The Development of a Thin Slice Methodology for Coding Scaffolding Between Siblings

by

Heather Prime

A thesis submitted in conformity with the requirements for the degree of Master of Arts in School and Clinical Child Psychology Graduate Department of Applied Psychology and Human Development University of Toronto

© Copyright by Heather Prime (2012)
The Development of a Thin Slice Methodology for Coding Scaffolding Between Siblings

Heather Prime
Master of Arts
Applied Psychology and Human Development
University of Toronto
2012

Abstract

The goal of the present study was to develop and compare two different methods for rating scaffolding between siblings: a thin slice approach and an interval coding approach. Fifty younger (age=3 years) and 50 older (age 3-7 years) siblings interacted for five minutes on a cooperation task and scaffolding during the task was coded for each child. Internal consistency was excellent for the thin slice measure and questionable for the interval measure. Inter-rater reliability was good for both. Thin-slicing was more strongly related to predicted variables (children’s theory of mind, language, age, cooperation, positive and negative behavior) than interval coding, and reduces demands on resources in terms of training and reliability. The development of a reliable and valid measurement for the assessment of child-to-child scaffolding, which involves limited training and is quick to code, will be a useful research and practice tool for developing children’s cooperation skills in applied settings.
Acknowledgments

I would like to thank the Jenkins lab for creating an environment full of laughter and productivity, and for keeping me moving forward at all times. I would like to give a special thank you to my supervisor, Jenny Jenkins, who has accompanied her high expectations with warmth and support.
# Table of Contents

Acknowledgments ........................................................................................................ iii

Table of Contents ........................................................................................................ iv

List of Tables ................................................................................................................ v

1 Background ................................................................................................................ 1
   1.1 Scaffolding in Siblings ..................................................................................... 1
   1.2 Comparing Two Methods .............................................................................. 3

2 Methodology .............................................................................................................. 5
   2.1 Participants ....................................................................................................... 5
   2.2 Procedure ........................................................................................................ 5
   2.3 Measures .......................................................................................................... 6
      2.3.1 Scaffolding ............................................................................................... 6
      2.3.2 Age ........................................................................................................... 7
      2.3.3 Language ................................................................................................. 7
      2.3.4 Theory of Mind ....................................................................................... 9
      2.3.5 Cooperation ............................................................................................ 9
      2.3.6 Positivity ................................................................................................. 10
      2.3.7 Negativity .............................................................................................. 10
   2.4 Results ............................................................................................................. 10
      2.4.1 Comparing the Efficacy of Thin Slice Versus Interval Coding ............. 10
      2.4.2 The Value of Two Coders Versus One ................................................. 13
   2.5 Discussion ....................................................................................................... 14

References ................................................................................................................. 17
List of Tables

Table 1. List of thin-slice and interval items, with respective wording.................................................. 6

Table 2 Inter-rater Pearson correlations for thin-slice coder 1 and thin-slice coder 2 for all items 8

Table 3. Means and standard deviations (SD) of thin-slice and interval coding measures, and inter-measure Pearson correlations........................................................................................................... 11

Table 4. Concurrent validity for thin-slice and interval composites with child-specific measures
........................................................................................................................................................................ 13
1 Background

1.1 Scaffolding in Siblings

Scaffolding refers to the idea that given appropriate assistance, a learner can reach a goal or function at a level that would be otherwise unattainable. Related to this concept is Vygotsky’s (1978) description of the zone of proximal development, the distance between a child’s developmental level and the level of potential development possible with guidance or support. Learning is optimal when it occurs in this zone. By continuously monitoring a learner’s level of understanding, a more skilled interactional partner can adjust support accordingly (Stone, 1998; Strauss, Ziv, & Stein, 2002). Over time, partners gradually adapt their scaffolding to encourage the learner to take more responsibility in the task until they are able to complete it independently (Wertsch, McNamee, McLane, & Budwig, 1980). The goal of the present study was to develop methods for assessing scaffolding between siblings and to compare the efficacy of two different methods of assessment.

Parent and teacher scaffolding of children has been shown to facilitate social and cognitive development (Landry, Miller-Loncar, Smith, & Swank, 2002; Landry, Swank, Smith, Assel, & Gunnewig, 2006; Bibok, Carpendale, & Muller, 2009). For example, maternal verbal scaffolding has been linked to children’s verbal and non-verbal cognitive skills (Smith, Landry, & Swank, 2000), play skills (Belsky, 1980), as well as early language and later reading skills (Dieterich, Assel, Swank, Smith, & Landry, 2005; Tamis-LeMonda, Bornstein, Kahana-Kalman, Baumwell, & Cyphers, 1998; Tamis-LeMonda, Bornstein, & Baumwell, 2001). Although scaffolding is a concept most commonly related to parents and teachers, children take on the teaching role for their siblings and peers as well (Azmitia & Hesser, 1993), showing individual differences in their level of instruction, the strategies they employ, and the approaches they take (Recchia & Howe, 2009). Siblings, in particular, have a unique relationship characterized by high levels of intimacy, emotionality, and familiarity (Dunn, 1983). As such, they are afforded many shared reciprocal (i.e. play; Dunn, 1983) and complementary (i.e. teaching; Azmitia & Hesser, 1993) interactions, both of which heighten opportunities for influencing one another’s sociocognitive development. That siblings influence one another’s development is supported by the positive association between family size and some aspects of cognitive development. For example, children with more siblings show higher levels of theory of mind development (ToM; Jenkins &
Asington, 1996; Perner, Ruffman, & Leekam, 1994; Peterson, 2000). Research that has directly studied teaching in siblings reveals that younger children are more likely to solicit teaching from older siblings as opposed to older peers, and that they are more active participants in the process (Azmitia & Hesser, 1993). Similarly, older siblings are more likely to provide explanations and feedback, and they spontaneously provide guidance and explanations more often than do older peers (Azmitia & Hesser, 1993). These results suggest a unique scaffolding process for siblings.

Although the term scaffolding usually refers to a more knowledgeable person helping a less knowledgeable person, we construed it more broadly in terms of a capacity to engage with the other’s mind with the goal of knowing and changing that mind. We hypothesized three critical elements of this skill in the preschool and early school aged period: being motivated towards ‘mutuality’, being able to represent internal states of the self and the other, and possessing linguistic skills that allow for negotiation of these internal states. Given the development of all of these capacities in the preschool period we expected older siblings to be more adept at scaffolding than younger siblings, but for nascent versions of this skill to be evident in younger children. Consequently, our goal was to develop a measure that was equally applicable to younger and older children. Given that scaffolding is in part about engaging with a partner’s mind, with the goal of knowing and changing that mind, we expected our measure of scaffolding to be associated with several aspects of socio-cognitive development that emerge in the preschool period. Cooperation, thought to be unique to humans (Tomasello, 2009), emerges around 18 months (Warneken & Tomasello, 2007). It involves the sharing of intentions and the coordination of behavior to achieve a joint goal (Brownell & Carriger, 1990; Warneken & Tomasello, 2007). As this involves being motivated towards ‘mutuality’, as well as the coordination of two people’s actions towards a joint goal, we expected it to be linked to children’s scaffolding.

Theory of mind refers to our ability to attribute mental states to other people. Understanding simple desires appears to develop around the age of 2, belief at 3 to 4, and then more sophisticated aspects of theory of mind in middle childhood and beyond (Wellman & Woolley, 1990; Wimmer & Perner, 1983; Carpendale & Chandler, 1996). As theory of mind understanding involves representing the internal states of the self and the other, this skill should be linked to scaffolding. Indeed, research has revealed an association between ToM development and children’s teaching abilities (Strauss et al., 2002; Davis-Unger & Carlson, 2008). Children
showing a developed capacity to understand mental states (i.e. ToM) are better able to impart knowledge and respond to learners’ errors in a sensitive fashion (Davis-Unger & Carlson, 2008).

Further, given that scaffolding relies on our ability to diagnose what another person knows, we expected linguistic skills to be associated with scaffolding. When the scaffoldor can use specific words to find out what is known and not known, this is likely to strengthen their ability to scaffold.

Finally, it has been argued that the motivation towards mutuality is affectively based (see Oatley, Keltner and Jenkins, 2006). We are defining scaffolding as a drive towards mutuality and, as such, we would expect to see higher rates of interpersonal positivity and lower rates of interpersonal negativity from children who are good at scaffolding.

1.2 Comparing Two Methods

In addition to developing a sibling scaffolding measure we sought to compare two different methodologies in measuring this construct. Specifically, we wanted to examine the viability of thin slice coding in comparison to interval coding, with the intention of developing a measure that minimizes coding demands. Thin slice judgments have been defined as judgments based on brief observations of a person’s expressive behaviours (Ambady & Rosenthal, 1992). They are characterized by automatic (as opposed to controlled or deliberative) evaluations (Ambady, Bernieri, & Richeson, 2000). Ambady and Rosenthal (1992) conducted a meta-analysis across 38 studies, examining the predictive accuracy of judgments based on thin slices of behaviour. The analysis revealed an overall effect size of .39, suggesting “thin slices of behaviour provide a great deal of information and permit significantly accurate predictions” (Ambady & Rosenthal, 1992, pg 267). Thin-slicing has been used to predict job performance (Hecht & LaFrance, 1995), intellect (Murphy, Hall, & Colvin, 2003; Reynolds & Gifford, 2001), and relationship quality (Harrigan & Rosenthal, 1986; Bernieri, Gillis, Davis, & Grahe, 1996). Ambady & Rosenthal (1993) showed that thin slice judgments of teachers’ attentiveness, warmth, and enthusiasm had predictive value for subsequent teacher evaluations.

Given the evidence for the accuracy of thin slice judgments, the question of their utility in coding observational data in general, and scaffolding between siblings specifically, comes to the fore. Behavioural observation provides a means to collect ecologically valid data, reduces respondent
biases, and allows researchers to gather data concerning constructs that are not easily and validly assessed through self-report (Margolin et al., 1998). Despite these advantages, however, the process is highly resource intensive. The procedure of coding data is particular demanding in terms of time and money, and often presents additional obstacles surrounding attaining and maintaining reliability. Efforts to reduce the demands associated with coding of behavioural observation data are thus a priority in the present development of a sibling scaffolding measure.

Traditionally, behavior is quantified using rating scales (i.e. completing a Likert-type scale on target variables; e.g. Deater-Deckard, Pylas, & Petrill, 1997, Matias, Scott, & O’Connor, 2006), or with observational methods such as event-based coding (i.e. noting each time a codable event of interest occurs; e.g. Crockenberg, Leerkes, & Barrig Jo, 2008) or interval coding (i.e. dividing a period of time into a number of relatively brief intervals and noting which codable events occur; e.g. Howes & Matheson, 1992). These traditional methods typically require coders to use narrow guidelines to classify behavior into predefined behavioural codes. Additionally, reliability of observers is a major concern, requiring considerable training to ensure they will produce essentially similar records (Bakeman & Gottman, 1987). The reliability and validity of observational methods and rating scale coding has been well established over the years. They have been used to code children’s free-play behaviours (Coplan, Prakash, O’Neil, & Armer, 2004), temperament (Goldsmith & Rothbart, 1996), joint attention (Carpenter, Nagell, Tomasello, Butterworth, & Moore, 1998), and cooperation (Wameken, Chen, & Tomasello, 2006), as well as maternal behaviours such as autonomy support (Bernier, Carlson, & Whipple, 2010), positive and negative control (Deater-Deckard et al., 1997), and sensitivity (Matias et al., 2006).

Thin-slicing offers an alternative to traditional coding methods that has the potential to reduce the time it takes to code and maintain reliability. Thin slice coding requires coders to make first impression judgments based on brief clips of behavior. As thin slice judgments suffer when information is processed more deliberately (Ambady, 2010), coders are encouraged to rate quickly based on intuition. Coders are trained (albeit, to a lesser extent than other observational methods) so as to promote familiarity with the constructs and thus facilitate judgment accuracy (Ambady et al, 2000). Data are coded by two, or more, independent coders. With this method, judgments of the coders are combined to obtain a group mean judgment, increasing the reliability
and decreasing the impact of a single observer’s judgment (Ambady et al., 2000; Rosenthal, 1987).

In the current study 50 sibling pairs interacted for five minutes on a cognitively challenging building task and the extent of scaffolding during the task was coded for each child. We compared a thin slice coding method with an observational method (interval coding). We further considered the need to aggregate codes from two thin slice coders versus using only one. If two coders achieve an acceptable level of agreement, and validity is not compromised, then, arguably, aggregation is not warranted.

2 Methodology

2.1 Participants

Subjects come from the Kids, Families, and Places (KFP) study, which is a birth-cohort longitudinal study that examines children’s development in a nested, multilevel framework. Families were recruited through a program called Healthy Babies Healthy Children, run by Toronto and Hamilton Public Health, which contacts the parents of all newborn babies within several days of the newborn’s birth. Criteria for participation included 2 children where one child was a newborn singleton (Sibling 1) and the other was born within four years of the newborn (Sibling 2). Additional inclusion criteria included the mother speaking English and the birth weight of the newborn was > 1500 grams. 501 families met criteria for the study. Scaffolding was assessed at Wave 3. Families used in the present study were a subset of the KFP sample, selected on a chronological basis. Data from 50 sibling pairs (N=100 children, 54 male and 46 female) and their families are described. Sibling 1 was approximately 3 years old (M=37.2 months, SD=1.9) and Sibling 2 was between 4 to 7 years old (M=65.8 months, SD=9.2).

2.2 Procedure

Data from time two (T2) and three (T3) was used for the current study and included questionnaire measures from mothers, child testing, and videos of mother-child and sibling interactions.

Sibling pairs were filmed engaging in a cooperative building task (Aguilar, O'Brien, August, Aoun, & Hektner, 2001). The camera was set up to film the faces, upper bodies and arms of the
children. Sibling 1 and Sibling 2 were instructed to sit on a “magic carpet” (yoga mat) and use Duplo building blocks to build a picture of a design together. They were told they only have 5 minutes to build it. Additionally, they were each only allowed to touch two colours (for example, Sibling 1 could touch light green and light blue blocks and Sibling 2 could touch dark green and dark blue blocks). To ensure understanding they were asked to point to the pieces they could each touch. If children finished the design before the end of 5 minutes, there were given a second model to build. All children were stopped after 5 minutes of the task, regardless of completion.

2.3 Measures

2.3.1 Scaffolding

The original ten scaffolding items can be seen in Table 1. Items were generated by investigators, based on the constructs that were outlined in the introduction. Items were based on the same construct for both methods but were worded to facilitate coding on the basis of an impression for the thin slice coding. For instance, the interval coding item ‘Positive Feedback’ became ‘This person gives positive feedback to reinforce his/her partner’ in the thin slice methodology.

<table>
<thead>
<tr>
<th>Item</th>
<th>Thin-slice</th>
<th>Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>This person gives ambiguous verbal directions.</td>
<td>Ambiguous verbal direction</td>
</tr>
<tr>
<td>2</td>
<td>This person gives clear and specific verbal directions.</td>
<td>Specific verbal direction</td>
</tr>
<tr>
<td>3</td>
<td>This person gives positive nonverbal directions.</td>
<td>Positive Non-Verbal Direction</td>
</tr>
<tr>
<td>4</td>
<td>This person reminds his/her partner about goals / rules of the task.</td>
<td>Meta Level Direction</td>
</tr>
<tr>
<td>5</td>
<td>When this person needs guidance, he/she is willing to ask for help.</td>
<td>Request for Help</td>
</tr>
<tr>
<td>6</td>
<td>This person is responsive to his/her partner’s request for help, even those that are subtle and/or nonverbal.</td>
<td>Response to a Request for Help</td>
</tr>
<tr>
<td>7</td>
<td>This person gives positive feedback to reinforce his/her partner.</td>
<td>Positive Feedback</td>
</tr>
<tr>
<td>8</td>
<td>If given a task, this person will try to complete it.</td>
<td>On task</td>
</tr>
<tr>
<td>9</td>
<td>If given a task, this person will try his/her best to follow the rules.</td>
<td>Non-compliance with rules (reversed for analyses)</td>
</tr>
<tr>
<td>10</td>
<td>This person reminds his/her partner when it’s his/her turn.</td>
<td>Turn Taking Directions</td>
</tr>
</tbody>
</table>

Table 1. List of thin-slice and interval items, with respective wording
2.3.1.1 Interval Coding

Two independent coders were trained in an interval coding strategy (drawn from work by Perlman & Ross, 1997; Howes & Matheson, 1992). Training involved joint coding as well as independent coding with subsequent discussion of discrepancies. Coders use narrowly defined descriptions to guide their coding. They were instructed to make notes and watch the clip as many times as necessary to ensure accurate coding. Each sibling interaction was coded in a series of 20s intervals and the presence or absence of each scaffolding item was noted within each interval. The proportion of episodes in which the behaviors occurred, for each child, are the final interval item scores. Reliability was checked at regular time points and final reliability on interval coding was high; kappas were .80 and .81 for Sibling 1 and Sibling 2, respectively, and percent agreement was above 85% for all items for both Sibling 1 and Sibling 2.

2.3.1.2 Thin-slice Coding

Two coders (different than the interval coders) were provided with brief, verbal descriptions of the scaffolding constructs they were coding and no additional training. They were directed to use all available information from the video and to rate quickly based on general impressions. The prompt for each of the 10 items was “Give your impression of how this person would interact with his/her partner on a day-to-day basis, based on what you have seen.” Coders watched the 5-minute film clip once per sibling pair and then each provided independent thin slice codes on a 5-point likert scale for each child, ranging from ‘Not at all true’ (1) to ‘Very true’ (5). Coders’ judgments on each item were combined to create mean judgments and used as the thin slice scores. Agreement between coders is presented in Table 2 and reviewed in the results section.

2.3.2 Age

Child’s age at testing, in months, as reported by the mother.

2.3.3 Language

The Peabody Picture Vocabulary Test (PPVT-III), a standardized test of receptive language skills (Dunn & Dunn, 1997), consists of 204 questions grouped into 17 sets of 12 items that are arranged in order of increasing difficulty. For each item, children are presented with 4 black and white pictures and asked to select the picture that best represents an orally presented word. The starting point of the test is determined by the child’s age. Item sets that are too easy or too
difficult for a child (as dictated by a standardized administration protocol) are not administered. The test yields one overall standardized summary score representing the child’s level of receptive language skills. The PPVT-III has an average correlation of .69 with the OWLS Listening Comprehension scale and .74 with the OWLS Oral Expression scale. Its correlations with measures of verbal ability are: .91 (WISC-III VIQ), .89 (KAIT Crystallized IQ), and .81 (KBIT Vocabulary).

<table>
<thead>
<tr>
<th>Scaffolding Code</th>
<th>S1</th>
<th>S2</th>
</tr>
</thead>
<tbody>
<tr>
<td>This person gives ambiguous verbal directions.</td>
<td>.07</td>
<td>.19</td>
</tr>
<tr>
<td>This person gives clear and specific verbal directions.</td>
<td>.26†</td>
<td>.24†</td>
</tr>
<tr>
<td>This person gives positive nonverbal directions.</td>
<td>.22</td>
<td>.38*</td>
</tr>
<tr>
<td>This person reminds his/her partner about goals / rules of the task.</td>
<td>.57**</td>
<td>.41**</td>
</tr>
<tr>
<td>When this person needs guidance, he/she is willing to ask for help.</td>
<td>.17</td>
<td>.32*</td>
</tr>
<tr>
<td>This person is responsive to his/her partner’s request for help, even those that are subtle and/or nonverbal.</td>
<td>.59**</td>
<td>.51**</td>
</tr>
<tr>
<td>This person gives positive feedback to reinforce his/her partner.</td>
<td>.31*</td>
<td>.63**</td>
</tr>
<tr>
<td>If given a task, this person will try to complete it.</td>
<td>.46**</td>
<td>.48**</td>
</tr>
<tr>
<td>If given a task, this person will try his/her best to follow the rules.</td>
<td>.69**</td>
<td>.47**</td>
</tr>
<tr>
<td>This person reminds his/her partner when it’s his/her turn.</td>
<td>.24†</td>
<td>.51**</td>
</tr>
</tbody>
</table>

†\textit{p} < .10; *\textit{p} < .05; **\textit{p} < .01

**Table 2 Inter-rater Pearson correlations for thin-slice coder 1 and thin-slice coder 2 for all items**
2.3.4 Theory of Mind

An adaptation of the scale described by Wellman & Liu (2004) was used. This scale presents various tasks in a sequential format that map closely onto the development of children’s theory-of-mind understanding. As children move through the scale, tasks become conceptually more difficult. Thus, progression further along the scale reflects more sophisticated theory-of-mind understanding. The first three tasks of the theory-of-mind scale assessed children’s understanding of diverse desires and beliefs, and knowledge and ignorance. This is followed by tasks that assessed more sophisticated theory-of-mind understanding such as false-belief, belief-based emotion, and real-apparent emotion. At the end of the scale, we added a second order belief question (Astington, Pelletier, Homer, 2002). If children failed two consecutive tasks on the theory-of-mind scale, testing was stopped. For all theory-of-mind tasks, stories were enacted for children with the use of puppets and props (or pictures, i.e. 2nd order false belief). For each of the tasks, the child is given a score of 0 (fail) or 1 (pass). A mean was taken across tasks and internal consistency of the scale was $\alpha = .85$.

2.3.5 Cooperation

Children’s cooperation skills were measured with two previously developed cooperation tasks, drawn from Warneken, Chen, and Tomasello (2006). The cooperation tasks included the trampoline and double tube tasks. The tasks measure the extent to which children cooperate with the tester to achieve a mutual goal. In the trampoline task, the child was invited by the tester to help make a bear dance on a hand-held trampoline. In the double tubes tasks, the child was invited to help the experimenter complete a sequence of actions in which she rolled a ball down one of two tubes and asked the child to catch it at the bottom. Discrete 10-second intervals were coded on a 5-point scale, up to a maximum of 80 seconds for each task. For the trampoline task, the scale ranged from 1 (no success) to 5 (high engagement). A mean of each child’s scores on the trampoline trials was derived. For the double-tubes task, the scale ranged from 1 (no attempt) to 5 (complete success). A mean of each child’s scores on the double tubes catch trials was derived. Coders then rated a global cooperation score along a 4-point scale, based on percentage of the task the child was cooperative (0-25%, 26-50%, 51-75%, 76-100%). Finally, after tasks were completed, coders also rated a global impression of the extent to which children were cooperative with the tester along a 4-point scale for each task, including 0 (uncooperative 3 or more times), 1 (uncooperative twice), 2 (uncooperative once), and 3 (cooperative at all times).
Ten percent of films were double coded by independent coders and the mean inter-rater reliability across all cooperation tasks was $\alpha = .86$ (ranging from .68 to .96). A composite cooperation variable was constructed by taking the mean of the standardized scores, and internal consistency of the five items making up the composite was $\alpha = .86$. Item total correlations ranged from .56 to .82.

2.3.6  

Positivity

We assessed child positivity during a mother-child interaction task. This ensured that any relationship between positivity and scaffolding did not occur simply because of the same task. Mothers and children were engaged in two 5-minute tasks. The first task, Free Play, required dyads to engage in unstructured play. Mothers were instructed to play as they normally would without toys. The Cooperative Building Task was the same task as that described for siblings (Aguilar et al., 2001). Child positivity was rated using 4 items, on a 7-point scale from the Coding of Attachment-Related Parenting (CARP; Matias et al., 2006): attention/on task, enjoyment of activity, social responsiveness, and positive affect. Ten percent of films were double coded by independent coders and the inter-rater reliability for the positivity composite was $\alpha = .92$. Internal consistency was $\alpha = .85$. Item total correlations ranged from .62 to .77.

2.3.7  

Negativity

This construct was also based on data from the mother-child interaction described above and coded with two scales from the CARP (Matias et al., 2006): child disruptive behavior and child’s negative affect. The composite showed an inter-rater reliability of $\alpha = .88$ and an internal consistency of $\alpha = .31$. Item total correlations ranged from .36 to .46.

2.4  

Results

2.4.1  

Comparing the Efficacy of Thin Slice Versus Interval Coding

Means and standard deviations of variables can be seen in Table 3. The majority of the distributions for the thin slice items were normally distributed for both children. The distributions for the interval measure items were highly skewed as particular behaviors were rarely seen in the observational period. For instance, for S1’s ‘positive feedback’ was never coded with the interval measure, presumably because S1’s saw their older sibling as more competent at the task such that praise was not needed. Similarly, responsiveness of Sibling 1 to
requests for help from Sibling 2 occurred less than 1% of the time, because Sibling 2 almost never asked for help.

<table>
<thead>
<tr>
<th></th>
<th>Thin-slice</th>
<th>Interval</th>
<th>Inter-measure correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>S1</td>
<td>S2</td>
<td>S1</td>
</tr>
<tr>
<td>Specific verbal direction</td>
<td>1.92(.63)</td>
<td>2.76(0.84)</td>
<td>.01(.02)</td>
</tr>
<tr>
<td>Positive Non-Verbal Direction</td>
<td>1.93(.66)</td>
<td>2.40(0.79)</td>
<td>.02(.04)</td>
</tr>
<tr>
<td>Meta Level Direction</td>
<td>1.68(.91)</td>
<td>2.51(0.87)</td>
<td>.06(.10)</td>
</tr>
<tr>
<td>Request for Help</td>
<td>2.77(.94)</td>
<td>2.68(0.77)</td>
<td>.10(.17)</td>
</tr>
<tr>
<td>Response to a Request for Help</td>
<td>2.65(1.11)</td>
<td>2.40(0.99)</td>
<td>.00(.01)</td>
</tr>
<tr>
<td>Positive Feedback</td>
<td>1.35(.51)</td>
<td>2.05(1.07)</td>
<td>.00(.00)</td>
</tr>
<tr>
<td>On task</td>
<td>3.04(1.05)</td>
<td>4.03(0.89)</td>
<td>.49(.40)</td>
</tr>
<tr>
<td>Follows rules</td>
<td>2.43(1.20)</td>
<td>3.07(1.04)</td>
<td>.78(.32)</td>
</tr>
<tr>
<td>Turn Taking Directions</td>
<td>1.7(.79)</td>
<td>2.96(1.16)</td>
<td>.02(.05)</td>
</tr>
</tbody>
</table>

*p < .05; **p < .01; a cannot be computed because at least one variable is a constant

Table 3. Means and standard deviations (SD) of thin-slice and interval coding measures, and inter-measure Pearson correlations
Second, we examined agreement between the two coders on thin slice judgments for Sibling 1 and Sibling 2 as shown in Table 2. Agreement between coders is not typically examined in thin slice methodology, but we were interested in whether any items were so poorly worded as not to be used in the composite. It can be seen that ‘ambiguous verbal directions’ showed the lowest level of agreement between coders. This agreement was so low as to suggest that the thin slice coders did not agree on what the item meant. This led us to drop this item from further consideration. With respect to all other individual codes, agreement was high enough (r > .2) for at least one sibling to suggest that coders did have something similar in mind when the description was read.

Third, we examined the agreement across the thin slice and interval codes on individual items. These are shown in the last two columns of Table 3. Agreement across measures for individual items ranges from low to high. Agreement is low for those items that were very infrequently coded in the interval coding. This is of course expected given the lack of variance on these items. For the same reason (i.e. frequency with which behavior was coded in interval coding), agreement across measures for Sibling 2 was higher than that for Sibling 1.

Fourth, we calculated the internal consistency for the thin slice and interval composites using Cronbach’s alpha. With the exception of ‘ambiguous verbal directions’, all items in Table 1 were included in the composites (N=9). Internal consistency was excellent (George & Mallery, 2003) for the thin slice composite, α = .90. Item total correlations ranged between .39 and .82. Internal consistency for the interval composite was questionable (George & Mallery, 2003), α = .58, with item-total correlations ranging from .02 to .55. This pattern would be expected given the low frequency of many items in the interval composite. Thin slice and interval composites were constructed by taking the mean of the items (N=9) for each measure, respectively. The thin slice composite was found to be strongly associated with the interval composite, r (100) = .78, p < .0001.

Independent Samples T-tests examining differences between Sibling 1 and Sibling 2 revealed that thin slice coders rated the Sibling 2 (M=2.76) as more likely to exhibit scaffolding than Sibling 1 (M=2.16), t(98) = -4.45, p < .001. Similarly, for interval coding, Sibling 2 (M=.22) was coded as providing significantly more scaffolding than Sibling 1 (M=.16), t(96) = -3.47, p < .001.

We examined the concurrent validity of the two scaffolding composites by correlating them with constructs hypothesized in the introduction to be related to scaffolding. Results are shown in
Table 4. An examination of the composites reveals that both the thin slice and the interval composites are significantly related to children’s age, theory of mind, linguistic skills, cooperation and interpersonal positivity and negativity (lower levels). Effect sizes were in the medium range for the thin slice measure and a bit less strong, but not significantly so, for the interval measure.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Thin-Slice</th>
<th>Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (in months)</td>
<td>.43**</td>
<td>.39**</td>
</tr>
<tr>
<td>Language</td>
<td>.29**</td>
<td>.21*</td>
</tr>
<tr>
<td>Theory of Mind</td>
<td>.39**</td>
<td>.36**</td>
</tr>
<tr>
<td>Cooperation</td>
<td>.39**</td>
<td>.24*</td>
</tr>
<tr>
<td>Positivity</td>
<td>.33**</td>
<td>.31**</td>
</tr>
<tr>
<td>Negativity</td>
<td>-.34**</td>
<td>-.33**</td>
</tr>
</tbody>
</table>

*p < .05; **p < .01

Table 4. Concurrent validity for thin-slice and interval composites with child-specific measures

2.4.2 The Value of Two Coders Versus One

Our last goal related to evaluating the practice of averaging thin slice codes across two coders. Internal consistency was acceptable for both coders (α = .79 and α = .91). Independent composite scores were calculated for coder 1 and coder 2, respectively. Coder agreement on these composite scores was acceptable at α = .78. An examination of each coder’s data independently, as opposed to averaged, showed that each rater’s scores were significantly correlated with constructs hypothesized to be related to scaffolding, although there were small differences in magnitude of correlations.
It took approximately 7 minutes per coder to code each siblings’ behavior using the thin slice methodology. As this methodology requires two coders, total time spent was 14 minutes. Training (i.e. discussion of concepts) took approximately 1 hour. Total time expenditure coding thin slice was 12.7 hours. For interval coding, each sibling dyad took approximately 20 minutes to code. Additionally, however, training of interval coders (including reliability) required an additional 30 hours (approximately), which results in a total expenditure of 46.6 hours of coding. The interval approach took 3.7 times longer to complete than the thin slice approach.

2.5 Discussion

Opportunity for siblings to contribute to one another’s socio-cognitive development stems from the intimate nature of their relationships, including shared experiences, collaboration, and pretense (Dunn, 2002). Research on sibling teaching has been growing in recent years (e.g. Recchia & Howe, 2009; Recchia et al., 2009; Howe, Brody, & Recchia, 2006; Perez-Granados & Callanan, 1997) and the present study contributes to this field of research both substantively and methodologically. We examined the efficacy of two methods of coding scaffolding in siblings, thin slice and interval coding. Although the children were instructed to work jointly on the task, it is not surprising that the Sibling 2’s exhibited greater levels of scaffolding behavior given their age, language ability, as well as the natural power balance between older and younger siblings (Recchia et al., 2009; Perlman, Siddiqui, Ram, & Ross, 2000).

Correlations between individual thin slice and interval items ranged from small to large. We found that behaviours that occurred in low or no frequency in one sibling group showed lower correlations between thin slice and interval codes for that sibling. In these instances, however, the opposite sibling group showed high correlations. Thin slice coders were able to make judgments on scaffolding, which manifests variably as a function of developmental level, which were less age-constrained. This resulted in a thin slice measure picking up a wider range of scaffolding behaviours, particularly in the Sibling 1 group, than the interval coding. Further, when examining the agreement between the two methods of coding, it is important to keep the task demands in mind. Thin slice coders are given the liberty to take what they have watched and infer the likelihood of each child exhibiting the target behaviour in general, whereas interval coders coded the exact behavior that occurred within each interval of the 5-minute interaction. For example, imagine a Sibling 1 who followed all of the Sibling 2’s directions, was on task,
compliant and generally contributing to the mutually reciprocal interaction at hand. The Sibling 2 may not ask the Sibling 1 for help, as he or she is competent and taking charge of the interaction. In this specific situation, the Sibling 2 may not have given the Sibling 1 any opportunity to be responsive to their sibling’s requests for help. When it comes to coding ‘responsive to request for help’, an interval coder would have an absence of data to code this accurately. However, a thin slice coder would take all available information as indicators that, if the situation were to arise, this Sibling 1 would be responsive to his/her Sibling 2’s request for help. Because coders are asked to use their impressions and to rate on the basis of stable elements of child characteristics (this person is likely to…) they can use a wider range of cues than is acceptable using the interval coding methodology. It is notable that thin slice coders were able to find enough information to form independent impressions of the likelihood of these low frequency behaviours reliably for the composite and with some level of agreement on individual items. This suggests that thin slice coding may have some advantage over interval coding for items that are low in frequency. For Sibling 1, this occurred with items such as ‘responsive to partner’s request for help’ and ‘gives positive feedback.’ For Sibling 2, this was the case for one item, ‘asks for help.’ This pattern of results is explained by the dynamics between older and younger siblings being based on differing levels of competence and power (Recchia et al., 2009; Perlman et al., 2000).

In terms of internal consistency, the interval measure showed lower internal consistency than the thin slice measure. This is partly explained by the poor distributional properties of the interval codes.

The concurrent validity of thin slice and interval coding was also examined. Based on previous work on scaffolding and children’s perspective taking, we expected to see higher levels of scaffolding as a function of children’s age, language ability, theory of mind, and positive behaviour, and lower levels as a function of negative behaviour. Correlations between the scaffolding measures and the hypothesized constructs were stronger for the thin slice coding than for the interval coding. These results suggest greater concurrent validity for the thin slice than the interval measure.

Another important comparison between thin slice coding and interval coding is their demands on resources. It took approximately 3.7 times longer to complete interval coding than thin slice coding, even when two coders are used for the thin slice approach. Using one coder for thin slice
essentially turned the thin slice method into a global rating scale. However, the difference is that it still relies on coders’ first impressions. Given that coders were reliable with one another on the composite measure, the present study suggests that impressionistic thin slice coding can provide a psychometrically sound method for coding scaffolding in siblings when using only one coder as opposed to two.

Although many studies examining sibling teaching connote one sibling as the “teacher” and one as the “learner”, this study gave dyads a cooperative task and examined the natural teaching that occurred. We argue that teaching relates to the ability of an individual to engage with the mind of the learner and that teaching and learning are bi-directional processes occurring between siblings. Thin-slicing allows for efficient coding of large quantities of data by novice researchers. This has important implications for applications of a thin slice approach in practical settings (i.e. home and school) as a tool for conducting research that will guide policy, practice, and evaluation of programs.
References


