Modelling children's Gear task strategy use with the Dynamic Overlapping Waves Model

de Bordes, Pieter F.; Boom, Jan; Schot, Willemijn D.; van den Heuvel-Panhuizen, Marja; Leseman, Paul P. M.

Published in:
Cognitive development

DOI:
10.1016/j.cogdev.2019.01.001

Cognitive Development 50 (2019) 237–247

Contents lists available at ScienceDirect

Cognitive Development

journal homepage: www.elsevier.com/locate/cogdev

Modelling children's Gear task strategy use with the Dynamic Overlapping Waves Model

Pieter F. de Bordes[a,b,*], Jan Boom[c], Willemijn D. Schot[d], Marja van den Heuvel-Panhuizen[e,f], Paul P.M. Leseman[e]

[a] *Department of Interdisciplinary Sciences, Faculty of Social and Behavioural Sciences, Utrecht University, the Netherlands*
[b] *Department of Psychology, University of Groningen, the Netherlands*
[c] *Department of Developmental Psychology, Faculty of Social and Behavioural Sciences, Utrecht University, the Netherlands*
[d] *Educational Consultancy & Professional Development, Faculty of Social and Behavioural Sciences, Utrecht University, the Netherlands*
[e] *Department of Education and Learning Sciences, Faculty of Social and Behavioural Sciences, Utrecht University, the Netherlands*
[f] *Nord University, Norway*

ARTICLE INFO

Keywords:
Problem solving
Dynamic Overlapping Waves Model
Perceptual learning
Embodied experience

ABSTRACT

The Dynamic Overlapping Waves Model (DOWM) can model strategy use in problem-solving tasks for strategies that can be construed as developmentally and hierarchically ordered (Boom, 2015). We observed children's (M_{age} = 11 years, *SD* = 6 months) strategy use during a task in which they had to find the rotation direction of the last gear in a series of connected gear chains, given the rotation direction of the first gear. Using DOWM, we found that strategy use was ordered as expected, from unskilled sensorimotor strategies to abstract strategies, and from less to more efficient in terms of speed and accuracy. This order aligns with the idea that perceptual learning is central to the emergence of abstract conceptual knowledge. Moreover, the current study shows that the DOWM does not preclude forward and backward transitions and even occasional transitions that skip certain strategies in the ordering. The DOWM seems a promising tool to developmentally capture the breadth of behavioral repertoire children display when they adopt new strategies for various problem-solving tasks.

1. Introduction

Many studies have investigated how children learn Science, Technology, Engineering and Mathematics (STEM) topics (Bolger, Kobiela, Weinberg, & Lehrer, 2010; Lehrer & Schaube, 1998; Martin & Schwartz, 2005; The National Research Council, 2014). In tailoring learning situations in which children could learn these topics, an important issue is how to support children's learning processes. While working on tasks, the students' strategy use reveals their progress in understanding the underlying STEM principles; see, for example, the torque principle in Siegler (1976) balance scale task. One of the difficulties in describing and modelling children's learning processes during problem solving is that this development is not always continuous or step-wise and does not occur in similar fashion across children (Fischer & Bidell, 2006; Siegler, 1996; Siegler et al., 2006; Van der Ven, Boom, Kroesbergen, & Leseman, 2012). In addition, children might vacillate between strategies (Boncoddo, Dixon, & Kelley, 2010), sometimes relapsing to less efficient strategies or make sudden jumps to more efficient strategies. This variability could be the result of relevant previous experience, different developmental pathways, or measurement error.

* Corresponding author at: Utrecht University, Department of Interdisciplinary Sciences, Heidelberglaan 1, 3584 CS, Utrecht, The Netherlands.
E-mail address: p.f.deBordes@uu.nl (P.F. de Bordes).

https://doi.org/10.1016/j.cogdev.2019.01.001
Received 15 May 2018; Received in revised form 18 January 2019; Accepted 19 January 2019
Available online 21 May 2019

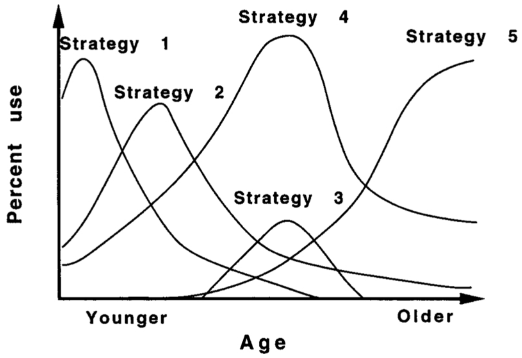

Fig. 1. Overlapping waves depiction of cognitive development from "Emerging Minds," by Siegler (1996). This well-known picture is just a heuristic model, for more realistic distributions see our Fig. 4.

The variability in progress of attaining different strategies is commonly treated as measurement error in developmental theory, supposedly occluding the 'true' development or learning that takes place (Fischer & Bidell, 2006; Siegler et al., 2006; Zheng & Fischer, 2002). In addition, this variability poses difficulties for using standard statistical techniques that rely on linearity of strong order effects of the data (Boom & ter Laak, 2007; Van der Ven et al., 2012). However, close inspection of the variable nature of children's microgenetic development in applying strategies in Siegler's Balance Beam (Boom & ter Laak, 2007) and children's developing numeracy skills (Van der Ven et al., 2012) suggests that this variability is rather an intrinsic part of learning and can be taken into account when modelling microgenetic development (see Van Geert & Van Dijk, 2002; Shrager & Siegler, 1998; Zheng & Fischer, 2002).

1.1. The Dynamic Overlapping Waves Model

According to Siegler (1996) and Siegler et al. (2006) children typically know and use several strategies to solve a given task at a given timepoint. With experience, the relative frequency in which certain strategies are used change, leading to increased or decreased use of certain strategies, abandoning old strategies, and discovering new ones (Siegler et al., 2006). So rather than conceptualizing children's developing strategy use in an all-or-non-fashion, children's strategy use can be seen as a competitive adaption of strategies at any given timepoint. Siegler (1996)) conceptualized this kind of development of strategy use as overlapping waves (Fig. 1). In his Overlapping Waves Model, the development of each strategy use is depicted in the form of a wave along a dimension of increasing maturation or sophistication (Siegler, 1996). Upon discovery of a strategy, children might use this strategy increasingly in respect to other strategies (depicted by a rise of the wave). But next, children might discover a newer strategy, slowly abandoning the now older one (depicted by a fall of the wave).

Boom and colleagues have extended Siegler (1996) Overlapping Waves Model statistically (Boom & ter Laak, 2007; Boom, 2015; Van der Ven et al., 2012) by combining Latent variable Growth curve Modelling (LGM) and Item Response Theory (IRT) modelling. The resulting Dynamic Overlapping Waves Model (DOWM) promises several advantages in comparison to step-wise models or continuous models. First, it can address the frequently observed variability in children's strategy use, incorporating observed variability into the model instead of ignoring it. Second, it can model the use of several qualitatively different strategies in one model, by assuming a latent single continuous dimension that is nonlinearly and probabilistically related to the relative frequency of use of each strategy, instead of just averaging over different strategies. Thirdly, it can model the microgenetic development of strategies as a dynamic process, meaning that strategy use on time $t + 1$ is at least based on strategy use on time t. Finally, the model can deal with (large) inter-individual variation in timing of the learning process, meaning that it can easily incorporate data from participants who are on beginners, intermediate or advanced level regarding their proficiency on the learning task, and/or who differ in amount of change or learning.

1.2. Development of strategy use during the gear task

In the current study, we applied the DOWM to capture children's development in strategy use on a problem-solving task involving force transmission through gear-chains, known as the Gear task (see, e.g., Dixon & Bangert, 2002). For this task, participants are typically asked to predict the direction of movement of a final gear (given the direction of the first gear) within a visually presented static chain of gears (Fig. 2) and are asked how they obtained their answer (e.g., Dixon & Dohn, 2003; Dixon & Kelley, 2007).The different strategies to solve the Gear task and transitions between them have been well documented and defined (Boncoddo et al., 2010; Dixon & Bangert, 2002, 2004; Dixon & Dohn, 2003; Dixon, Holden, Mirman, & Stephen, 2012; Dixon & Kelley, 2007; Dixon, Stephen, Boncoddo, & Anastas, 2010; Stephen, Boncoddo, Magnuson, & Dixon, 2009; Trudeau & Dixon, 2007), allowing these strategies to be used for complex modelling.

The results of the aforementioned studies using the Gear task typically show that most of the participants start out by literally tracing presumed local force transmission across gears by means of eye movements and/or gestures, known as the Force Tracing

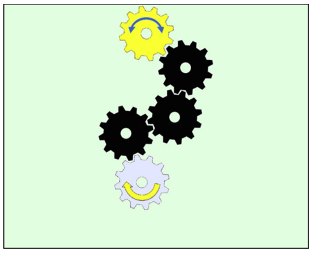

Fig. 2. An example of a trial during the Gear task.

strategy (e.g., Dixon & Bangert, 2002). This suggests that participants use sensorimotor resources in order to simulate the local force transmission from each gear to the next (Alibali, Spencer, Knox, & Kita, 2011; Boncoddo et al., 2010; Dixon & Kelley, 2007; Trudeau & Dixon, 2007). According to Dixon and Kelley (2007) classifying each subsequent gear as either left turning or right turning (i.e., Classification strategy) is usually developed directly out of the Force Tracing strategy. This abstraction occurs, according to them, because the actions that coincide with force tracing (i.e., moving hands or eyes along the gear chain from left to right and vice versa) contain alternation information. In support of this claim, Boncoddo et al. (2010) found that the number of children's gestures associated with force tracing predicted later emergence of the Classification strategy (also see Trudeau & Dixon, 2007, for similar findings). This illustrates how finding an abstract strategy like the Classification strategy is perhaps ultimately grounded in sensorimotor actions whereby the *movement* of a gear is abstracted in terms of *directionality* (either left or right) of each gear along the chain.

A more efficient strategy that is seldom used by both adults and children is the Skipping strategy and consists of classifying only every other gear as turning into the same direction, starting from the first gear (e.g., Dixon & Bangert, 2002). This strategy eliminates the need to label (including pointing towards) each gear individually, making it more efficient in its execution. The directionality of the gears seems to be central to this strategy as with the Classification strategy and it already seems to hint towards the mathematical notion of parity. However, in contrast to the Parity strategy, the gears do not have to be counted in order to apply the Skipping strategy. Using the Parity strategy entails applying the rule that the first and last gears turn in opposite direction when the number of gears is even (even parity) and in the same direction otherwise (uneven parity). When using this strategy, the movement of each gear and directionality of the gears along the chain, central to the Force Tracing strategy and the Classification strategy, respectively, seem to be inherent to the *number of gears* within the chain that is used to apply the Parity strategy. Indeed, participants usually apply this strategy only after using the Force Tracing strategy, the Classification strategy, the Skipping strategy or all of them (Dixon & Bangert, 2002, 2004; Stephen et al., 2009). Like the shift from the Force Tracing strategy to the Classification strategy described earlier, applying the Classification strategy and/or the Skipping strategy might yield information about the parity (i.e., left and right alternate along the chain of gears) that participants can use to discover the Parity strategy.

1.3. Embodied experience

Children up to 12 years of age rarely use the Parity strategy spontaneously within the Gear task (Dixon & Bangert, 2002). Because sensorimotor actions seem to play a pivotal role in developing abstract strategies for solving the Gear task (see Dixon et al., 2010) and learning abstract concepts in general (Rambusch & Ziemke, 2005), the use of the Parity strategy might be increased for this age group if they can manually explore energy transmission through gears prior to the Gear task. This would stress the importance of embodied experience in learning abstract strategies. In addition, it might allow us to include the Parity strategy in the overlapping waves model, elucidating the full range of development possible for this task. To this end, we devised an exploration task in which children could make either curved or straight gear tracks, using plastic gears. The alternating rotational direction of gears along the gear chain would be clearly visible when they would make straight tracks compared to curved tracks. Making the straight tracks therefore might lead children to already discover something about the functional relation of interlocked gears in terms of alternating rotational movements, possibly leading them to discover and apply more abstract strategies during the Gear task than children who made curved tracks.

1.4. Hypotheses

Taken together, we expect strategy use for the Gear task to be ordered developmentally in which the discovery of abstract strategies is ultimately grounded in embodied experience and specifically the sensorimotor actions that coincide with executing less advanced strategies. Our first hypothesis is therefore that the strategies can be ordered in terms of becoming more efficient, meaning that the more abstract the applied strategy is, the less sensorimotor input is required (as indicated by shorter response times) and the more trials are correct. Participants attune to less (and possibly different) perceptual information in order to perform well when abstract strategies are used compared to sensorimotor strategies, making them more efficient to solve the Gear task trials. Our second hypothesis is that there is a specific order of microgenetic development in which these strategies (i.e., the Force Tracing strategy, Classification strategy, Skipping strategy, Parity strategy, respectively) are used. Our third hypothesis is that children learn from embodied experience with energy transmission through gears by which they can discover and apply more abstract strategies on the

Gear task afterwards. Specifically, we expect that if they would make straight tracks, they will apply more abstract strategies on the subsequent Gear task in comparison to those that made curved gear tracks.

2. Methods

2.1. Design

The present study has a microgenetic design. First, children's strategy use during the Gear task was assessed repeatedly (by means of interviews) over a short period of time. Second, we measured the number of (in)correct trials and the response time per trial of the Gear task. Third, we assessed whether making straight gear tracks in the exploration task would lead children to make more use of abstract strategies in the subsequent Gear task than if they would make curved tracks during the exploration task.

2.2. Participants

The sample consisted of 69 primary school children (M_{age} = 11 years; 0 months, SD = 6 months) from two different schools in the Netherlands. We excluded the data of one participant from the analysis since she had an arm in a cast. In addition, we excluded the data of two participants from the analysis because one child was not speaking loud enough for coding, and one child did not understand the questions during the Gear task (answers were about whether the child correctly solved trials or not, instead of *how* he solved it). Excluding these participants left 66 participants in the sample of which 29 were boys (M_{age} = 11 years; 1 month, SD = 6 months and 37 were girls (M_{age} = 10 years; 11 months, SD = 5 months).

2.3. Materials

2.3.1. Gear task

A total of 30 trials were displayed on a 17" computer screen using E-prime 2 (Schneider, Eschman, & Zuccolotto, 2002). Each trial contained a gear-chain with a purple gear with a yellow arrow pointing clock-wise, a yellow gear with a black, bidirectional arrow, and one to six black gears (5 of each, presented in random order across trials and across children) in between (Fig. 2). Children indicated whether they thought the yellow gear would turn clockwise or counterclockwise by pressing one of two response buttons on the computer keyboard that were labelled with a symbol of a rotating arrow (one with a counter clock-wise arrow on the left and one with a clock-wise arrow on the right). After each response, auditory and visual feedback was provided for 2000 ms. If the trial was solved correctly, a smiling mouse would appear saying "*very good*". If the trial was solved incorrectly, a mouse with a neutral expression would appear saying "*incorrect*". After each third trial, a mouse and question mark were presented, indicating a verbal response by the child was required in which the child explained how he or she solved the previous trials.

2.3.2. Exploration task

The materials for this task consisted of yellow, plastic gears, colored pegs (blue, yellow, and red) and plastic white pegboards in which yellow pegs could be placed as to attach the gears.

2.4. Procedure

2.4.1. General procedure

Testing took place in a quiet room in a school building. As a cover story, children were told that they would play two different games in succession, after which they had to decide which game they favored most.

2.4.2. Procedure Gear task

The Gear task was in part inspired by the task involving gears as used in several studies by Dixon and colleagues (e.g., Boncoddo et al., 2010; Dixon & Bangert, 2002, 2004; Dixon & Dohn, 2003). Children were seated, facing the computer screen at a distance of 70–80 cm, while a test trial displayed two adjacent gears; one purple and one yellow. The researcher explained that on each trial, there would be one purple gear (with an arrow always pointing clockwise) and one yellow gear (with bidirectional arrows) that were interlocked by one or more gears. Next, the researcher asked the child to which direction the yellow gear would rotate. Children usually answered this question correctly (three children gave an initial incorrect answer but revised their answer after a simple "*are you sure?*" question) by indicating (verbally or by means of pointing) that the yellow gear would go counterclockwise. Next, the response buttons were shown on the computer keyboard. The researcher explained that after every *three* trials, a screen with a mouse with the question mark would appear and the children had to state how they solved the last three trials. If children had no further questions, the task started. Questions were for example: "How did you solve the puzzle?" or "How did you know to what way the final gear was rotating", or after having asked this "How did you solve it this time?" Whenever the children's answer was vague or incomplete, follow-up questions were: "Can you explain that a bit more?" or "Can you explain that in different words?"

2.4.3. Procedure exploration task

The experimenter introduced children to the task by presenting a plastic toy pegboard with five interlocking gears attached to it with pegs. The experimenter replaced the middle gear of the gear chain as to demonstrate how the gears can be placed in and out of

Table 1
Strategy Codes, Definitions and Examples.

Strategy	Definition	Example
1. Guessing/other	The child either used an unknown strategy or stated that he/she guessed.	*"The curve in the gear chain is causing all the gears to go left" or "I don't know!"*
2. Force Tracing	Enacting local force transmission across gears by means of eye movements and/ or gestures.	*"This gear is going that way and it pushes the next gear that way "*
3. Classification	Classifying each subsequent gear as either left turning or right turning.	*"If this gear goes left, the next gear goes right and the one after goes left"*
4. Skipping	Classifying every other gear as turning into the same direction until the last gear is reached, starting with the first gear	*"This gear turns the same way as that gear because there is one in between"*
5. Parity	Applying the rule that the first and last gears turn in opposite direction when the number of gears is even (even parity) and in the same direction otherwise (uneven parity).	*"It's an uneven number so the first and last go the same way" or "If the first gear goes left, the third gear also goes left"*

Note: Strategies were coded with both verbal and non-verbal behavior considered.

the pegboard. The experimenter instructed children to make a gear-chain as such, that the first and last gear would rotate into the similar direction. While rotating the gears, the researcher stressed that the first and last gears (attached with the blue pegs) were rotating in the same direction to illustrate what the end state of each trial should look like. In addition, they were told not to touch the red pegs already present on the board with the gears. By using specific configurations of the red pegs, children were constrained to make either straight tracks or curved tracks in three different trials in a fixed order. Finally, the children were told that they had to indicate when the gear track was finished. After this instruction, the first pegboard with two gears on it (attached with blue pins), yellow pegs, and gears were placed in front of the child and the Exploration task started. Whenever children finished the gear track, they typically turned one of the gears that would cause for all the gears within the chain to move. We then asked whether the gears attached with the blue pegs rotated in the same direction. If this was the case and if the child confirmed this, we would praise the child and replace the board with the next one until all three parts of the task were finished.

2.5. Measures and coding data

We videotaped all ten answer fragments for subsequent coding of verbal and nonverbal behavior in terms of strategy use and we assigned a code of one to five for Guessing/ unknown strategy, Force Tracing strategy, Classification strategy, Skipping, and Parity strategy, respectively (see Table 1). To assure inter-rater reliability, we recoded 20% of the data on strategy use that resulted in a kappa of 0.84, which is high (Landis & Koch, 1977). So, a participant was credited with using one out of five mutually exclusive strategies for each of our ten measurement points (items) or was set to missing when the response could not be classified according to the strategy codes (a total of 5.09% had to be set to missing), resulting in 10 *item* codes per child. Next, we calculated the modus of strategy use on the gear task over the 10 measurement occasions per child in order to analyse the relation between the most frequently used strategy, amount of trials correct, and average reaction time per trial.

We registered responses as correct '1' or incorrect '0' and summed them, resulting in scores that could range from 0 to 30. In addition, we counted the number of trials correct per set of three trials after the children's strategy use was inquired and transformed this number into percentages of items correct per used strategy across children. With this, we could analyze the relation between the percentage of trials correct per strategy use. Accompanying response times (in milliseconds) on each trial were also registered, using E-prime 2 (Schneider et al., 2002). The responses and response times were averaged within children for each set size (number of gears in trials). Of this dataset, two values were omitted from further analysis as they were more than 3 standard deviations from the average response time of children for that set size. Finally, no behavioral measures were used during the exploration task as the aim of this task was to test whether the condition contrast would affect strategy progress in the subsequent gear task.

2.6. Analysis

Participants were asked the same questions ten times, with three trials of gear making in between questions. The coding system described above was used for each to these ten repeated measurements (the items). The core of our DOWM model is a hypothetical latent ability that is different for each participant (inter-individual differences), but also possibly changing within participant (intra-individual differences) over the ten measurement occasions. To examine level and change in this ability a Latent Growth Modelling (LGM) approach was used. An LGM presupposes a steady increase or decrease in such an ability over a small number of regularly spaced measurement occasions for each person, with the baseline or average over all measurement occasions denoted as the intercept, and the change (if any) from one measurement occasion to the next denoted as the slope. More on LGM, as a variety of longitudinal Structural Equation Modeling (SEM), can be found in Bollen and Curran (2006).

However, this ability itself cannot be directly observed: it is an underlying, hypothetical, hence latent ability. What only can be observed are the ten responses of the children, scored as Guessing/ unknown strategy, Force Tracing strategy, Classification strategy, Skipping strategy, and Parity strategy (coded as 1–5). So, we need to map the manifest display of these five strategies, in the participants recorded answers, to the latent ability of these participants.

Fortunately, Item Response Theory models can estimate the probability of choosing a certain strategy as a function of the properties of that strategy and the latent ability of the respondent, both expressed on a common scale, with interesting properties. With increasing ability, the likelihood of using the strategy 1 will decrease, the likelihood of using the strategy 5 will increase and the likelihood of using in-between strategies will rise and fall (depending on where you start; See Fig. 4). The properties of that strategy are represented in Fig. 4 by the location (difficulty) and by the relative flatness/steepness (discrimination) of the three shapes. The difficulty of the strategy is characterized by a parameter (the threshold) that defines the border between two subsequent categories (strategies), the discrimination (sensitivity) is characterized by the residual variance for the category. The larger the positive difference between the ability of the participant and the difficulty of the strategy, the larger the likelihood of using that strategy. The larger the negative difference between the ability and the difficulty of the strategy, the smaller the likelihood of using that strategy. Fig. 4 illustrates the distributions of each of the five strategies so that the likelihood of responding according to one of the included strategies can be read off easily: e.g. with latent ability 1.0 (the scale has an arbitrary origin or zero point, so the absolute value is not important, but relative distances are) the likelihood of using strategy 3 and strategy 2 is almost the same, while the other strategies are not likely to be used by a hypothetical child with this level of ability. However, with a latent ability of e.g. 4.0, the highest three strategies (3, 4, and 5) have an almost equal chance of being used.

Fig. 4 applies to all 10 repeated items equally because analyzing such strategy change requires measurement invariance. This means that all items are supposed to have the same properties, and only respondents can change over the 10 measurement occasions/items. Therefore, in the analysis, categories were restricted to have the same difficulty and discrimination[1] over the 10 measurement occasions/items. A general increase in how well participants do, would not be surprising, as would be individual differences in both starting point (position on the latent ability x-axis in Fig. 4) and in change (how much a child would move to the right along this latent ability axis) over the 10 occasions/items. We used a Graded Response Model (GRM; Ostini & Nering, 2006; Samejima, 1969) appropriate for multi-category items. An important assumption of such a model is that strategies are orderd as a series of steps that are mastered in sequence when solving trials on the gear task. Overall fit will indicate whether this assumption is tenable. The model reveals furthermore whether the strategies are equally difficult or not, and how much overlap there is, as can be seen in Fig. 4. The model thus provides important information about the strategies.

In sum: We combined and integrated Item Response Theory (IRT) modeling and Latent Growth Modelling (LGM) to model the responses of the participants on the ten questions. The IRT part is not only needed because the data are categorical and the LGM part is not only needed to account for the repeated nature of the measurements[2]. The combined and integrated analysis offers more: the model allows to relate differences in latent ability of participants to differences in the likelihood of using the particular strategies (in an inter-individual sense) but also to model increases in ability to shifts in using the particular strategies (in an intra-individual sense). Learning in this sense implies a shift in the probability distribution of strategy use: chances increase of using a more advanced strategy while chances decrease of using a less advanced strategy.

Although we conceptually distinguished two components in the model, the entire analysis was done as one statistical model in Mplus 8, with the Weighted Least Squares Means and Variances adjusted estimator (WLS-MV), a PROBIT link, and Theta parameterization (Muthén & Muthén, 1998). An advantage of using the WLS-MV estimator in Mplus is that absolute overall fit measures are available. This overall fit of the model was evaluated in terms of the indices Root Mean Square Error of Approximation (RMSEA), Comparative Fit Index (CFI), and Tucker Lewis Index (TLI), following commonly applied cut-off criteria. Specifically, model fit was considered good if RMSEA was below 0.05, and CFI and TLI were above 0.90 (Little, 2013).

3. Results

3.1. Descriptive statistics

On the Gear task, children had on average 23 trials ($SD = 5.6$) out of 30 trials correct and spend on average 10 s ($SD = 5.25$ s.) per trial. Making straight or curved gear tracks during the exploration task did not affect performance on the Gear task in terms of number of trials correct ($M_{difference} = -2.52$, $p = .07$), modus of strategy use ($M_{difference} = -0.015$, $p = .95$), and average response times ($M_{difference} = 523.80$ ms, $p = .64$). The average response time per child per trial was related to the set size (number of gears visible trials), $r = 0.45$, $p < .001$. Gender did not affect this relation as split sample correlations did not deviate from this outcome or from each other. There was no relation between the set size and whether it was correctly solved, $r = -0.003$, $p = .88$. In sum: while accuracy *did not* increase with increasing set size, the average response time and the standard deviation of the response time *did* increase with increasing set size (see Table 2).

Frequencies of strategy use per measurement point during the Gear task are shown in Fig. 3. A total of 41 11-year-olds (62.12%) used the Force Tracing strategy at least once, 38 children (57.58%) used the Classification strategy at least once, 16 children (24.24%) used the Skipping strategy at least once, and 6 children (9.10%) used the Parity strategy at least once.

As can be seen in Fig. 3, the frequency of children using the Force-tracing and Classification strategies slightly decreases over the

[1] Implemented by requiring that thresholds are the same across occasions/items and that residual variances are the same by fixing them to 1 for all items. We also fixed the intercept mean to zero, as this was a convenient way to anchor the scale which is needed for model identification. More detailed information can be obtained from the second author.

[2] Because measurements are repeated, correlations between items are to be expected, which would violate IRT assumptions. However, because LGM accounts for this correlational structure between items, the IRT model can still be used. See example 6.5 in the Mplus users manual (Muthén & Muthén, 1998).

Table 2
Group results of percentage of trials correct, average response time (in ms) and standard deviations of the response time (in ms) for each set size (amount of gears displayed in a trial) of the Problem-solving task.

	Set size					
	3	4	5	6	7	8
Percentage trials correct	0.78	0.80	0.79	0.80	0.77	0.78
Average Response Time	5738	7687	8848	10240	11412	13412
SD Response Time	2765	4301	5171	5423	5396	7530

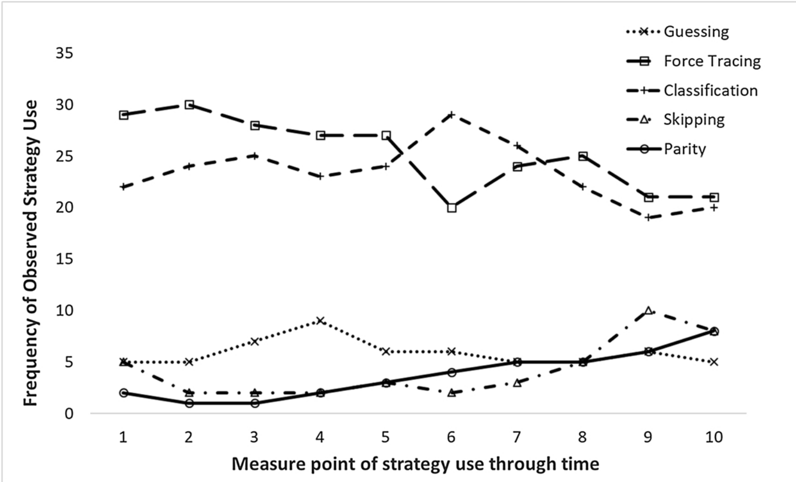

Fig. 3. Frequency of observed strategy use per measure point during the Gear task.

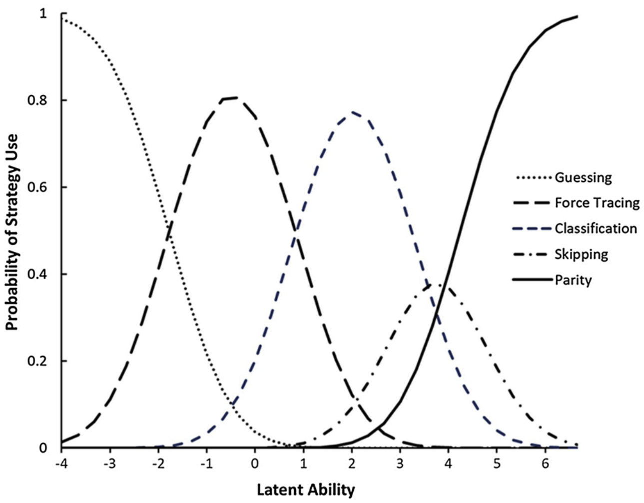

Fig. 4. Category Characteristic Curves (CCCs) with Probability of Strategy Use on the Y-axis and Latent Ability on the X-axis. The Latent Ability scale may stand for inter-individual differences, but also for intra-individual learning differences.

course of measurements whereas the frequency of children using the Skipping and Parity strategies increases slightly. However, it is unclear how all these strategies relate to each other intra-individually in terms of developmental sequence: note that if the frequency at a next timepoint is the same (or higher) it does not imply that the same children are involved. Therefore, in order to test whether the strategies have the developmental order as hypothesized in the introduction, we modelled strategy use in a manner that takes the individual developmental trajectories into account (see Section 3.3).

Table 3
Correlation Matrix Between Gender, Age, Number of Trials Correct, Average Response Time, and Modus of Strategy Use.

	1	2	3	4	5
1. Gender	–	−0.09	−0.16	−0.03	−0.13
2. Age		–	−0.01	−0.04	0.16
3. Number of trials correct			–	−0.11	0.37**
4. Average response time				–	−0.44**
5. Modus of strategy use					–

** $p < 0.01$ (2-tailed).

Table 4
Percentage of trials correct per strategy use across children.

Strategy	Proportion correct
Guessing	0.57
Force Tracing	0.73
Classification	0.86
Skipping	0.94
Parity	0.94

3.2. Number of trials correct and response time

Interindividual correlations between gender, age, number of trials correct, average response time, and modus of strategy use are shown in Table 3. As can be seen, there was a positive correlation between the modus of strategy use and number of trials solved correctly (r = .37, p < .001), meaning that children who applied more abstract strategies yielded more correct answers. Table 4 further reveals that more abstract strategies yield a larger percentage of trials correct. In addition, there was a negative correlation between modus of strategy use and response time (r = −0.44, p < .001), meaning that children who applied more abstract strategies solved the task faster. Whether strategy use is related to number correct and response time on the intraindividual level, is addressed in the next section.

3.3. Dynamic Overlapping Waves Model

Using the DOWM as analysis tool revealed several interesting results. In Fig. 4 the Category Characteristic Curves (CCCs) of the final model are given. As explained in the method section, the likelihood of responding according to one of the strategies can be read off easily: e.g. with latent ability 1.0 the likelihood of using strategy 3 and strategy 2 is almost the same, while with a latent ability of e.g. 4.0, the highest three strategies (3, 4, and 5) have an almost equal chance of being used.

The scale on the X-axis represents latent ability and learning and pertains to the children. Whereas such a figure is customarily used for charting inter-individual ability differences, and that still is possible, it can also be used for charting intra-individual learning differences. With increasing learning level (or ability, or maturity, or development, or moving to the right) the likelihood of responding according to a particular strategy changes: for the guessing strategy it clearly goes down, for the parity strategy it increases, while for the in-between strategies it first rises and then falls. The thresholds determine the position of the CCCs as they demarcate where a next strategy becomes dominant. The thresholds and curves are invariant and fixed over items, and thus over time, and define the ruler that makes it possible to gauge intra-individual change and inter-individual differences. When it comes to inter-individual differences within our sample of 11-year-olds, it seems that most children with the average ability (0 on the x-axis of Fig. 4) apply the Force Tracing strategy. In addition, children in this sample were least likely to use the Skipping strategy (indicated by the lowest peak on the y-axis, compared to the other strategies).

The fit of initial model was not acceptable (RMSEA 0.111, CFI 0.958, and TLI 0.967), but because the sample size is small and the results seemed promising, we accepted two modifications to improve overall fit: item two has been discarded entirely due to improper behavior, such as a negative residual variance, and item 4 and 6 were allowed to co-vary (based on the modification indices). The improved model had acceptable to good fit (RMSEA 0.076, CFI 0.979, and TLI 0.984). The average slope increased 0.134 units per occasion[3]. So, the average increase over all 9 measurement occasions/items is (8*0.134) 1.072 units on the X-axis scale in Fig. 4. There was a significant variance for both the intercept (1.87 units) and the slope (0.062 units). In addition, the individual variance of the slope could be predicted from school-type (school A or B; beta 0.54), but not by Condition (straight or curved gear-chains in the exploration phase; ns). However, since their interaction had a considerable effect (beta −0,423) we kept Condition in the model (fixed at zero). We centered the school and condition indicators and used these for creating an interaction term. The intercept variance could not be explained by these same predictors.

There was a positive correlation between the level of ability (intercept) and number correct (r = 0.50, p < .001) and a negative

[3] For the initial model all the parameters had very similar estimates.

correlation between ability (intercept) and response time ($r = -0.565, p < .001$) confirming and accentuating what was reported in 3.2. In addition, there is a positive correlation between the change in ability (slope) and number correct ($r = 0.30, p < .001$). A negative correlation between change in ability (slope) and response time was not significant ($r = -0.147, p = .303$).

4. Discussion

In this study we measured children's reasoning and performance on a problem-solving task involving gears. We predicted that the strategies can be ordered in terms of becoming more efficient as using more abstract strategies yielded shorter response times per trial and a larger number of trials correct. We found that children using a more abstract strategy had more trials correct. One explanation for this might be that more abstract strategies like the Parity strategy relies less on sensorimotor actions and different perceptual information, like when applying the Force Tracing strategy, that is prone for errors. For example, when children apply the Force Tracing strategy, they visually track the rotational direction of each individual gear and the direction the gears interlock along the whole gear chain. When applying the Parity strategy, however, they simply count each gear along the chain without tracking the movement of each gear or the way the gears interlock. In support of the claim that abstract strategies require fewer sensorimotor actions, we found that more abstract strategies require less time to solve the Gear task trials. Finally, although the set size of the trials was positively related with the time needed to solve those trials, set size did not affect whether children found the correct answer.

Next, we assumed that the different strategies for solving this task would have a specific order in microgenetic development, running from strategies that draw heavily on sensorimotor actions (i.e., Force Tracing strategy) to strategies that are more abstract (Parity strategy). That is to say, we expected the probability of using each of the strategies to rise and fall as experience with the task increases, with abstract strategies gaining probability to be used and more rudimentary strategies losing probability to be used. This is important in that it is consistent with the idea that strategy use can be seen as a competitive adaption of strategies at any given timepoint and as experience on the task increases, more abstract strategies will be likely applied. We tried to fit the five strategies into the model in this presupposed order. The order itself therefore, strictly speaking, cannot be tested, but the good fit we found supports this presupposed order as at least plausible. Although the model presupposes that the strategies are developmentally ordered in general, the, strategy use on the individual level is not always sequential or even progressive. This is often considered to be error variance but here we entertain the possibility that ability itself on the task can deteriorate (e.g. by inattentiveness) or make sudden jumps (e.g. after experiencing a sudden flash of insight). This natural variability could be seen as a process of repeated building up and collapse of task-ability (i.e., "scalloping", Zheng & Fischer, 2002) and is perceived to be crucial for learning to occur (Siegler et al., 2006; Zheng & Fischer, 2002). The DOWM allows for a limited amount of such backward transitions and even transitions that skip certain strategies on the individual level.

Taken together, we perceive this microgenetic development as a result of perceptual learning in which children are able to differentiate useful information for a more advanced strategy while working on the task with a certain strategy (see Adolph & Kretch, 2015; Adolph, Joh, Franchak, Ishak, & Gill, 2008; Gibson & Pick, 2000). For example, the Classification strategy is usually found after applying the Force Tracing strategy because alternation information that is used for the Classification strategy is nested within the actions that coincide with force tracing (Dixon & Kelley, 2007). From a perceptual learning account, using a more *abstract* strategy in this sense does not mean that it is based outside a physical realm (e.g., mental representation) but rather that it is nested within the prolonged interaction with the task materials (also see Dixon et al., 2009; Trudeau & Dixon, 2007 for a similar account).

Finally, manual exploration of gears in the exploration task might have helped children to discover and apply abstract strategies in the Gear task later on as a considerable number of children in the current study used the Skipping strategy (24.24%) and the Parity strategy (9.10%), in comparison to children of comparable ages in the study of Dixon and Bangert (2002). This allowed us to adopt the Parity strategy in DOWM. In addition, this seems to stress the importance of embodied experience in acquiring an advanced understanding of STEM related topics (Abrahamson & Lindgren, 2014; Weisberg & Newcombe, 2017). However, the effect of other differences in task design between our study Dixon and Bangert (2002) study cannot be ruled out. Hence, the effect of embodied experience with gears in the exploration task on the discovery and application of abstract strategies in the subsequent Gear task is solely speculative. We did not find any differences in strategy use on the Gear task between children constructing curved or straight gear tracks during the exploration phase, possibly because the difference between the conditions might have been too subtle or because the sample was too small as the found difference was close to significance. This still leaves the question open of what kind of sensorimotor input is most beneficial to acquiring an abstract notion of the system at hand (see Pouw, Van Gog, & Paas, 2014 for an extended discussion on this issue). For further investigation of this matter, it might be helpful to increase the salience of the alternation in rotational movements along the gear chain by assigning each alternating gear a specific color (see Dixon & Dohn, 2003).

4.1. Limitations of the study

A limitation of our study is that some children had difficulties initiating the Gear task. This was evidenced by a large variation in both response times as performance (correct/incorrect) of the second trial and difficulty in resuming the task at the fourth trial, just after they received the first question on how they resolved the first three trials. This seems to indicate that the training trail was insufficient for children to get properly accustomed to the task demands and that the first items might have measured task understanding, rather than task proficiency. Ideally, we would have preferred to drop the first 6 trials (=2 items) from the analysis to be assured that task proficiency instead of task understanding was measured. However, given the limited number of trials, this was not feasible. For future research, the training session should be extended and should include the interview questions that were used during the Gear task to get children accustomed to these questions. Next, it is known that younger children, unlike adults, hardly arrive at using the parity strategy (see Dixon & Bangert, 2002), meaning that the frequencies of strategy use are age dependent. However, including younger and older children would have been interesting as that would make it possible to analyze whether development of strategy use is affected by age, which is feasible using DOWM as age can simply be added to the model as a predictor. Finally, it would have been interesting to perform an analysis that could reveal around which point(s) of measurement children are likely to progress in strategy use. However, this analysis would not yield robust results because of the small sample used in this study.

4.2. Strengths of the DOWM

As the DOWM can deal with (large) inter-individual variation in timing of the learning process, it can easily incorporate data from participants who are on beginners, intermediate or advanced level within the proposed microgenetic development. DOWM takes the observed variability of strategy use of all participants into account in constructing the Category Characteristic Curves for each strategy separately. Importantly, these Category Characteristic Curves are placed along one latent single continuous dimension that is nonlinearly and probabilistically related to the relative frequency of use of each strategy, instead of just averaging over different strategies. Therefore, it can display the full breath of how a *repertoire* of strategy use can develop across children within a single model instead of proposing a single possible developmental pathway along the different strategies or the development of a single strategy. With typical item response functions for polytomous items the personal latent ability can be linked to a profile of responding in one of the categories. Therefore, the most important advantage of applying the DOWM might be that the relations between the use of the strategies in tasks like this, that have always been difficult to conceptualize, are now easily accessible both conceptually as well as computationally.

We have found an effect of school on strategy use in the DOWM, meaning that children from one school seemed to apply more abstract strategies during the Gear task than children going to the other school. We did not predict this difference and to our knowledge, neither schools have taught children anything specifically on force transmission by means of gear chains. Although this unexplained variability can be perceived as a weakness of the current study, it can also be perceived as a strength of the DOWM in that it can incorporate data of children from different schools in modelling children's microgenetic development.

For example, it can be used to test hypotheses on the effectiveness of different teaching methods in terms of what children gain in understanding STEM topics over time. In addition, it could assist teachers in designing educational activities that provide children the right challenges at the right time, in the right sequence, and at the proper pace. Finally, it can be used to test predictions from theories on the amount and developmental order of discrete steps in which children can learn such as strategy use (e.g., for the Balance Beam, see Siegler & Chen, 2002) or skill level (Fischer & Bidell, 2006; Van Der Steen, Steenbeek, Van Dijk, & Van Geert, 2014). Therefore, we suggest that next to modelling strategy progression, DOWM is a promising tool that can be used on a wider scope in education and research. To further discover the applicability of DOWM and exploit its merits, we suggest that for future research the DOWM is applied on children's strategy use on different tasks and with larger populations.

Acknowledgements

We would like to thank Suzanne van Vliet, Nelleke Doelman, and Lineke Spiele for helping with the data-collection and the Curious Minds Platform, Platform Delta, and professor Tom ter Bogt for supporting this study.

References

Abrahamson, D., & Lindgren, R. (2014). Embodiment and embodied design. *The Cambridge Handbook of the Learning Sciences, 2*, 358–376. https://doi.org/10.1017/CBO9781139519526.

Adolph, K. E., & Kretch, K. S. (2015). Gibson's theory of perceptual learning. *International Encyclopedia of the Social and Behavioral sciences, 10*, 127–134. https://doi.org/10.1016/B978-0-08-097086-8.23096-1.

Adolph, K. E., Joh, A. S., Franchak, J. M., Ishak, S., & Gill, S. V. (2008). Flexibility in the development of action. In J. Bargh, P. Gollwitzer, & E. Morsella (Eds.). *Oxford handbook of human action.* New York: Oxford University Press.

Alibali, M. W., Spencer, R. C., Knox, L., & Kita, S. (2011). Spontaneous gestures influences strategy choices in problem solving. *Psychological Science, 22*, 1138–1144. https://doi.org/10.1177/0956797611417722.

Bolger, M. S., Kobiela, M., Weinberg, P. J., & Lehrer, R. (2010). Embodied experiences within an engineering curriculum. *Proceedings of the 9th International Conference of the Learning Sciences (ICLS, 2010), Vol. 1*, 706–713.

Bollen, K. A., & Curran, P. J. (2006). *Latent curve models.* Hoboken, NJ: Wileyhttps://doi.org/10.1002/0471746096.

Boncoddo, R., Dixon, J. A., & Kelley, E. (2010). The emergence of a novel representation from action: Evidence from preschoolers. *Developmental Science, 13*, 370–377. https://doi.org/10.1111/j.1467-7687.2009.00905.x.

Boom, J. (2015). A new visualization and conceptualization of categorical longitudinal development: Measurement invariance and change. *Frontiers in Psychology, 6*(289), https://doi.org/10.3389/fpsyg.2015.00289.

Boom, J., & ter Laak, J. (2007). Classes in the balance: Latent class analysis and the balance scale task. *Developmental Review, 27*, 127–149. https://doi.org/10.1016/j.dr.2006.06.001.

Dixon, J. A., & Bangert, A. S. (2002). The prehistory of discovery: Precursors of representational change in solving gear-systems problems. *Developmental Psychology, 38*, 918–933. https://doi.org/10.1037/0012-1649.38.6.918.

Dixon, J. A., & Bangert, A. S. (2004). On the spontaneous discovery of a mathematical relation during problem solving. *Cognitive Science, 28*, 433–449. https://doi.org/10.1207/s15516709cog2803_6.

Dixon, J. A., & Dohn, M. C. (2003). Redescription disembeds relations: Evidence from relational transfer and use in problem solving. *Memory & Cognition, 31*, 1082–1093. https://doi.org/10.3758/BF03196129.

Dixon, J. A., & Kelley, E. (2007). Theory revision and redescription: Complementary processes in knowledge acquisition. *Current Directions in Psychological Science, 16*, 111–115. https://doi.org/10.1111/j.1467-8721.2007.00486.x.

Dixon, J. A., Stephen, D. G., Boncoddo, R., & Anastas, J. (2010). The self-organization of cognitive structure. *Psychology of Learning and Motivation, 52*, 343–384. https://doi.org/10.1016/S0079-7421(10)52009-7.

Dixon, J. A., Holden, J. G., Mirman, D., & Stephen, D. G. (2012). Multifractal dynamics in the emergence of cognitive structure. *Topics in Cognitive Science, 4*, 51–62. https://doi.org/10.1111/j.1756-8765.2011.01162.x.

Fischer, K. W., & Bidell, T. R. (2006). Dynamic development of action and thought. In W. Damon, R. M. Lerner, & R. M. Lerner (Eds.). *Handbook of child psychology. Vol. 1: Theoretical models of human development* (pp. 313–399). (6th ed.). Hoboken, NJ: Wiley.

Gibson, E. J., & Pick, A. D. (2000). *An ecological approach to perceptual learning and development.* USA: Oxford University Press.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159–174. https://doi.org/10.2307/2529310.

Lehrer, R., & Schaube, L. (1998). Reasoning about structure and function: Children's conceptions of gears. *Journal of Research in Science Teaching, 35*, 3–25. https://doi.org/10.1002/(SICI)1098-2736(199801)35:1%3c3::AID-TEA2%3e3.0.CO;2-X.

Little, T. D. (2013). *Longitudinal structural equation modeling.* New York: Guildford Press.

Martin, T., & Schwartz, D. L. (2005). Physically distributed learning: Adapting and reinterpreting physical environments in the development of fraction concepts. *Cognitive Science, 29*, 587–625. https://doi.org/10.1207/s15516709cog0000_15.

Muthén, L.K., & Muthén, B.O. (1998–2017). Mplus User's Guide. Eighth Edition. Los Angeles, CA: Muthén & Muthén.

Ostini, R., & Nering, M. L. (2006). *Polytomous item response theory models.* Thousand Oaks, CA: Sagehttps://doi.org/10.4135/9781412985413.

Pouw, W. T., Van Gog, T., & Paas, F. (2014). An embedded and embodied cognition review of instructional manipulatives. *Educational Psychology Review, 26*(1), 51–72. https://doi.org/10.1007/s10648-014-9255-5.

Rambusch, J., & Ziemke, T. (2005). The role of embodiment in situated learning. July *Proceedings of the 27th Annual Conference of the Cognitive Science Society,* 1803–1808.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime user's guide.* Pittsburgh: Psychology Software Tools Inc.

Shrager, J., & Siegler, R. S. (1998). SCADS: A model of children's strategy choices and strategy discoveries. *Psychological Science, 9*(5), 405–410. https://doi.org/10.1111/1467-9280.00076.

Siegler, R. S. (1976). Three aspects of cognitive development. *Cognitive Psychology, 8*, 481–520. https://doi.org/10.1016/0010-0285(76)90016-5.

Siegler, R. S. (1996). *Emerging minds. The process of change in children's thinking.* New York: Oxford University Press.

Siegler, R. S., & Chen, Z. (2002). Development of rules and strategies: Balancing the old and the new. *Journal of Experimental Child Psychology, 81*(4), 446–457. https://doi.org/10.1006/jecp.2002.2666.

Siegler, R. S. (2006). Microgenetic analyses of learning. In W. Damon, R. M. Lerner, D. Kuhn, & R. S. Siegler (Eds.). *Handbook of child psychology: Vol. 2: Cognition, perception, and language* (pp. 464–510). (6th ed.). . https://doi.org/10.1002/9780470147658.chpsy0211.

Stephen, D. G., Boncoddo, R. A., Magnuson, J. S., & Dixon, J. A. (2009). The dynamics of insight: Mathematical discovery as a phase transition. *Memory & Cognition, 37*(1132-), 1149. https://doi.org/10.3758/MC.37.8.1132.

The National Research Council (2014). *STEM integration in K–12 education status, prospects, and an agenda for research.* The National Academic Press.

Trudeau, J. T., & Dixon, J. A. (2007). Embodiment and abstraction: Actions create relational representations. *Psychonomic Bulletin & Review, 14*, 994–1000. https://doi.org/10.3758/BF03194134.

Van Der Steen, S., Steenbeek, H., Van Dijk, M., & Van Geert, P. (2014). A process approach to children's understanding of scientific concepts: A longitudinal case study. *Learning and Individual Differences, 30*, 84–91. https://doi.org/10.1016/j.lindif.2013.12.004.

Van der Ven, S. H., Boom, J., Kroesbergen, E. H., & Leseman, P. P. (2012). Microgenetic patterns of children's multiplication learning: Confirming the overlapping waves model by latent growth modelling. *Journal of Experimental Child Psychology, 113*, 1–19. https://doi.org/10.1016/j.jecp.2012.02.001.

Van Geert, P., & Van Dijk, M. (2002). Focus on variability: New tools to study intra-individual variability in developmental data. *Infant Behavior and Development, 25*, 340–374. https://doi.org/10.1016/S0163-6383(02)00140-6.

Weisberg, S. M., & Newcombe, N. S. (2017). Embodied cognition and STEM learning: Overview of a topical collection in CR: PI. *Cognitive Research, 2*(1), https://doi.org/10.1186/s41235-017-0071-6.

Zheng, Y., & Fischer, K. (2002). Always under construction: Dynamic variations in adult cognitive microdevelopment. *Human Development, 45*, 141–160. https://doi.org/10.1159/000057070.