Coding Speech Samples: Combining Qualitative and Quantitative Methods
Jessica Pleuss, Amy Susman-Stillman, Anna Shkolnik
Center for Early Education and Development, University of Minnesota

Introduction & Background

• Quantitative data analysis methods allow for advanced statistical analyses but can fail to capture the complexity of data.
• Qualitative data analysis methods remain true to the complex nature of data, but can fail to provide a uniform basis for comparing data and drawing conclusions that generalize beyond a sample.
• In this study, we address this problem by creating a coding system that capitalizes on positive elements of both qualitative and quantitative methods.

Ounce Research Project (ORP)

• A randomized control intervention study examining use of an observational assessment tool—The Ounce Scale™ (Meisels, Marsden, Dombro, Weston, & Jewkes, 2003)—with child care providers and parents of infants and toddlers
  • It was hypothesized that using The Ounce Scale™ would result in a deeper and more complex understanding of young children’s development, which might be expressed when a parent or provider describes an individual child.
  • We developed a new coding system to measure this, via speech samples collected from the parents and providers.

Speech Sample Data

• Semi-structured phone interviews
  – Data collected from 98 child care providers at (up to) 3 timepoints each 6 months apart (baseline, post-training, and post-implementation)
  – Asked to speak for 5 min about an infant or toddler in their care so that we could “get to know” child
  – Standard prompts used as needed, e.g. “If you were to have a substitute caregiver for the day, what would you want him/her to know about (name)?” Tell us enough so that a substitute caregiver would feel like she/he knows the child.
• These results based on a subsample of 102 provider speech samples (34, 35, & 33 at Time 1, 2, and 3 respectively)

Coding System Overview

• Our coding scheme assesses the speaker’s complexity of thought regarding the child
  – reveals speaker’s understanding and interpretation of child behavior
  – adopts the same basic structure of the system used by Thomas and Englund (1996), adapted to reflect the content of the ORP speech samples
• Transcripts are divided into “thought units” which are coded:
  – Within a category (child description, adult practices, goals, priorities)
  – Within a level (1-3: less to more abstract/developmentally appropriate)
• Relating one thought unit to another is coded as a “connection”: either:
  – Within category
  – Across category
• Combination of qualitative & quantitative methods yields both scores and a chart which diagrams the speech sample

The Coding System (Examples)

<table>
<thead>
<tr>
<th>Thoughts</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Child behavior</td>
<td>&quot;The child looks at the toy and reaches for it.&quot;</td>
</tr>
<tr>
<td>Thought units are connected</td>
<td>&quot;The child grabs the toy and brings it to its mouth.&quot;</td>
</tr>
<tr>
<td>Child behavior</td>
<td>&quot;The child looks at the toy and reaches for it.&quot;</td>
</tr>
<tr>
<td>Thought units are connected</td>
<td>&quot;The child grabs the toy and brings it to its mouth.&quot;</td>
</tr>
</tbody>
</table>

Quantitative Coding: Scores

• Scores are given for each thought unit and each connection
  – Connections are weighted, so that the scores reflect complexity of thought rather than amount
  – Within category = 5
  – Across category = 10
  – Subtotals and totals are calculated

Qualitative Coding: Charts

• Entire speech sample is charted to create a visual representation which shows at a glance the nature and complexity of thoughts expressed
  – Each thought unit is represented by a dot (●)
  – Each connection is represented by a line or “link” connecting the corresponding thought units (—)

Achieving Reliability

• Inter-rater reliability was assessed on 15 of 112 cases, using:
  – Pearson correlations (do scores vary together?)
  – paired sample t-tests (are scores different between coders?)
• Coders achieved adequate reliability on total scores and the majority of subscale scores

Are Scores Useful?

• Yes, charts indicate at a glance:
  – Types of thoughts expressed (categories)
  – Depth of thought expressed (levels)
  – Complexity of thought expressed (connections)

Are Charts Interpretable/meaningful?

• Yes, charts indicate at a glance:
  – Types of thoughts expressed (categories)
  – Depth of thought expressed (levels)
  – Complexity of thought expressed (connections)

Sensitivity: Change Over Time?

• Individual scores do show variation over time, suggesting that providers’ interpretation and understanding of child behavior is not static

Overview of Categories, Levels Used to Code Speech Samples

<table>
<thead>
<tr>
<th>Categories &amp; Levels Used to Code Speech Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Levels</td>
</tr>
<tr>
<td>--------</td>
</tr>
<tr>
<td>3 (conceptual, more dev. appr.)</td>
</tr>
</tbody>
</table>

Distribution of Thought Unit Subtotal Scores

<table>
<thead>
<tr>
<th>Distribution of Thought Unit Subtotal Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>One provider’s change over time</td>
</tr>
</tbody>
</table>

Distribution of Connections Subtotal Scores

<table>
<thead>
<tr>
<th>Distribution of Connections Subtotal Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>One provider’s change over time</td>
</tr>
</tbody>
</table>

Are Scores Useful?

• Yes, all scores except ‘goals,’ ‘priorities,’ and ‘within category,’ show approximately normal distributions with wide variation
• Total scores ranged from 23 to 196
• As planned, the ‘across category’ score heavily influences the total score, ensuring that it is the complexity of thought, not merely the amount of talking, that is measured