

Review

Lessons Learned from 10 Experiments That Tested the Efficacy and Assumptions of Hypothetical Learning Trajectories

Arthur J. Baroody^{1,*}, Douglas H. Clements²  and Julie Sarama²¹ College of Education, University of Illinois Urbana, Champaign, IL 61820, USA² Development and Research in Early Math Education, University of Denver, Denver, CO 80208, USA; douglas.clements@du.edu (D.H.C.); julie.sarama@du.edu (J.S.)

* Correspondence: baroody@illinois.edu

Abstract: Although reformers have embraced learning trajectories (LT, also called learning progressions) as an important tool for improving mathematics education, the efficacy and assumptions of LT-based instruction are largely unproven. The aim of a recently completed research project was to fill this void. Fulfilling this aim was more challenging than many supporters of LT-based instruction might imagine. A total of 10 experiments were undertaken, of which 5 demonstrated that LT-based instruction was significantly more efficacious than a counterfactual involving either a Teach-to-Target/Skip-Level approach (Assumption 1) or the same unordered activities (Assumption 2). The results of the remaining studies were non-significant either for theoretical (2) or methodological (3) reasons. In the five indicating LTs' efficacy, we found that some LTs consists of levels that are facilitative conditions for the next higher level and, thus, may be helpful but perhaps not necessary for the subsequent level.

Keywords: cardinality; early childhood; (hypothetical) learning trajectories; learning progressions



Citation: Baroody, A.J.; Clements, D.H.; Sarama, J. Lessons Learned from 10 Experiments That Tested the Efficacy and Assumptions of Hypothetical Learning Trajectories. *Educ. Sci.* **2022**, *12*, 195. <https://doi.org/10.3390/educsci12030195>

Academic Editors: Colby Tofel-Grehl, Beth L. MacDonald, Kristin A. Searle and James Albright

Received: 22 December 2021

Accepted: 25 February 2022

Published: 10 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

A hypothetical learning trajectory (HLT) is an extension of a learning progression or learning trajectory (LT) that also includes instructional goals and activities [1]. Specifically, HLTs in mathematics education consist of three components [2–4]:

1. A *goal* is the target developmental level. Goals are based on the structure of mathematics, societal needs, and research on children's thinking about and learning of mathematics and require input from experts in mathematics, mathematics education, educational policy, and developmental psychology [5–7].
2. A *developmental progression* is a sequence of theoretically and research-based increasingly sophisticated patterns of thinking that most children pass on the way to achieving the goal or target. Theoretically, each level serves as a foundation for successful learning of subsequent levels.
3. *Instructional activities* include theory and research-based curricular tasks and pedagogical strategies designed explicitly to promote the development of each level.

The conventional wisdom in the mathematics education community holds that HLTs are an important tool in improving mathematics education. Indeed, it may seem obvious that instruction (a) should promote lower levels of knowledge to lay the foundation for a goal at higher level compared to teaching to a target (focusing directly on a goal) and (b) is more efficacious than using a project approach that entails instructional activities without regard to developmental order.

Such assumptions are consistent with the conclusions of an Institute of Education Sciences (IES) Practice Guide: Teaching math to Young Children [8]. The purpose of the IES Practice Guide was to review the research literature and make instructional recommendations based on this evidence and expert opinions. Frye et al. found moderate evidence

for their first recommendation, “Teach number and operations using a developmental progression”. Moderate evidence was defined as multiple studies with “high internal validity but moderate external validity (i.e., studies that support strong causal conclusions, but generalization is uncertain) or high external validity but moderate internal validity” (i.e., studies that support the generality of a relation but the causality is uncertain; see also https://ies.ed.gov/ncee/wwc/Docs/Multimedia/wwc_pg_loe_022718.pdf accessed on 1 March 2022). Frye et al. found minimal evidence for their second recommendation, “Teach geometry, patterns, measurement, and data analysis using a developmental progression”. Minimal evidence was defined as “evidence from studies that do not meet the criteria for moderate evidence (e.g., case studies, qualitative research)”. Frye et al. concluded that there was no direct evidence that instruction based on a developmental progression was efficacious.

Citing Corcoran et al. [9], Shavelson and Karplus [10] similarly concluded that school reformers and educational researchers have embraced instruction based on a learning progression as an important but unproven tool for reform: “CCII (Center on Continuous Instructional Improvement) views learning progressions as potentially important, but as yet unproven tools for improving teaching and learning and recognizes that developing and utilizing this potential poses some challenges” (p. 5). Shavelson and Karplus warned:

Learning progressions have captured the imaginations and rhetoric of school reformers and education researchers as one possible elixir for getting K-12 education “on track” . . . Learning progressions and research on them have the potential to improve teaching and learning; however, we need to be cautious . . . The enthusiasm gathering around learning progressions might lead to giving heavy weight to one possible solution when experience show single solutions to education reform come and go.

More recently, Lobato and Walters [1] noted the empirical evidence supporting the efficacy and assumptions of learning progression/LT-based instruction is (still) surprisingly limited.

To provide such evidence, we proposed and IES funded the HLT Project, “Evaluating the Efficacy of Learning Trajectories in Early Mathematics”. Sections 2–4 summarize the project’s rationale, methods, and results, respectively. Spoiler alert: Corroborating the efficacy and basic assumptions of HLT-based instruction was challenging. Section 5 discusses theoretical reasons for our inconsistent findings and underscores why developing and utilizing the potential of HLT-based instruction is challenging, and Section 6 focuses on methodological reasons for some findings, with implications for future research projects. Section 7 summarizes our conclusions.

2. Rationale of the HLT Project

2.1. Goals

The overarching goal of our HLT Project was to rigorously evaluate the efficacy of using LTs as a curricular and pedagogical tool and the key assumptions on which HLT-based instruction is based. To ensure the findings were generalizable, we conducted multiple experiments across various mathematical topics and age groups. We studied the preschool and kindergarten ages because HLTs are particularly important for early childhood mathematics education for several interrelated reasons. One is that early childhood educators too often have minimal, if any, training on mathematics development and education. As a result, they frequently underestimate young children’s (informal) mathematical knowledge, mechanically teach the lessons specified in a curriculum guide or textbook, and focus only on the most basic numeracy content, which many, or even most, children have already learned [11–15]. Indeed, because of a negative disposition towards mathematics instructions, many early childhood teachers do not set any mathematical goals, use any mathematical curriculum or resources, and rely on (hit-or-miss) opportunities that emerge from children’s play or routine activities [13,16–20]. The pedagogical knowledge and learning expectations of teachers of academically at-risk children are particularly unlikely

to foster numeracy [21–25]. Another reason is that LTs have been well developed for the preschool–kindergarten age range.

We now turn to the two key assumptions of HLT-based instruction.

- **Assumption 1.** *Instruction in which LT levels are taught consecutively (e.g., for children at level n , using instructional activities to foster level $n + 1$ and then $n + 2$ before instruction on a goal or target-level knowledge at level $n + 3$) results in greater learning than instruction that immediately and solely targets level $n + 3$ (or higher levels), namely the “Skip-Level” or “Teach-to-Target” approach.*
- **Assumption 2.** *Instruction aligned with an LT sequence results in greater learning than instruction that either uses a traditional curriculum’s activities and sequence (business as usual) or uses the same activities as those of the LT but chosen and ordered to fit a theme-based project.*

The arguments pro and con for each assumption and the evidence regarding the efficacy of HLT-based instruction are addressed in turn.

2.1.1. Assumption 1

The first assumption is that instruction should move children from their present level to the next higher level and continue in this manner until the instructional goal is reached. Proponents of traditional didactic instruction (a Teach-to-Target approach) continue to argue that teaching to a skill—direct instruction and drill of target knowledge—is the most mathematically rigorous and efficient way to ensure accurate target-level knowledge (see [7,26–30]). Such an approach avoids promoting the informal and error-prone strategies of lower levels and the slow movement through these lower levels. An example of this “Teach-to-Target” approach is the “worked examples” method—explicitly describing and illustrating how to solve a new type of problem, including the why (conceptual rationale) for each step [31–33]. Some evidence supports the Teach-to-Target approach [27,34–37], although the research designs often do not include other research-validated approaches.

In contrast, those interested in educational reform have long recommended building on prior knowledge as a means overcoming limitations of rote memorization engendered by traditional, didactic instruction. For example, in his 1892 “Talk to Teachers”, the eminent psychologist William James [38] advocated meaningful memorization:

“When we wish to fix a new thing in a pupil’s [mind], our . . . effort should not be so much to *impress* and *retain* it as to *connect* it with something already there . . . If we attend clearly to the connection, the connected thing will . . . likely . . . remain within recall”. (pp. 101–102)

In a similar vein, Piaget [39] argued that “the fundamental relation from the point of view of pedagogical . . . application” is not associations, but assimilation: “the integration of any sort of reality into [an existing] structure”(p. 16).

Since the late 19th century, when research on the development of mathematical knowledge exploded, educational reformers became increasingly interested in developing, promoting, and using such an approach [40]. A basic assumption for using HLT-based instruction is that it is more efficacious than teaching a target-level competence directly [2,8].

Consider, for example, the classic example of achieving fluency with basic sums such as $3 + 4 = 7$. Traditional didactic instruction focuses on direct imposition of the knowledge: repeated exposure and practice of the basic facts, frequently accompanied by suppression of children’s existing slow and sometimes error-prone informal strategies [41,42]. If such efforts fail to result in memorization by rote, exposure and practice are increased and the correct answer is provided if a child responds incorrectly or does not respond quickly [43].

In contrast to this one-phase approach, mathematics educators have long recommended achieving meaningful memorization using three phases [44–46]. In Phase 1, children are encouraged to develop efficient counting strategies to better detect patterns and relations among basic sums. In Phase 2, children next use discovered mathematical regularities to devise reasoning strategies such as the near-doubles reasoning strategy (e.g.,

$3 + 4 = [3 + 3] + 1 = 6 + 1 = 7$). This strategy builds on prior knowledge by relating an unknown near double ($3 + 4$) to a known double ($3 + 3 = 6$) and a known add-1 combination ($6 + 1 = 7$). If this effort fails, knowledge of these prerequisites would be checked, and if need be, remedied. For example, if a child was not fluent with the add-1 combination $6 + 1 = 7$, prerequisite knowledge for its fluency—number-after relations (e.g., when we count, the number after six is seven)—would then be checked. Once fluency with number-after relations was achieved, remedial instruction would next focus on encouraging a child to recognize the connection between adding 1 and the structure of the counting sequence—that is, the number-after rule for adding 1 (e.g., the sum of $7 + 1$ is the number after seven in the counting sequence or eight). In Phase 3, children achieve fluency by either automatizing reasoning strategies or internalizing families of related facts [2,47–49].

Similar to this example, HLTs can highlight developmentally appropriate and important goals (e.g., the importance of number-after relations as a basis for fluency with basic sums) and help focus instructional efforts on them. HLTs underscore how children typically develop and the need to consider what they must already know to make progress and what level of instruction is within their comprehension (e.g., a child who does not know number-after relations is unlikely to achieve fluency with add-1 combinations, let alone near doubles such as $3 + 4$) [50,51]. HLTs, then, spotlight the need for formative assessment to determine where children are developmentally on a progression, so that instruction can target their learning needs with meaningful and effective learning tasks. For these reasons and more, researchers, educators, and policy makers have recommended HLTs as a useful tool for teachers in helping them to understand, promote, and assess children's mathematical learning [2,4,8,47,52,53].

2.1.2. Assumption 2

The second assumption of an LT approach is that there is a sequence of such levels of learning and teaching that is determined by research-based developmental progressions and that instruction is more efficacious if it promotes each level in turn. Postulating that each level of knowledge builds hierarchically on the concepts and processes of the previous levels stands in contrast to some traditional early childhood curricular organizations: theme, project, and emergent approaches [54–59]. In these approaches, a theme (e.g., “colors”), a project (e.g., visiting an apple orchard and making applesauce), or an emergent issue (e.g., building a bus when children expressed interest in buses spontaneously) determines the sequencing of activities. For example, if the theme is colors, children are asked to sort by color; if it involves apples, children might count the seeds in an apple or cut them and talk about “halves”. Thus, the activity is chosen for its fit to the classroom work, which is ostensibly more meaningful and connected for the child and thus will lead to greater learning.

In *Experience and Education*, Dewey [60] summarized the lessons he learned from his own efforts to reform education. He argued that instruction cannot simply consist of a hodgepodge of activities without clear educational purposes. Teachers must strive to provide *educative experiences* (experiences that lead to worthwhile learning or a basis for later learning), not *mis-educative experiences* (activities for the sake of activity and that may impede development). According to Dewey's “principle of interaction,” educative experiences result “from an interaction of external factors (e.g., the nature of the subject matter and teaching practices) and internal factors (e.g., a child's developmental readiness and interests). Unless a theme, project, or emergent issue is carefully chosen and developed with important goals and students' range of developmental levels in mind, instruction may violate Dewey's principle of interaction and, thus, be inefficient, ineffective, or even detrimental. Some, many, or even most children may not be developmentally ready or developmentally too advanced for the instruction. Although careful integration of mathematics into daily routines and instruction in other areas can be valuable, doing so without regard to the mathematical goals and developmental progressions of an LT may be

mis-educative. Although children's interest should guide instructional decisions, children's interests are malleable, and teachers can inspire new interests.

2.2. Existing Evidence of Efficacy and Its Limitations

Before we began the HLT Project, the following critical question had yet to be answered causally: "Which approach, HLT-based, Teach-to-Target, or Theme/Project/Emergent-based results in better mathematical outcomes for preschool children?" Although LT-based instruction is often recommended as a valuable educational tool, there is surprisingly little empirical support for this belief or its underlying assumptions [1,8]. Most research has focused on empirically validating the developmental levels of HLTs by using a cross-sectional methodology or tracking the progress of individuals over time (e.g., [61–64]). Relatively little research has involved closely examining the impact of instructional scaffolding on children's movement along an HLT compared to not doing so.

Moreover, although considerable research has shown interventions that have HLTs as a component are efficacious in promoting numeracy, little research has directly or systematically examined their unique contribution or assumptions [8]. For instance, a preschool curriculum based on HLTs promoted numeracy significantly more than did business-as-usual instruction (effect size, 1.07) or an intervention organized by mathematical topics (effect size, 0.47 [65]). Although the HLT and topically based interventions were closely matched in terms of content and superior performance of the former might be due to using an HLT, the two curricula had other differences (e.g., different activities and integrated versus discrete content) that might account for the performance difference.

3. Methods

The HLT Project entailed scientific and rigorous tests not heretofore conducted on the HLT construct by designing experiments that had the following three characteristics:

1. Ensured causal interpretation of the findings via Randomized Control Trials.
2. Ensured a control group received an intervention that was as similar as possible to the HLT intervention, except for a single defining attribute of the HLT construct.
3. Identified each participant's location on a LT at pretest and ensured an equivalent baseline for posttest comparison of interventions on the dependent measure(s).

3.1. Research Design to Test Assumption 1

To test the assumption that progressively teaching one level above a child's existing level on an LT should be more efficacious than skipping a level and directly teaching to the target level, seven experiments were undertaken that involved a comparison of a LT-based instruction and a control group, which received the same target-level instruction but skipped prior levels.

3.2. Research Design to Test Assumption 2

To test the assumption that presenting instruction in the developmental order hypothesized by a LT presumably matters, three experiments involved comparing an experimental group received LT-based training (activities ordered by an LT) with a counterfactual group, which, involved the same activities but not ordered by a LT and (typically) a business-as-usual (BAU) control group, which received only classroom experiences.

4. Results

4.1. Results for Assumption 1

Table 1 shows that the results of the seven experiments that evaluated Assumption 1 produced different results. Four found that progressively teaching one level above a child's existing level on an LT was more efficacious than skipping a level and directly teaching to the target level: Experiment 3 [66], Experiment 4 [67], Experiment 7 [68], and Experiment 10 [69]. Unpublished Experiments 1, 2, and 9 had a positive impact but not above and beyond the Teach-to-Target intervention) due to methodological problems.

The results of Experiments 3, 4, and 7 indicate that the LT-based instructional approach is efficacious in various ways. For example, Experiment 3 indicated that children in the Teach-to-Target group showed an aversion to math, whereas the HLT-taught children exhibited engagement. HLT participants in Experiments 4 and 7 involving arithmetic showed growth not only in correct answers but also in the use of more sophisticated strategies. Indeed, although the Teach-to-Target intervention in the Experiment 7 had a heavier dosage of target-level arithmetic instruction, the HLT-based instruction produced significantly and (as measured by effect size) substantially more accurate solution at and above the target level. This is striking in that the counterfactual children spent *all* their instructional time at the target level, far more than the HLT children. Nevertheless, the HLT children scored higher on items measuring that level (and those measuring levels above) than the counterfactual children.

4.2. Results for Assumption 2

As Table 1 shows, the results of the three experiments that evaluated Assumption 2 similarly produced mixed results. Experiment 8 [70] indicated that using activities ordered by an HLT was more efficacious than with using the same, albeit unordered, activities. That is, the experiments testing Assumption 1 used mostly different activities. However, the child in these experiments experiences the same activities. Thus, the results specifically showed the importance of following the developmental progression.

However, Experiment 5 [71] and Experiment 6 [72,73] found that the HLT-based intervention produced significant learning but not significantly better than that involving the same unordered activities. The next two sections discuss possible reasons for the mixed results.

5. Discussion of Theoretical Issues

Why—despite our own belief in LT-based instructional approaches—was it so difficult to corroborate the efficacy and underlying assumptions of such an approach? Four theoretical factors might account for the inconsistent results.

5.1. Nature of the Relation between Successive Levels

Earlier levels in an HLT may support later levels either by facilitating the latter or serving as a developmental prerequisite (a necessary condition) for the target knowledge. As an example of a developmental prerequisite, consider two concepts in object counting: The *count-to-cardinal* concept, also known as the cardinality principle (CP), entails understanding that the last number word said when counting a set indicates the total number of items in that set (e.g., counting a set of five blocks as “one, two, three, four, five” and recognizing that there are “five” blocks in all). The *cardinal-to-count* concept serves as the conceptual basis for counting out a specified number of items: to produce a given quantity, count object objects to that number. Fuson [74] hypothesized that count-to-cardinal concept (or CP) serves as developmental prerequisite for the cardinal-to-count concept: indicates that a cardinal label of a set such as “five” indicates what the last number word would be if the set were counted. In essence, the cardinal-to-count concept is the inverse of the count-to-cardinal concept and serves as the rationale for the counting-out procedure. For instance, “five” in the request “give me five blocks” specifies that the counting-out process should stop when the count reaches five. With facilitative relations, a messy middle can be expected. That is, though success on an earlier facilitative level increases the probability of success on a later target knowledge, knowledge of any one facilitator may or may not be evident before the target knowledge emerges. With modest facilitators particularly, a child might skip one or even more levels, or *appear* to do so, and still learn higher target knowledge. This might account for why, in Experiment 5 [71] and Experiment 6 [72,73], the experimental intervention based on an HLT resulted in significantly improved patterning knowledge but not significantly better than the counterfactual intervention, which involved

the same unordered activities. Teaching the levels in order was *not* crucial for promoting an advanced level of patterning knowledge.

For Experiment 3 [66], Experiment 4 [67], Experiment 7 [68], and Experiment 8 [70], the experimental intervention based on an HLT resulted in significant improvement above and beyond that of the counterfactual. Nevertheless, in these experiments, some control participants achieved success on the target level without instruction on precursor levels. Such results are consistent with an HLT that embodies strong facilitators, but not prerequisites, for the target knowledge. (Both prerequisite or necessary and facilitative relationships are postulated by Hierarchical Interactionalism [2].)

In Table 1, the evidence indicated that the HLT in Experiment 7 was nearly a necessity for kindergartners with lowest entry level. These results suggest that the earliest levels in the HLT are more critical than later levels and probably unwise to skip and/or that the greater the distance between a child and the target level, the more important is the adjustment of instruction to the child's level.

For Experiment 10 [69], the HLT involved a hypothesized conceptual prerequisite for a target concept and skill. With one exception (described in the fourth bullet below), participants pretested at a level below the conceptual prerequisite had negligible or no success on target tasks. The HLT-based experimental intervention resulted in significant improvement on both conceptual and procedural fluency dependent measures above and beyond the improvement of the counterfactual (Teach-to-Target intervention). Specific findings include

- Five of the seven participants who received the HLT-based intervention, which included prior training on the conceptual prerequisite, had (some) success on the target-concept measure; six of seven, on the target procedural-fluency measure.
- The one HLT participant who was unsuccessful on both the conceptual and the procedural-fluency task had negligible success learning the conceptual prerequisite.
- Seven of the eight participants who were trained on the target concept and skill but not the prerequisite concept had (almost) no success learning the target knowledge.
- Finally, post hoc analysis indicated that the exceptional Teach-to-Target participant who mastered both the target concept and skill not only exhibited the best pretest performance of the sample but appeared to have learned the prerequisite concept during the pretesting.

Overall, then, of the seven children who exhibited knowledge of the prerequisite concept before the target training, six appeared to benefit from the target-level training and exhibited some success on the measure of target understanding (see Table 2). Of the eight children who did not exhibit knowledge of the prerequisite concept before the target training, the target-level training resulted in no success on the measure of target understanding in seven (Teach-to-Target) cases and negligible success in another (HLT participant). The corresponding results for the target skill were all seven prerequisite knowers achieved (some) success, whereas seven not-knowers had no success and one had minimal success (see Table 3). The lack of a messy middle is strongly consistent with prerequisite knowledge involving a necessary relation and, in such cases, instructional order (including not skipping the lower level) is important.

Table 1. Summary of the Research for the HLT Project.

Experiment: Domain	Published	Assumption	Method	Statistically and Practically Significant ^b	Reason for Non-Significance	Relation
Experiment 1 (<i>n</i> = 76 preschoolers): Counting/subitizing/ cardinality	-	1	LT vs. TtT/Skip	No	Methodology	-
Experiment 2 (<i>n</i> = 180 pre-K): Counting/subitizing/ cardinality ^a	-	1	LT vs. TtT/Skip	No	Methodology	-
Experiment 3 (<i>n</i> = 152 preschoolers): Shape composition	[66]	1	LT vs. TtT/Skip	Yes ES = 0.55 <i>p</i> = 0.016	-	Not necessary but strongly facilitative
Experiment 4 (<i>n</i> = 26 kindergartners): Addition and subtraction	[67]	1	LT vs. TtT/Skip	Yes Multiple qualitative indicators of a gain of 24% or greater	-	Not necessary but strongly facilitative
Experiment 5 (<i>n</i> = 16 preschoolers): Patterning pilot	[71]	2	LT vs. Unord	No ES = 0.238 for main variable, <i>p</i> = 0.48	Type of relation, faulty LT	Somewhat facilitative
Experiment 6 (<i>n</i> = 48 preschoolers): Patterning	[72,73]	2	LT vs. Unord	No Unord scored higher on some measures, <i>ns</i>	Type of relation, faulty LT	Somewhat facilitative
Experiment 7 (<i>n</i> = 291 kindergartners): Early arithmetic	[68]	1	LT vs. TtT/Skip	Overall: Yes; small for those with highest entry level Target: Yes	-	Not necessary but facilitative; near necessary for those with lowest entry level
Experiment 8 (<i>n</i> = 189 kindergartners): Length measurement	[70]	2	LT vs. REV vs. BAU	Yes/No ES = 0.32 (LT vs. REV)	-	Not necessary but highly facilitative
Experiment 9 (<i>n</i> = 20 preschoolers): Cardinality	-	1	LT vs. TtT/Skip	No	Methodology	-
Experiment 10 (<i>n</i> = 15 preschoolers): Cardinality	[69]	1	LT vs. TtT/Skip	Yes ES = 1.3, <i>p</i> = 0.032 (procedural fluency) ES = 1.68 <i>p</i> = 0.016 (conceptual understanding)	-	Necessary (or necessary and sufficient?)

Note. LT = HLT-based intervention. TtT/Skip = Teach-to-Target/Skip Level(s) counterfactual; Unord = same but unordered instructional activities counterfactual; BAU = business-as-usual (passive) control condition. ^a The 180 children were assigned to one of the three sub-experiments depending on their initial (pretest) level of development. ^b Slavin and Smith [75] caution that effect sizes for small-*n* studies, such as Experiments 4, 5, and 11, are more variable than those of large-*n* studies. Thus, the former produce less reliable and replicable estimates of program impact than the latter. They further note that the most important source of this greater variability may be what Cronbach et al. [76] call “superrealization”. Superrealization refers to high implementation fidelity due to better monitoring and more input by experimenters than would be available at scale. Slavin and Smith conclude that, although this variable may not impact internal validity, it can appreciably affect external validity.

Table 2. Knowledge of the hypothesized prerequisite (count-to-cardinal) concept before training \times posttest performance on the target (cardinal-to-count) concept in Experiment 10.

		Understanding of Target Concept at Posttest	
		No	Yes
Knowledge of the prerequisite (count-to-cardinal) concept before target training on the target (cardinal-to-count) concept	Yes	A 1	B 7
	No	C 8	D 0

Table 3. Knowledge of the hypothesized prerequisite (count-to-cardinal) concept before training \times posttest performance on the target (counting-out) skill in Experiment 10.

		Fluency of Target Skill at Posttest	
		No or Little	Modest or Good
Knowledge of the prerequisite (count-to-cardinal) concept before target training on the target (counting-out) skill	Yes	A 0	B 8
	No	C 8	D 0

5.2. Qualitative Differences between Successive Levels

Even with a succession of prerequisite levels that involve necessary conditions, if two successive levels are highly similar, children may spontaneously construct the higher level from the lower level learned with the support of instruction. That is, with little or no external help, students may generalize learning to the next level. Achieving the lower level (via instruction) may effectively be a necessary and sufficient condition for achieving the higher level. Alternatively, children might spontaneously construct a lower but “skipped” level as they learn the level higher with the support of instruction, “filling in” the knowledge of the skipped level [2,77]. In such cases, skipping instruction on the next level and focus on the next higher level would be warranted at least for some students.

5.3. Number of Paths to Target Knowledge

Various scholars have questioned whether there is a single path for all key ideas—whether an HLT can be considered the only or even the best path to a goal [2,5,77,78]. For example, Lesh and Yoon [79] proposed that some knowledge domains might be characterized as the diametrically opposite of a linear, ladder-like LT, namely a web of knowledge. With multiple pathways of facilitators, the middle ground between initial knowledge and the target knowledge can be especially messy.

5.4. Validity of the LT

Some domains such as early patterning have been researched less than other domains such as counting, number, and arithmetic development. Thus, the relations among levels of knowledge or thinking of the former are less clear than those of the latter. Experiment 5 [71] entailed evaluating the LT for early knowledge of for repeating patterns summarized in Figure 1.

Levels of HLT	Explanation
<i>Level 1: Intuitive Patterner</i>	Implicitly detect and may even use simple repeating patterns but may not recognize alternating or sequential patterns explicitly or accurately (e.g., “I’m wearing a pattern when wearing a striped shirt of two alternating colors or a multicolored shirt with no repeating pattern).
<i>Level 2: Perceptual Patterner</i>	Operates first with concrete AB repeating patterns (e.g., ●●●●■●●●●...) and then with more complex patterns such as ABB (e.g., ●●●●●■●●●●■●●●●) or ABC patterns (e.g., ●●▲●●▲●●▲●●▲...).
<i>Sublevel 2a: Pattern Fixer and Duplicator</i>	Can find the missing element of a repeating pattern and fix it (e.g., given ■●●●■●●●_ or ■●●●■_■●, fills in the blank with a circle). Can copy a pattern model such square, circle, square, circle, square, circle....
<i>Sublevel 2b: Pattern Extender</i>	Can continue a repeating pattern (e.g., given square, circle, square, circle, square, circle, adds on a square and circle and another square and circle).
<i>Level 3: Generalized Patterner</i>	Can abstract a pattern and translate it into new media (e.g., given a repeating pattern such as ■●●●■●●● can translate into the movements up, down, up, down, up down; the colors red, blue, red, blue, red, blue), or the letters ABABAB.
<i>Level 4: Unit Patterner</i>	Can identify the core of a repeating pattern (the smallest portion of the pattern that repeats to create the rest of the pattern).

Note. The portion of the HLT shown is relevant to repeating patterns only (not, e.g., growing patterns) and is hypothesized to apply to the simplest repeating (AB patterns) first and then more complex ones (ABB or ABC patterns).

Figure 1. A Modified Version of the HLT for Initial Patterning Instruction.

One unresolved question of particular interest was: Where should translating a repeating pattern into letters fit in a patterning HLT? Logically, such a competence fits the definition of Level 3 (Children can abstract a pattern and translate it into new media), which in Sarama and Clements’ [2] original LT was combined with Level 4 (Children can identify the core of a repeating pattern (the smallest portion of the pattern that repeats to create the rest of the pattern)). For the LT training, then translating repeating patterns was postponed until after participants received Level-2 training. Interestingly, two popular early childhood mathematics curricula—Building Blocks [80] and Mathematics Their Way [81]—regularly use letters to label patterns from the beginning of patterning instruction. This approach was used in the counterfactual training.

Baroody et al. [71] observed that children in both conditions struggled mightily with translating patterns into different materials (e.g., translating the circle-square-circle-square-circle-square pattern depicted above into triangles-hexagon-triangle-hexagon-triangle-hexagon or—in a few cases—even a circle-square repeating pattern involving different colors). In contrast, they quickly learned to translate repeating patterns into letters (e.g., translating ■●■●■● into the plastic alphabet letters: ABABAB). Using letters to label the elements of a pattern or its core, then, seems to be a distinct form of translating patterns—qualitatively different from translating a pattern into other objects (see also [72]). As a result, the counterfactual (“unordered”) intervention may have conferred two advantages:

1. The early use of letters to label the elements of a pattern may have fostered the Level-2 competencies (e.g., extending a repeating pattern) by counterfactual participants.
2. Early use of letters to label the core of a pattern may have helped some such participants achieve Level-4 competence (identifying the core of a repeating pattern). (Par-

enthetically, translating a pattern into different objects (listed as a Level-3 competence in Figure 1) may be more challenging and facilitated by an explicit understanding of the concept of a core unit (listed as a Level-4 competence in Figure 1). This conjecture is consistent not only with Baroody et al.'s [71] observations but with Fyfe et al.'s [82] finding that using letters to identify unit cores was efficacious in promoting the ability to translate a pattern into different objects. Although an implicit consideration of unit may naturally help some children to translate a repeating pattern into different materials, more explicit instruction that entails systematic instruction that first involves using letters to label the elements of a pattern (Level-2) and then the core of a pattern may provide a better basis for most children to tackle this challenging task.)

Yilmaz et al. [73] reported eye-tracking data that indicated Level-2 children *implicitly* attend to the core when, say, extending a pattern and only later construct the explicit knowledge that permits success on the core-identification task used to assess Level 4. That is, experiences constructing Level 2 implicitly draws attention to the core and can facilitate explicit attention to the core during Level-4 training whether conducted simultaneously or afterward. So, another reason for indistinct impact of HLT-based instruction and instruction based on the same unordered activities is that existing patterning LTs, such as that in Figure 1, may have been based on incomplete information—on research that did not adequately examine children's implicit patterning knowledge.

6. Discussion of Methodological Issues

Another barrier to confirming the efficacy of HLT-based instruction and its assumptions are the methodological challenges of such research. We first discuss five general challenges and then illustrate these issues with a description of our efforts to study a particular domain (early cardinality development). These are by no means the only challenges. However, we believe that their explication may increase the quantity and quality of future research.

6.1. General Methodological Challenges

6.1.1. Issues with the Starting Level

When evaluating the efficacy and assumptions of HLT-based instruction, careful attention must be paid in identifying a participant's starting developmental level, ensuring enough participants are at an appropriate starting level to achieve significant statistical power, and equating the learning conditions on this variable. For example, Baroody et al. [71] reported that, unlike type of intervention, starting level was significantly related to learning the target knowledge (core identification). The two HLT and three unordered participants who exhibited partial Level-2 competence at pretest all achieved success on the target (Level-4) task at posttest. In contrast, among participants who were at Level 1 at pretest, only three of the six HLT-like participants and one of the five non-HLT children did so. Given that Level 3 should perhaps follow Level 4, the five participants who started with partial knowledge of Level 2 were already close to the target level, whereas those who started at Level 1 were a full level away from it. With a larger sample of children who start at Level 1, then, type of intervention might have made a significant difference.

6.1.2. Sacrifice of Ecological Validity

Research requires trade-offs between internal and external validity (e.g., between controls that permit a clear conclusion and results that can be generalized to actual classrooms). The positive impact of the HLT-based instruction may be greater outside of a controlled sequence of activities used in the present project. For example, in the experiments that compared an HLT-based intervention with an intervention using the same unordered activities [70–72], the former involved a fixed sequence of activities, regardless of a child's progress. This was necessary to equate coverage and dosage and eliminate these factors as possible confounds or alternative explanations. However, HLTs are recommended as resources to support more flexible instruction based on formative assessment [2,8]. That is,

typically the use of HLTs involves immediately moving to the next higher level once a level is attained and only after this level is attained.

6.1.3. Small Sample Size

A possible explanation for the insignificant finding of Experiment 5, for example, was the small sample insufficient power to detect a real difference. However, a follow-up with three the number of participants per group (Experiment 6) also yielded a non-significant difference [72].

6.1.4. Entangling Lower with Higher of Levels of Instruction

An analysis of Experiment 1 revealed two inter-related reasons for the lack of a significant difference between the HLT-based intervention and the Teach-to-Target intervention. One was that the target-level activities form the Building Blocks curriculum used for both interventions involved both target-level and lower-level competencies. Another is that—despite their research protocol training—the trainers naturally did what educators do, which was they help a child with both levels of competencies regardless of a child's assignment (having trainers teach only one condition would have introduced a possible confound). In effect, the two types of intervention were not clearly distinct. A lack of fidelity to both the HLT and counterfactual also plagued Experiment 2. The plan was to have 180 children starting at the same level, but the population was so diverse that children were assigned to three different levels and thus three instructional conditions. Despite additional professional development, trainers found it difficult to accurately enact the six different instructional conditions (often doing 3 or more each day with different children).

6.1.5. Imprecise Dependent Measures

The operational definition of target-level competencies needs to be precise. The dependent measures for Experiment 1 and Experiment 2 (in Table 1)—the first two efforts to examine a cardinality LT—involved tasks drawn from the TEMA-3 [83] and REMA [84]. Two 'how many?' tasks—cardinality rule with 8 items (after counting 8 items, asking a child how many) and how many pennies (after counting 8 pennies)—served to gauge prerequisite knowledge (Level 2, the count-to-cardinal concept, or CP). A give-*n* task (put 5, 7, and 10 boxes in a cart) served partly to gauged target-level knowledge (Level 4, cardinal-to-count, producing a set). Unfortunately, these tasks do not precisely measure conceptual understanding at Levels 2 and 4.

Whereas the count-to-cardinal concept or cardinality principle (CP) entails understanding that the last number word used to count a collection also indicates its total number of items, Fuson [74] observed that many children can learn the cardinality rule (stating the last number word is an acceptable response to the how many question) by rote—without recognizing that it represents the total. Thus, children successful on the cardinality rule with 8 items may or may not have constructed the hypothesized prerequisite (Level-2) knowledge for Level 4.

Children successful on the give-*n* task ('put 5 [then 7, and finally 10] boxes in a cart') almost certainly understand the Level 4 (cardinal-to-count) concept (a cardinal term such as "seven" indicates what the last number word would be if a collection is counted). This advanced cardinal concept is the basis for knowing when to stop the counting-out process (e.g., put 7 boxes in the cart, stop counting out boxes when "seven" is reached). However, the task involves executing a counting-out procedure that requires remembering the requested number, counting items as they put in the cart, comparing a count to the requested number. In brief, a child might understand the cardinal-to-count concept but respond incorrectly because of a procedural slip up.

6.2. A Case in Point: Cardinality Development

Although some researchers agree with Fuson's [74] hypothesis that ability to count out a requested number of items and its conceptual rationale (Level 4 in Table 4) should

build on earlier level of cardinal-number understanding (Levels 1 to 3 in Table 4 [8,85]; others do not [86,87].

Table 4. A possible learning progression of key aspects of pre-counting and counting-based cardinal number knowledge and their type of mapping, conceptual basis, and direct measure [8,74,85].

Aspect of Cardinal Number	Conceptual Basis	Mapping	Direct Measure
Pre-meaningful counting (verbal subitizing-based) cardinality development			
Level 1A: number recognition (n -knower levels)	Cardinal representation of a small number underlies immediate subitizing of 1, 2, or 3	Quantity-to-word (via subitizing)	How-many task
Level 1B: putting out a requested n (also commonly called n -knower levels)	Cardinal representation of small numbers used to subitize when 1, 2 or 3 have been put out	Word-to-quantity (via subitizing)	Give- n task
Counting-based cardinality development			
Level 2: cardinality-principle knower [CP-knower] level) ^a	Count→cardinal concept or cardinality principle (last number word = total)	Quantity-to-word (via counting)	How-many task
Level 3: applications of CP: Number-constancy concepts	3A. Counting-based conservation of cardinal identity: Addition or subtraction, but not irrelevant physical transformations, changes total 3B. Counting-based cardinal equivalence: Sets with same number are equal despite looking different	Quantity-to-word over a quantity transformation Comparing two quantity-to-word mappings	1. Conservation of cardinal identity 2. Cardinal equivalence
Level 4: Counting out a requested n	Cardinal→count concept (a cardinal number = the last number word used if a set is counted)	Word-to-quantity (via counting)	Predict last n word and give- n tasks

^a Meaningfully attaining Level 2 may be preceded by learning the last-word rule, which can be applied without understanding to achieve success on the how-many task.

6.2.1. Experiment 9: Lessons Learned, Part 1

To evaluate the validity of the hypothesized LT (Table 4), the lessons learned in Experiments 1 and 2 were then applied to Experiment 9. Specifically, a conservation of numerical identity task was added to check whether correct responses on the ‘how many?’ task were due to a cardinality rule learned by rote or the meaningful count-to-cardinal concept (cardinality principle). This task required a child to not only generate the cardinal number for a collection of 5 or 6 by counting but apply this outcome meaningful—to recognize whether a transformation affected the total (addition or subtraction of 1) or not (change in appearance). The scoring of the give- n task was modified to distinguish between errors that violate the cardinal-to-count concept (e.g., counting out all the available items or counting out more than requested number) and minor errors that do violate the principle.

Overview of Experiment 9. This effort entailed randomly assigning 10 participants to the HLT condition (4 boys, mean age = 3.55 years, 5 African American, 3 multiracial, 8 free/reduced lunch) and 10 in the Teach-to-Target condition (4 boys, mean age 3.8, 3 African American, 2 multiracial, 9 free/reduced lunch). An analysis revealed that both groups improved significantly and substantially at delayed posttest on the give- n task but that the HLT-Like group did not significantly improve more than the Teach-to-Target group.

Methodological issues with Experiment 9. Three issues appeared to account for the non-significant difference. Two involved the starting level. One compromising issue was that children were included in Experiment 9 regardless of how far below the target level they were developmentally. For the example, the lowest-functioning child in the experiment could not initially subitize even one and two and had trouble counting one-to-one with

collections beyond two. Despite focused remedial efforts, this HLT-assigned child did not improve on these foundational competencies. This makes sense given the relatively long time needed to construct verbal concepts of “one” and “two” [88,89]. Training on the hypothesized prerequisite cardinality concept (count-to-cardinal concept) and, thus, the more advanced target-level knowledge (cardinal-to-count concept and counting-out procedure) had no impact.

A second issue with starting level was that, at pretest, half of the children included in Experiment 9 could occasionally count out a collection of 5 or more upon request. That is, although highly inconsistent in their performance on the give- n task, they sometimes appeared to apply the cardinal-to-count concept (i.e., stopped their counting-out process at the requested number).

A third issue was assessing the cardinal-to-count concept (Level 4) directly and reliably. Despite the more lenient scoring of the give- n task, a performance failure due to the cognitive demands of implementing the counting-out procedure might still have underestimated understanding of the cardinal-to-count concept. For instance, even if a child understood the concept, the demands on attention and memory required to remember the requested number, count out items, and/or compare the count to the requested number might cause a slip up [40].

6.2.2. Experiment 10: Lessons Learned, Part 2

Building on Experiments 1, 2, and 9, Experiment 10 was undertaken.

Methodological improvements to Experiment 10. Three modifications were implemented:

First, children who could not recognize 1 and 2 were excluded from the experiment as developmentally unready.

Second, to better test the hypothesis of whether skipping a level makes a difference, only children who had not already achieved Level 2 *and* who did not have more than minimal success counting out 5 to 7 items were included.

Third, observations during the training phase of Experiment 10 suggested that a stop-at- n task, which involved asking a child to stop a Muppet’s counting-out process at the requested number, might serve as an effective measure of the cardinal-to-count concept. A child who recognizes that the requested number represent the cardinal value of the requested collection and should be the stopping point of the counting-out process (i.e., understands the cardinal-to-count concept) should be successful on the stop-at- n task. Unlike the give- n task, this task relieved children of the demands of counting out a collection themselves (minimized cognitive demands and performance failure). The stop-at- n task, then, was adopted as the dependent conceptual measure and the give- n task was retained as in the dependent procedural fluency measure in Experiment 10.

Experiment 11: Results and limitations. As noted previously, the results of Experiment 10 clearly indicated that the count-to-cardinal concept (cardinality principle) is a developmental prerequisite for the cardinal-to-count concept and counting-out collections beyond the subitizing range. Unclear is whether the prerequisite Level 2 is a necessary condition for Level-4 competencies, as hypothesized by Fuson [74], or a necessary and sufficient condition, which is essentially equivalent to Sarnecka and Carey’s [87] hypothesis that the concepts are indistinct or develop simultaneously. For a prerequisite involving a necessary condition, all the data should be distributed among cells A, B, and C with cell D = 0, as it is Table 2 [90]. For a necessary and sufficient condition, all the data should be distributed between cells B and C with cells A and D = 0, as it is in Table 3.

Aside from the conflicting results, the problem is that the sample is too small to be sure what the distribution would be in each table for the population of young children. (The COVID pandemic interrupted data collection midstream.) There is another reason Sarnecka and Carey’s [87] alternative hypothesis that the count-to-cardinal concept (CP) underlies both meaningful one-to-one counting and fluency with counting out a specified number of items (i.e., that the count-to-cardinal and the cardinal-to-count concepts are indistinct) cannot be discounted. According to this alternative hypothesis, the HLT participants

significantly and substantially outperformed the Teach-to-Target children, because the former received training on the count-to-cardinal concept (CP) and the latter did not and, thus, the former had a greater dosage of counting-based cardinality training overall.

One way to critically test Fuson's [74] hypothesis against Sarnecka and Carey's [87] alternative hypothesis would be to track longitudinally whether an understanding of count-to-cardinal and the cardinal-to-count concepts evolve sequentially or simultaneously. Another would be to train children who have achieved Level 1 in Table 1 (i.e., are developmentally ready for Level 2) on the count-to-cardinal concept (CP). If the count-to-cardinal concept (CP) is a necessary condition for the cardinal-to-count concept, as Fuson hypothesizes, *and* the two concepts are clearly distinct (i.e., involve a significant conceptual leap), then participants should significantly improve on the former but not the latter and skipping Level 2 to achieve Level 4 would not be an option. (Currently, there is too little evidence to determine whether the number-constancy concepts—extensions of the count-to-cardinal concept—are a necessary or facilitative condition for cardinal-to-count concept and counting out, thus it unclear whether Level 3 can be skipped.) If the count-to-cardinal concept (CP) is indistinct from the cardinal-to-count concept as Sarnecka and Carey hypothesize (i.e., the former is effectively a necessary and sufficient condition for the latter), then theoretically participants should improve on both tasks to an equal degree. If the count-to-cardinal concept is a necessary condition for the cardinal-to-count concept but the two concepts are only somewhat distinct, then the results could be messy: significantly more participants may or may not improve on the former than on the latter. If—contrary to what the present results indicate—the count-to-cardinal concept (CP) is only a facilitative condition for the cardinal-to-count concept, then Level 2 may be skippable in achieving Level 4, and some portion of the comparison group trained only Level 4 may achieve the cardinal-to-count concept. In brief, corroborating the efficacy and assumptions of HLT-based instruction, in general, and the validity Fuson's hypothesis, in particular, is challenging for both theoretical and methodological reasons.

7. Conclusions

Overall, then, the evidence of the HLT Project corroborates the efficacy and basic assumptions of an HLT-based approach. Nevertheless, as the case of testing Fuson's hypothesis [74] about cardinality development illustrates, much research still needs to be performed to evaluate whether an HLT's developmental progression consists of facilitative levels or developmental prerequisites. Except in cases where an LT involves prerequisite knowledge (i.e., at least a necessary condition) for a higher level that is qualitatively different, a messy middle can be expected, and some children without a lower level of knowledge can be expected to achieve a higher level of knowledge. This is consistent with the theory upon which the LTs examined in the HLT Project are based, which does not require that levels are prerequisites to be educational useful (and that different types of developmental progressions exist [2]). Further, the theory recognizes that some children can learn multiple contiguous levels of thinking in parallel. However, recall that the HLT in Experiment 7 was nearly necessary for kindergartners with the lowest entry level. This implies that at least for some children learning some topics, the greater the distance between a child and the target level, the more important the adjustment of instruction to the child's level.

Different topics may have quite distinct conceptual structures. For example, consider Rittle-Johnson et al.'s [91] view that existing LTs for early knowledge of for repeating patterns might better be characterized as a "construct map"—a probabilistic continuum of knowledge rather than distinct phases of knowledge. Indeed, in our patterning experiments (Experiments 5 and 6), the evidence indicates a series of facilitative factors, refinement of the patterning HLT (Figure 1), and methodology may yet identify one or more prerequisite levels for later levels.

Further, Hierarchical Interactional theory [2] states that HLTs are hypothetical in two ways. First, they must be realized with teachers and children. Second, they should

continually be improved based on new information. Therefore, we interpret the null results of the patterning studies (Experiments 5 and 6) as a valid caution that an LT approach is only as good as the LT it uses. However, our analyses also indicate that the null results were due to faults not so much in the LT approach itself but in the LT (which has already been substantially revised, e.g., see LearningTrajectories.org). The more research on a given topic, the more valid future versions of that topic's LT.

In summary, creating and evaluating HLTs are challenging but worthwhile tasks, as the HLT Project illustrates. Even the three Assumption 1 experiments that did not significantly favor the HLT-based instruction for methodological reasons, and thus cannot be considered a valid test, served as pilot studies to work out the intricate methodology needed for the successful corroboration of the cardinality LT (in Table 4). HLT Project results also indicate that the benefits justify meeting the challenges.

Author Contributions: All authors contributed equally to the conceptualization of this manuscript. Writing—original draft preparation, A.J.B.; writing—reviewing and editing, D.H.C. and J.S.; funding acquisition, D.H.C., J.S. and A.J.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by U.S. Department of Education, Institute of Education Sciences), Grant No. R305A150243.

Institutional Review Board Statement: The Experiment was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board of the University of Denver (766112-16, 2/5/2019) for studies involving humans.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the Experiment.

Data Availability Statement: Data for the experiments, after all planned research is published, will be processed by the Scholarly Commons office of the University of Denver Library system and shared on the Inter-University Consortium for Political and Social Research (ICPSR) system.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lobato, J.; Walters, C.D. A taxonomy of approaches to learning trajectories and progressions. In *Compendium for Research in Mathematics Education*; Cai, J., Ed.; National Council of Teachers of Mathematics: Reston, VA, USA, 2017; pp. 74–101.
2. Sarama, J.; Clements, D.H. *Early Childhood Mathematics Education research: Learning Trajectories for Young Children*; Routledge: New York, NY, USA, 2009.
3. Maloney, A.P.; Confrey, J.; Nguyen, K.H. (Eds.) *Learning Over Time: Learning Trajectories in Mathematics Education*; Information Age Publishing: Charlotte, NC, USA, 2014.
4. Simon, M.A. Reconstructing mathematics pedagogy from a constructivist perspective. *J. Res. Math. Educ.* **1995**, *26*, 114–145. [[CrossRef](#)]
5. Clements, D.H.; Sarama, J. Learning trajectories in mathematics education. *Math. Think. Learn.* **2004**, *6*, 81–89. [[CrossRef](#)]
6. Fuson, K.C. Pre-K to grade 2 goals and standards: Achieving 21st century mastery for all. In *Engaging Young Children in Mathematics: Standards for Early Childhood Mathematics Education*; Clements, D.H., Sarama, J., DiBiase, A.-M., Eds.; Erlbaum: Mahwah, NJ, USA, 2004; pp. 105–148.
7. Wu, H.-H. *Understanding Numbers in Elementary School Mathematics*; American Mathematical Society: Providence, RI, USA, 2011.
8. Frye, D.; Baroody, A.J.; Burchinal, M.R.; Carver, S.; Jordan, N.C.; McDowell, J. *Teaching Math to Young Children: A Practice Guide*; U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance (NCEE): Washington, DC, USA, 2013.
9. Corcoran, T.; Mosher, F.A.; Rogat, A. *Learning Progressions in Science: An Evidence-Based Approach to Reform*; Center on Continuous Instructional Improvement, Teachers College—Columbia University: New York, NY, USA, 2009.
10. Shavelson, R.J.; Karplus, A. Reflections on learning progressions. In *Learning Progressions in Science: Current Challenges and Future Directions*; Alonzo, A.C., Gotwals, A.W., Eds.; Sense: Rotterdam, The Netherlands, 2012; pp. 13–26.
11. Clarke, B.; Clarke, D.; Cheeseman, J. The mathematical knowledge and understanding young children bring to school. *Math. Educ. Res. J.* **2006**, *18*, 78–102. [[CrossRef](#)]
12. Engel, M.; Claessens, A.; Finch, M.A. Teaching students what they already know? The (mis)alignment between mathematics instructional content and student knowledge in kindergarten. *Educ. Eval. Policy Anal.* **2013**, *35*, 157–178. [[CrossRef](#)]

13. Ginsburg, H.P.; Lee, J.S.; Boyd, J.S. Mathematics education for young children: What it is and how to promote it. *Soc. Policy Rep.* **2008**, *22*, 3–23. [CrossRef]
14. Kilday, C.R.; Kinzie, M.B.; Mashburn, A.J.; Whittaker, J.V. Accuracy of teachers' judgments of preschoolers' math skills. *J. Psychoeduc. Assess.* **2012**, *30*, 48–158. [CrossRef]
15. Lee, J.S.; Ginsburg, H.P. What is appropriate mathematics education for four-year-olds? Pre-kindergarten teachers' beliefs. *J. Early Child. Res.* **2007**, *5*, 2–31. [CrossRef]
16. Balfanz, R. Why do we teach children so little mathematics? Some historical considerations. In *Mathematics in the Early Years*; Copley, J.V., Ed.; National Council of Teachers of Mathematics: Reston, VA, USA, 1999; pp. 3–10.
17. Hachey, A. The early childhood mathematics education revolution. *Early Educ. Dev.* **2013**, *24*, 419–430. [CrossRef]
18. Lee, J.S. Preschool teachers' shared beliefs about appropriate pedagogy for 4-year-olds. *Early Child. Educ. J.* **2006**, *33*, 433–441. [CrossRef]
19. Lee, J.S.; Ginsburg, H.P. Early childhood teachers' misconceptions about mathematics education for young children in the United States. *Australas. J. Early Child.* **2009**, *34*, 37–45. [CrossRef]
20. Li, X.; Chi, L.; DeBey, M.; Baroody, A.J. A Experiment of early childhood mathematics teaching in the U.S. and China. *Early Educ. Dev.* **2015**, *26*, 37–41. [CrossRef]
21. Ferguson, R.F. Can schools narrow the black-white test score gap? In *The Black-White Test Score Gap*; Jencks, C., Phillips, M., Eds.; Brookings Institution Press: Washington, DC, USA, 1998; pp. 318–374. ISBN 9780815746102.
22. Layzer, J.L.; Goodson, B.D.; Moss, M. Life in preschool: Observational Experiment of early childhood programs for disadvantaged four-year-olds. *Final Rep.* **1993**, *1*, ED366468.
23. Lee, J.S. Multiple facets of inequality in racial and ethnic achievement gaps. *Peabody J. Educ.* **2004**, *79*, 51–73. [CrossRef]
24. Lee, J.S.; Ginsburg, H.P. Preschool teachers' beliefs about appropriate early literacy and mathematics education for low- and middle-SES children. *Early Educ. Dev.* **2007**, *18*, 111–143. [CrossRef]
25. Lubienski, S.T.; Shelley, M.C., II. A closer look at U.S. mathematics instruction and achievement: Examinations of race and SES in a decade of NAEP data. In Proceedings of the Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL, USA, 21–25 April 2003.
26. Bereiter, C. Does direct instruction cause delinquency? Response to Schweinhart and Weikart. *Educ. Leadersh.* **1986**, *44*, 20–21.
27. Clark, R.E.; Kirschner, P.A.; Sweller, J. Putting students on the path to learning: The case for fully guided instruction. *Am. Educ.* **2012**, *36*, 6–11.
28. Kirschner, P.A.; Sweller, J.; Clark, R.E. Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educ. Psychol.* **2006**, *41*, 75–86. [CrossRef]
29. Rosenshine, B. The empirical support for direct instruction. In *Constructivist Theory Applied to Instruction: Success or Failure?* Tobias, S., Duffy, T.M., Eds.; Taylor & Francis: London, UK, 2009; pp. 201–220.
30. Rosenshine, B. Principles of instruction: Research-based strategies that all teachers should know. *Am. Educ.* **2012**, *36*, 12.
31. Clark, R.C.; Nguyen, F.; Sweller, J. *Efficiency in Learning: Evidence-Based Guidelines to Manage Cognitive Load*; Pfeiffer: San Francisco, CA, USA, 2006.
32. Renkl, A. The worked-out examples principle in multimedia learning. In *The Cambridge Handbook of Multimedia Learning*; Mayer, R.E., Ed.; Cambridge University Press: Cambridge, UK, 2005.
33. Schwonke, R.; Renkl, A.; Krieg, C.; Wittwer, J.; Alven, V.; Salden, R. The worked-example effect: Not an artefact of lousy control conditions. *Comput. Hum. Behav.* **2009**, *25*, 258–266. [CrossRef]
34. Borman, G.D.; Hewes, G.M.; Overman, L.T.; Brown, S. Comprehensive school reform and achievement: A meta-analysis. *Rev. Educ. Res.* **2003**, *73*, 125–230. [CrossRef]
35. Carnine, D.W.; Jitendra, A.K.; Silbert, J. A descriptive analysis of mathematics curricular materials from a pedagogical perspective: A case Experiment of fractions. *Remedial Spec. Educ.* **1997**, *18*, 66–81. [CrossRef]
36. Gersten, R. Direct instruction with special education students: A review of evaluation research. *J. Spec. Educ.* **1985**, *19*, 41–58. [CrossRef]
37. Heasty, M.; McLaughlin, T.F.; Williams, R.L.; Keenan, B. The effects of using direct instruction mathematics formats to teach basic math skills to a third grade student with a learning disability. *Acad. Res. Int.* **2012**, *2*, 382–387.
38. James, W. *Talks to Teachers on Psychology: And to Students on Some of Life's Ideals*; Talk Originally Given in 1892; W.W. Norton & Company: New York, NY, USA, 1958.
39. Piaget, J. Development and learning. In *Piaget Rediscovered*; Ripple, R.E., Rockcastle, V.N., Eds.; Cornell University Press: Ithaca, NY, USA, 1964; pp. 7–20.
40. Resnick, L.B.; Ford, W.W. *The Psychology of Mathematics for Instruction*; Erlbaum: Mahwah, NJ, USA, 1981.
41. Hasselbring, T.S.; Goin, L. Research Foundation and Evidence of Effectiveness for FASTT Math. 2005. Available online: <http://www.tomsnyder.com/reports/> (accessed on 16 September 2005).
42. Thorndike, E.L. *The Psychology of Arithmetic*; Macmillan: New York, NY, USA, 1922.
43. Bezuk, N.S.; Cegelka, P.T. Effective mathematics instruction for all students. In *Effective Instruction for Students with Learning Difficulties*; Cegelka, P.T., Berdine, W.H., Eds.; Allyn and Bacon: Boston, MA, USA, 1995; pp. 345–384.
44. Rathmell, E.C. Using thinking strategies to teach basic facts. In *Developing Computational Skills*; Suydam, M.N., Reys, R.E., Eds.; 1978 Yearbook; National Council of Teachers of Mathematics: Reston, VA, USA, 1978; pp. 13–50.

45. Thornton, C.A. Emphasizing thinking strategies in basic fact instruction. *J. Res. Math. Educ.* **1978**, *9*, 214–227. [CrossRef]
46. Thornton, C.A. Solution strategies: Subtraction number facts. *Educ. Stud. Math.* **1990**, *21*, 241–263. [CrossRef]
47. Baroody, A.J. Using number and arithmetic instruction as a basis for fostering mathematical reasoning. In *Reasoning and Sense Making in the Mathematics Classroom: Pre-K—Grade 2*; Battista, M.T., Ed.; National Council of Teachers of Mathematics: Reston, VA, USA, 2016; pp. 27–69.
48. National Research Council. *Adding It Up: Helping Children Learn Mathematics*; Kilpatrick, J., Swafford, J., Findell, B., Eds.; National Academy Press: Washington, DC, USA, 2001.
49. National Mathematics Advisory Panel (NMAP). *Foundations for Success: The Final Report of the National Mathematics Advisory Panel*; U.S. Department of Education: Washington, DC, USA, 2008.
50. Vygotsky, L.S. *Thought and Language*; Hanfmann, E., Vakar, G., Eds.; MIT Press: Cambridge, MA, USA, 1962.
51. Vygotsky, L.S. *Mind in Society*; Harvard University Press: Cambridge, MA, USA, 1978.
52. Baroody, A.J. Curricular approaches to introducing subtraction and fostering fluency with basic differences in grade 1. In *The Development of Number Sense: From Theory to Practice*; Bracho, R., Ed.; University of Granada: Granada, Spain, 2016; pp. 161–191. [CrossRef]
53. Butterfield, B.; Forrester, P.; McCallum, F.; Chinnappan, M. Use of Learning Trajectories to Examine Pre-Service Teachers' Mathematics Knowledge for Teaching Area and Perimeter. 2013. Available online: <https://files.eric.ed.gov/fulltext/ED572797.pdf> (accessed on 21 December 2021).
54. Broderick, J.T.; Hong, S.B. *From Children's Interests to Children's Thinking: Using a Cycle of Inquiry to Plan Curriculum*; National Association for the Education of Young Children: Washington, DC, USA, 2020.
55. Edwards, C.; Gandini, L.; Forman, G.E. *The Hundred Languages of Children: The Reggio Emilia Approach to Early Childhood Education*; Ablex: New York, NY, USA, 1993.
56. Helm, J.H.; Katz, L.G. *Young Investigators: The Project Approach in the Early Years*, 3rd ed.; Teachers College Press: New York, NY, USA, 2016.
57. Hendrick, J. (Ed.) *First Steps toward Teaching the Reggio Way*; Prentice-Hall: Upper Saddle River, NJ, USA, 1997.
58. Katz, L.G.; Chard, S.C. *Engaging Children's Minds: The Project Approach*, 2nd ed.; Greenwood Publishing Group: Westport, CT, USA, 2000.
59. Tullis, P. The death of preschool. *Scientific Amer. Mind* **2011**, *22*, 36–41. [CrossRef]
60. Dewey, J. *Experience and Education*; Collier: New York, NY, USA, 1963.
61. Barrett, J.E.; Battista, M.T. Two approaches to describing the development of students' reasoning about length: A case Experiment for coordinating related trajectories. In *Learning Over Time: Learning Trajectories in Mathematics Education*; Maloney, A.P., Confrey, J., Nguyen, K.H., Eds.; Information Age Publishing: Charlotte, NC, USA, 2014; pp. 97–124.
62. Murata, A. Paths to learning ten-structured understanding of teen sums: Addition solution methods of Japanese Grade 1 students. *Cogn. Instr.* **2004**, *22*, 185–218. [CrossRef]
63. Steffe, L.; Cobb, P. *Construction of Arithmetical Meanings and Strategies*; Springer: Berlin/Heidelberg, Germany, 1988.
64. Steffe, L.; Olive, J. *Children's Fractional Knowledge*; Springer: Berlin/Heidelberg, Germany, 2010.
65. Clements, D.H.; Sarama, J. Experimental evaluation of the effects of a research-based preschool mathematics curriculum. *Am. Educ. Res. J.* **2008**, *45*, 443–494. [CrossRef]
66. Clements, D.H.; Sarama, J.; Baroody, A.J.; Joswick, C.; Wolfe, C. Evaluating the efficacy of a learning trajectory for early shape composition. *Amer. Educ. Res. J.* **2019**, *56*, 2509–2530. [CrossRef]
67. Clements, D.H.; Sarama, J.; Baroody, A.J.; Joswick, C. Efficacy of a learning trajectory approach compared to a teach-to-target approach for addition and subtraction. *ZDM Math. Educ.* **2020**, *52*, 637–649. [CrossRef]
68. Clements, D.H.; Sarama, J.; Baroody, A.J.; Kutaka, T.S.; Chernyavskiy, P.; Joswick, C.; Cong, M.; Joseph, E. Comparing the efficacy of early arithmetic instruction based on a learning trajectory and teaching-to-a-target. *J. Educ. Psych.* **2021**, *113*, 1323–1337. [CrossRef]
69. Baroody, A.J.; Clements, D.H.; Sarama, J. Does Use of a Learning Progression Facilitate Learning an Early Counting Concept and Skill? University of Illinois Urbana: Champaign, IL, USA, 2022; *submitted*.
70. Sarama, J.; Clements, D.H.; Barrett, J.E.; Cullen, C.J.; Hudyma, A. Length measurement in the early years: Teaching and learning with learning trajectories. *Math. Think. Learn.* **2021**, 1–24. [CrossRef]
71. Baroody, A.J.; Yilmaz, N.; Clements, D.H.; Sarama, J. Evaluating a basic assumption of learning trajectories: The case of early patterning learning. *J. Math. Educ.* **2021**, *13*, 8–32. [CrossRef]
72. Yilmaz, N.; Baroody, A.J.; Clements, D.H.; Sarama, J.; Sahin, V. Does a Learning Trajectory Facilitate Learning to Recognize the Core Unit of a Repeating Pattern. In Proceedings of the Exploring Cognitive Processes in Mathematics; American Educational Research Association Annual Meeting, San Francisco, CA, USA, 18 April 2020.
73. Yilmaz, N.; Baroody, A.J.; Sahin, V. What do eye-tracking data say about the cognitive mechanisms underlying the pattern extension skills of young children? In Proceedings of the Poster presented at the Stanford Educational Data Science Conference, Stanford, CA, USA, 18 September 2020.
74. Fuson, K.C. *Children's Counting and Concepts of Number*; Springer: Berlin/Heidelberg, Germany, 1988.
75. Slavin, R.; Smith, D. The relationship between sample sizes and effect sizes in systematic reviews in education. *Educ. Eval. Policy Anal.* **2009**, *31*, 500–506. [CrossRef]

76. Cronbach, L.J.; Ambron, S.R.; Dornbusch, S.M.; Hess, R.O.; Hornik, R.C.; Phillips, D.C. *Toward Reform of Program Evaluation: Aims, Methods, and Institutional Arrangements*; Jossey-Bass: San Francisco, CA, USA, 1980.
77. Sarama, J.; Clements, D.H.; Barrett, J.E.; van Dine, D.W.; McDonel, J.S. Evaluation of a learning trajectory for length in the early years. *ZDM*. **2011**, *43*, 667–680. [[CrossRef](#)]
78. Duncan, R.G.; Gotwals, A.W. A tale of two progressions: On the benefits of careful comparisons. *Sci. Educ.* **2015**, *99*, 410–416. [[CrossRef](#)]
79. Lesh, R.; Yoon, C. Evolving communities of mind—Where development involves several interacting and simultaneously developing strands. *Math. Think. Learn.* **2004**, *6*, 205–226. [[CrossRef](#)]
80. Clements, D.H.; Sarama, J. *Building Blocks, Volumes 1 and 2*; McGraw-Hill: Columbus, OH, USA, 2013.
81. Baratta-Lorton, M. *Workjobs: Activity-Centered Learning for Early Childhood*; Addison-Wesley: Menlo Park, CA, USA, 1972.
82. Fyfe, E.R.; McNeil, N.M.; Rittle-Johnson, B. Easy as ABCABC: Abstract language facilitates performance on a concrete patterning. *Child Dev.* **2015**, *86*, 927–935. [[CrossRef](#)]
83. Ginsburg, H.P.; Baroody, A.J. *Test of Early Mathematics Ability*, 3rd ed.; Pro-Ed: Austin, TX, USA, 2006.
84. Clements, D.H.; Sarama, J.; Wolfe, C.B.; Day-Hess, C.A. *REMA—Research-Based Early Mathematics Assessment*; Kennedy Institute, University of Denver: Denver, CO, USA, 2008.
85. Baroody, A.J.; Purpura, D.J. Early number and operations: Whole numbers. In *Compendium for Research in Mathematics Education*; Cai, J., Ed.; National Council of Teachers of Mathematics: Reston, VA, USA, 2017; pp. 308–354.
86. Le Corre, M.; van de Walle, G.A.; Brannon, E.; Carey, S. Revisiting the performance/competence debate in the acquisition of counting as a representation of the positive integers. *Cogn. Psychol.* **2006**, *52*, 130–169. [[CrossRef](#)]
87. Sarnecka, B.W.; Carey, S. How counting represents number: What children must learn and when they learn it. *Cognition* **2008**, *108*, 662–674. [[CrossRef](#)]
88. Palmer, A.; Baroody, A.J. Blake’s development of the number words “one,” “two,” and “three”. *Cogn. Instr.* **2011**, *29*, 265–296. [[CrossRef](#)]
89. Wynn, K. Children’s acquisition of the counting words in the number system. *Cogn. Psychol.* **1992**, *24*, 220–251. [[CrossRef](#)]
90. Dixon, J.A.; Moore, C.F. The logic of interpreting evidence of developmental ordering: Strong inference and categorical measures. *Dev. Psychol.* **2000**, *36*, 826–834. [[CrossRef](#)] [[PubMed](#)]
91. Rittle-Johnson, B.; Fyfe, E.R.; Loehr, A.L.; Miller, M.R. Beyond numeracy in preschool: Adding patterns to the equation. *Early Child. Res. Q.* **2015**, *31*, 101. [[CrossRef](#)]