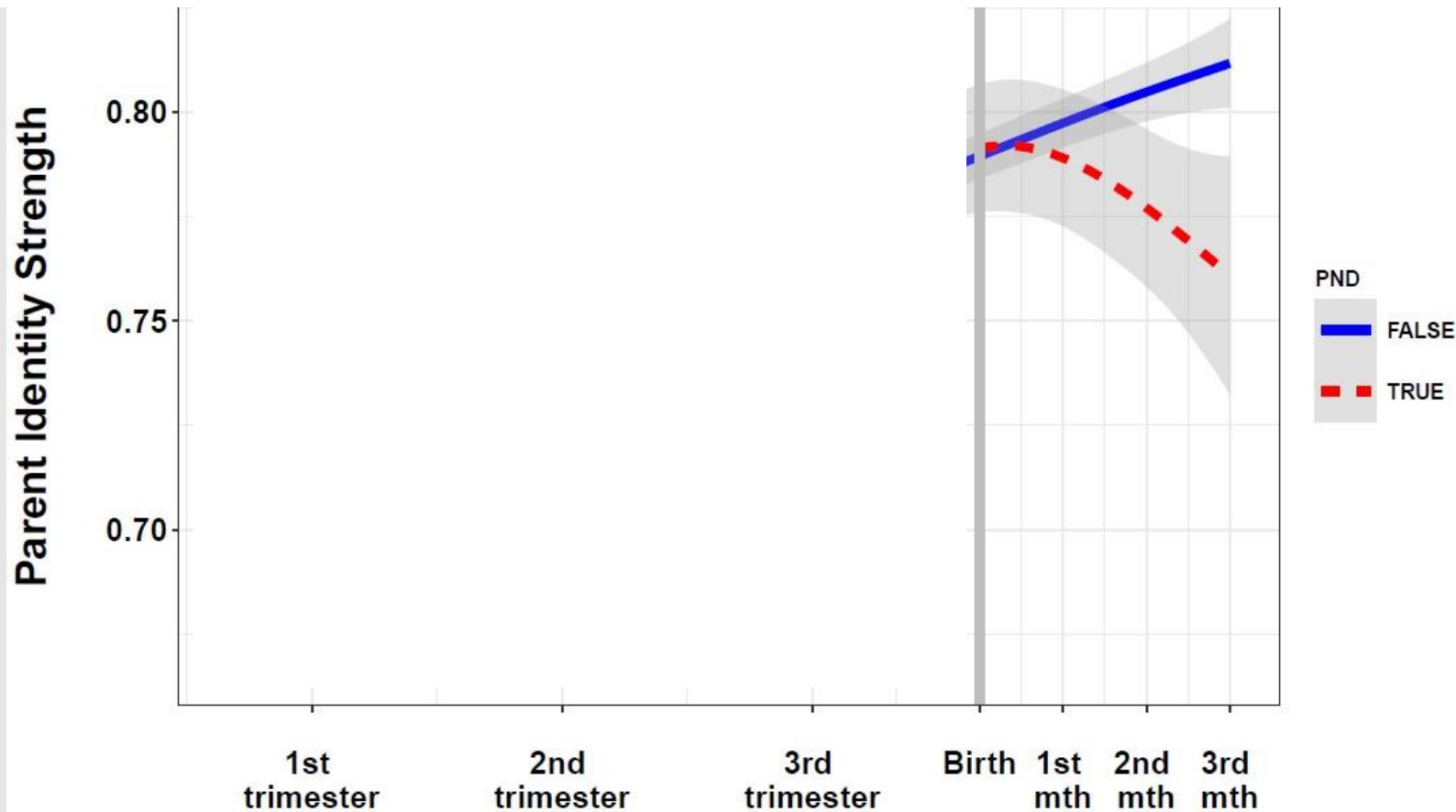
Netmums UK data

N = 326 first-time mothers
(55 with PND, 17%)

Growth curve model:

Post-natal linear effect for PND:
b=-0.06, *SE*=0.03, *p*=.019