Augmenting intelligence: Developmental limits to learning-based cognitive change

Article · May 2016

DOI: 10.1016/j.intell.2016.02.005

4 authors, including:

Constantinos Christou
University of Cyprus
125 PUBLICATIONS 856 CITATIONS

SEE PROFILE

George Spanoudis
University of Cyprus
60 PUBLICATIONS 597 CITATIONS

SEE PROFILE

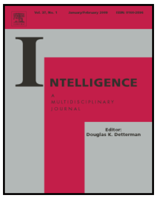
Andreas Demetriou
University of Nicosia
118 PUBLICATIONS 1,609 CITATIONS

SEE PROFILE

Intelligence 56 (2016) 16–27

Contents lists available at ScienceDirect

Intelligence



Augmenting intelligence: Developmental limits to learning-based cognitive change

Eleni Papageorgiou [a], Constantinos Christou [b], George Spanoudis [b], Andreas Demetriou [c,*]

[a] *Pedagogical Institute, Cyprus*
[b] *University of Cyprus, Cyprus*
[c] *University of Nicosia, Cyprus*

ARTICLE INFO

Article history:
Received 21 April 2015
Received in revised form 13 February 2016
Accepted 13 February 2016
Available online xxxx

Keywords:
Mathematics learning
Intelligence
Development
Reasoning
Attention control
Working memory

ABSTRACT

We investigated if learning relational reasoning in mathematics generalizes to other domains and general intelligence, including speed, attention control, and working memory. A total of 118 10-year olds were involved, allocated to an experimental and a control group. The experimental group was involved in 12 learning sessions addressed to various aspects of relational reasoning. Various analyses, including Rasch scaling, growth modeling and structured means analysis, showed significant but not sustainable learning gains in the ability trained. However, learning transferred to similar processes in analogical reasoning and also to attention control and working memory, indicating sustainable effects on mechanisms underlying general intelligence. An upper developmental constraint to learning was found. Implications for psychometric and developmental theories of intelligence and for education are discussed.

Learning is of central concern to many disciplines. In the psychology of intelligence, researchers focus on a double face problem. On the one hand, they try to specify how learning is constrained by general intellectual ability (i.e., g or its measured manifestation, IQ). On the other hand, they examine how learning may change general intellectual ability itself, if at all (Hunt, 2011; Jensen, 1998). In developmental psychology, this problem is restated in terms of developmental constraints. That is, it is examined, on the one hand, if learning possibilities vary as a function of developmental level (or stage) of cognitive processes (Brainerd, 1977; Piaget, 1964). On the other hand, it is also examined if learning may accelerate transition across developmental levels and elevate individuals higher on a developmental hierarchy than it would be possible by spontaneous development (Brainerd, 1977; Efklides, Demetriou, & Gustafsson, 1991; Klauer, 1998, 2014; Klauer & Phye, 1994). In educational science concerns are more practical, focusing on the stability of learning gains and their transfer to other domains (Csapó, 1999; Greiff et al., 2014; Klauer & Phye, 2008). This study is related to all of these concerns: We examine if learning to use general cognitive processes (e.g., classification and induction of relations) in a specific domain (i.e., mathematics) (i) augments general intelligence (defined as a latent construct underlying several domains in addition to mathematics), (ii) transfers to domain-free representational and processing capacities, such as processing speed, attention control, and working memory, (iii) varies over time, and (iv) is constrained by developmental level.

There is general agreement that g (or IQ, as a global measurement of g) is systematically related to learning. On the one hand, high g implies faster, deeper, and more stable learning than low g (Jensen, 1998). On the other hand, learning (school-based or experimentally induced) influences intelligence positively. There is evidence that each extra year of schooling augments IQ by 2–4 IQ units (Ceci, 1991; Gustafsson, 2008; Gustafsson & Undheim, 1996). However, it is disputed if this effect reflects a better handling of the test itself or a real increase in intelligence. Jensen (1998) suggested that these effects are shallow, primarily reflecting improvement in test taking skills rather that a change in g itself. There is empirical support for this view. For instance, te Nijenhuis, van Viane, and van der Flier (2007) claimed that test–retest gains and gains related to systematic learning experiences addressed to the abilities related to various intelligence tests are not related to g. It is also claimed that gains in IQ from long-term programs, such as the Head Start program, did not relate to g because they do not affect the underlying processing and inferential mechanisms of g (te Nijenhuis, Jongeneel-Grimen, & Kirkegaard, 2014).

The assumption that g is impervious to learning was invoked to explain the finding that increases in IQ because of learning fade out with time. However, this interpretation ignores the possible developmental variation of g. That is, developmental theory assumes that development transforms the underlying g construct in both its representational and

[*] Corresponding author at: University of Nicosia Research Foundation, University of Nicosia, Makedonitissas Avenue, P.O. Box 24005, 1700 Nicosia, Cyprus.
E-mail addresses: ademetriou@ucy.ac.cy, demetriou.a@unic.ac.cy (A. Demetriou).

http://dx.doi.org/10.1016/j.intell.2016.02.005

inferential efficiency. Therefore, when intervention is delivered only at a given time T but measurements are taken at both time T and a relatively remote time T + 1, what appears to be a fade out effect because performance at T + 1 is lower than performance at T simply reflects the fact that g at T + 1 is not identical to g that was affected by the learning experience. This possibility renders conclusions regarding the depth of learning effects (test taking expertise or underlying g-loaded mental processes) unfounded. A critical test of this assumption would be to examine if g-related learning generalizes to underlying processing and representational mechanisms, such as attention control and working memory.

Developmental research is only partially in agreement with psychometric research. On the one hand, somehow echoing Jensen's position about g-bound constraints of learning, Piaget (1964) (see also Inhelder, Sinclair, & Bovet, 1974) himself postulated that learning is constrained by the current mental structure. That is, inferential patterns and concepts exceeding the assimilatory possibilities of the current structure cannot be learned, because this structure would reject or distort patterns and concepts that cannot be meaningfully understood. On the other hand, Piaget did accept that learning directed to the integration and consolidation of the mental operations underlying the current mental structure may both accelerate the development of this structure and generalize to concepts drawing upon it. In psychometric terms, this would be equivalent to change in mental age as a result of learning.

Research in this tradition investigated the effects of learning on all sorts of Piagetian structures and concepts (Brainerd, 1977; Efklides et al., 1991; Inhelder et al., 1974; Shayer & Adey, 2002; Strauss, 1972). In line with Piaget, this research found that learning focusing on the integration of mental operations was more successful, stable, and transferrable than learning focusing on the acquisition of specific skills and processes. Also, it was found that progress within a stage is much easier to attain than progress across stages. Along these lines, Klauer and his colleagues (Klauer and Phye, 1994; Klauer, Willmes, and Phye, 2002) developed a program that trained children to reason inductively, drawing from both the developmental and the psychometric approach. This program adopted the Piagetian assumption that processing of similarities and differences between objects or representations, inducing their underlying relations, and integrating them into classificatory or relational schemes is crucial for operational development (Inhelder et al., 1974). Notably, this assumption coincides with the psychometric assumption that induction of relations between objects or representations and of relations between relations is the substance of g (Carroll, 1993; Spearman, 1904) or fluid intelligence (Cattell, 1963). Klauer and colleagues maintained that their program permanently increased fluid intelligence and improved academic performance (Klauer, 1998, 2014; Klauer & Phye, 2008; Klauer et al., 2002).

A stricter test of the effects of learning would be to specify if an intervention transfers to fundamental representational and processing capacities underlying the ability trained, such as attention control or working memory. This is because individual differences in fluid intelligence are assumed to reflect differences in these fundamental processes. Specifically, fast processing, (Jensen, 1998), attention control (Diamond, 2013), and working memory (Kyllonen and Christal, 1990) are associated with higher intelligence. In developmental research, changes in each of these processes were associated with changes in thought and problem solving (Case, 1985; Demetriou, Christou, Spanoudis, and Platsidou, 2002; Kail, 1991, 2007; Pascual-Leone, 1970; Pascual-Leone and Johnson, 2011). It was suggested that these processes relate in a cascade fashion such that increasing speed facilitates attention control, which facilitates working memory, which facilitates transition to higher levels of reasoning and problem solving (Fry and Hale, 1996, 2000; Kail, 2007).

There is research examining if modifying these processes transfers to g. Findings so far are inconclusive. Several studies showed that training executive processes in working memory, such as information binding and attention control, did transfer to fluid intelligence (Jaeggi, Buschkuehl, Jonides, and Perrig, 2008) and every day and school performance (Barnett, 2011; Diamond, 2013). However, extensive evaluation of this literature suggested that training executive processes confounds changes in the command of these processes per se with changes in inferential processes shared by working memory and Gf (Melby-Lervag & Hulme, 2013; Shipstead, Redic, & Engle, 2012). That is, what is supposed to be transfer of learning effects from WM to Gf it is actually learning directly affecting Gf. Along the same line Nutley et al. (2011) showed that training nonverbal Gf related reasoning processes did raise Gf in 4 years old children; however, training working memory processes, although effective to improve working memory performance, did not transfer to Gf. On the contrary, Rueda, Checa, and Combita (2012) found that training attention control did transfer to Gf in 5 years old children.

Incongruence between studies may be apparent rather than real. That is, it might be the case that the possible impact of learning varies with age, because the role of different processes varies with development. In this case, differences between studies may simply reflect differences in the processes addressed vis-à-vis participants' age. Demetriou and colleagues (Demetriou et al., 2013; Demetriou, Spanoudis, & Shayer, 2014; Demetriou et al., 2014) advanced a model of intellectual development postulating that these relations vary systematically with developmental phase. According to this model, fluid intelligence develops through four major reconceptualization cycles (the ReConceP sequence), with two phases in each. In succession, the four cycles operate with episodic representations (birth to 2 years), realistic mental representations (2–6 years), rule-based reasoning integrating mental representations (6–11 years), and principle-based reasoning integrating rules (11–18 years). Transitions within cycles occur at 4 years, 8 years, and 14 years, when relations between the representational units of the present cycle are metarepresented into the representational units of the next cycle (Christoforides, Spanoudis, & Demetriou, in press). These cycles were specified on the basis of performance on a large variety of tasks addressed to reasoning and problem solving in various domains. Many of these tasks were used here to test the reasoning in various domains (see Method). These include pragmatic and conditional reasoning, categorical and analogical reasoning expressed through verbal, numerical, and figural content, scientific reasoning addressed to various aspects of hypothesis formation and testing, and various aspects of spatial reasoning, such as mental rotation and orientation in space (Demetriou & Kyriakides, 2006).

Demetriou et al. (2013, 2014) showed that changes in Gf in the first phase of each cycle (i.e., at 6–8 years and 11–13 years) are related to changes in processing efficiency. Measures of processing speed, such as choice reaction times and Stroop-like tasks of attention control were used to measure processing efficiency. Changes in the second phase of each cycle (i.e., 4–6 years, 8–10 years, and 13–16 years) are related to changes in working memory. Tasks addressed to various aspects of short-term memory and executive processes in working memory were used (Demetriou et al., 2002; Demetriou, Mouyi, & Spanoudis, 2008; Demetriou et al., 2013; Demetriou et al., 2014). They suggested that this pattern reflects differences in the processing requirements of developmental acquisitions. At the beginning of cycles processing speed is a better index because it reflects changes in the facility of using the new mental units. Later in the cycle, when networks of relations between representations are established, working memory is a better index because alignment and inter-linking of representations both requires and facilitates working memory.

In short, this model posits that intelligence is a universe of processes which give meaning to the world, handling change sensibly and adaptively. The main meaning-making processes are abstracting, aligning and relating, and filling in gaps of information and evaluating them by inference and reasoning. It is a developmental process that accomplishes these aims under the representational and processing constraints of the current phase, finding ways to minimize the constraints and enhance possibilities. In so doing it causes development in representational and processing possibilities. Individual differences at any phase may come from differences in representational and processing possibilities related to underlying brain structures but also from lack of knowledge related to encounters and lack of evaluation sensitivities that would grasp what is needed and flexibly adjust available representations and processes.

This model allows more specific predictions about learning effects because these may be tuned to the network of relations associated with the developmental phase concerned. For instance, in the early phase of a cycle, there should be a stronger flow of effects between Gf and speed or attention control rather than with working memory. In the second phase the opposite pattern should be expected. In time windows spanning over two phases, these patterns may be inverted from one measurement to the next. For instance, in the transition from the rule-based to the principle-based cycle, working memory dominates at the beginning because Gf-WM relations are stronger in the second phase of the rule-based cycle and speed-control dominates at the end because in the first phase of the principle-based concepts Gf-speed relations are stronger. Also, progression along ReConceP may vary across different domains, because abstraction and inferential processes underlying ReConceP are more likely within rather than across domains (Demetriou et al., 2014). Thus, transfer of learning across domains may vary as a function of the proximity of domains to the processes affected by learning: The closer the better, because naturalization into the informational and procedural specificities of the other domain is easier. For instance, training relational and analogical reasoning in mathematics would be easier to transfer to relational and analogical reasoning in other domains rather than to causal reasoning. This later domain would require experimentation and isolation of variables skills, in addition to relational reasoning that would connect outcomes with possible causes.

However, processing and representational efficiency as captured by speed of processing, attention control, and working memory tasks are good indexes of transfer possibilities. When speed and attention control signify that an individual operates in the early phase of a cycle, transfer from a domain trained to another domain would be difficult because mental units are not yet fluent. When working memory signifies that an individual operates in the second phase of a cycle, transfer would be easier because the individual is already aligning mental units across domains.

This study involved 10 to 11-year old primary school children. This is a transition age between the cycle of rule-based reasoning to the cycle of principle-based reasoning. Therefore, this age allows studying the alternation of relations between processes that come as a result of change in developmental cycle. Moreover, it allows examining if training addressed to processes related to the last phase of one cycle (e.g., relational thought) may transfer to processes related to next cycle (e.g., hypothesis testing by experimentation). Specifically, these children were systematically trained to use relational thought in the domain of mathematics. In addition to mathematics, these children were examined by tasks addressed to analogical, deductive, spatial, and scientific reasoning and also to speed, attention control, and working memory. Based on their performance at pretest, a control and an experimental group were formed, matched on all of these processes on the group level. The experimental group was involved in the learning experiment to be described below and then both groups were examined immediately after the experiment and several months later.

The models discussed above lead to different predictions about the possible effects of learning. First, the Jensen-based psychometric model assumes no transfer to g. Therefore, in line with this assumption, change would be limited in the domain trained (mathematics) and it should fade out with time. In concern to the possible relations between Gf and measures of processing and representational efficiency (PREM), such as speed, attention control, and working memory, learning gains may depend on working memory or any other of the PREM, to reflect the dependence of inferential processes on more fundamental processes. However, transfer of learning effects Gf-related processes to PREM processes would not be expected because inferential processes do not condition the operation of PREM processes.

Second, the classic developmental model predicts transfer of effects to the processes related to those trained. That is, process-directed learning may generalize to domains primarily reflecting process use, such as analogical and deductive reasoning. This model is silent about the Gf-PREM relations. Its neo-Piagetian version, however, would make a prediction equivalent to the psychometric prediction. That is, any effect of learning or any transfer of it would depend on PREM measures but it would not affect them.

Third, the ReConceP model aligns with developmental theory in concern to transfer to other domains. In fact, it specifies that this generalization would be discernible at the level of a latent construct defined over the various domains. In concern to the domains, it predicts that it would be proportional to their proximity to the domain trained, being maximum in analogical and deductive reasoning, and minimal in spatial and scientific reasoning. In addition, it predicts that learning would also transfer to working memory and speed and attention control. In this later case, the relations between learning-based change and these efficiency indexes would be phase-sensitive. That is, it would depend on working memory but it would generalize to all three of them. This is so because the present age phase was transitional between the conceptual and the principles-based cycle, when working memory is the index of change in the first phase and speed in the second phase. If the learning experiment simulates spontaneous development, it would generate the patterns expected according to normal age progression.

1. Method

1.1. Participants

A total of 118 (54 males) children were involved. At first testing, these children came from fifth grade primary school classes randomly selected from five schools in Nicosia, Cyprus's capital. All children were native speakers of Greek. Their mean age at first testing was 10.5 years. Within schools, one class was assigned to the experimental group (three classes in total; 31 females, 29 males) and at least one class was assigned to the control group (four classes in total; 33 females, 25 males). Classes in Cypriot schools involve an average of 21 students. Also, all classes in all schools are mixed ability classes. That is, children are randomly assigned to classes so that all ability levels are represented in the classes of each grade. The performance of the two groups on mathematics ($t = -.20$, $p > .84$), the cognitive domains addressed by the battery described below ($t = -1.56$, $p > .12$), working memory, $t = 1.16$, $p > .24$), speed, ($t = -.74$, $p > .94$), and attention control ($t = -1.89$, $p > .10$) did not differ significantly at first testing on any of the abilities of interest. It is noted that two participants from the control group were dropped as outliers from all analyses presented below, because their contribution to normalized multivariate kurtosis was very large (Bentler, 2006). This was due to the fact that these children fluctuated between ceiling at pretest and near floor at the first posttest on the mathematics battery. The control group was a “no-contact” group, which received no intervention other than regular instruction in school. The learning experiment (see below) and the two testing waves after learning took place when students were at sixth grade. The immediate posttest took place within one month after the end of the intervention (mean age at immediate posttest was 11.2 years). The delayed posttest for each classroom started four months after the completion of the immediate posttest (mean age at the delayed posttest was 11.6 years).

1.2. Task batteries

Three task batteries were used: The first addressed mathematical reasoning; the second addressed several other domains of reasoning (i.e., analogical, deductive, spatial, and scientific reasoning) to abstract a latent Gf factor and examine possible generalization of the effects of learning; the third addressed processing and representational efficiency (PRE) (i.e., speed, attention control, and working memory).

1.2.1. Mathematical reasoning

This battery included 22 tasks, focusing on processing similarities and differences between numbers and ensuing relations. Ten of these tasks were concerned with number grouping according to common attributes and 12 tasks were concerned with number seriation according to their relations. Specifically, some problems required participants to identify common attributes between numbers (e.g., what is common between 4, 16, 8, 32, 20, 100, and 40), form sets based on common attributes (e.g., "select what is common between these numbers {12, 14, 7, 56, 28, 36, 84, 54, 49, 19} and spell it out explicitly"), extrapolate number sets by adding new objects sharing their defining property (e.g., which one of the numbers {9, 12, 6, 7, 3, 8} belongs to the set {24, 36, 18, 15, 63, 30}, identify numbers that do not belong to a set (e.g., 9, 21, 11, 15, 12, 6, 35) because they do not share the set's defining property, detect similarities and differences in two-dimensional series (e.g. complete with the right number: (8, 4, 2 to 1, 1/2, 1/4 to 1/8, 1/16, …). In some of the tasks, children were asked to explicitly specify the relations running in two or more items.

These tasks are developmentally patterned. Specifically, tasks requiring participants to specify an explicitly present relation (e.g., each next number in a series is the double of the previous one) address early rule-based reasoning. Tasks requiring to map relations of this kind onto each other require late rule-based reasoning. Finally, tasks requiring to specify relations between relations and explicitly spell them out require principle-based reasoning.

Each item was scored on a fail (0) – pass (1) basis. The battery was very reliable (Cronbach's alpha was .92, .93, and .93 for the three testing waves, respectively).

1.2.2. Reasoning in other domains

The second battery was based on the test of cognitive development presented by Demetriou and Kyriakides (2006) because it has good psychometric and developmental properties. This battery involved the following tasks.

1.2.2.1. Analogical reasoning. Five verbal analogies and three Raven-like matrices of increasing complexity addressed analogical reasoning. Complexity was specified in reference to the familiarity, the abstractness, and the order of the relations involved. Specifically, there were three analogies of the a : b :: c: d type, where the thinker was asked to specify the d component. Some of them were concrete and familiar (i.e., ink : pen :: paint :: - [color, *brush*, paper]) and some of them were abstract (i.e., picture : painting :: word : [paper, speech, *literature*]). The rest required the construction of third and fourth order relations {(tail : fish :: feed : mammals) ::: - [*movement*, animals, vertebrates]} :::: {(propeller : ship :: wheels : car) ::: - [vehicles, *transportation*, carriers]}. The participant's task was to choose the correct word (shown here *in italics*) among the three alternatives provided for each missing element.

Eleven Raven-like matrices of increasing complexity addressed figural analogical reasoning. Matrices varied in complexity according to the number of the dimensions and transformations involved. Specifically, five matrices required to grasp the pattern organizing several figures varying along a single dimension (e.g., systematic change in size, shape, size and shape). Six matrices required to grasp the relation between two or more dimensions (e.g., size and shape); some matrices required to grasp the relation between transformations altering the dimensions involved (one figure integrated into another while background shades changed according to a certain rule).

1.2.2.2. Deductive reasoning. Six class inclusion tasks and five syllogisms addressed deductive reasoning. Difficulty of class inclusion tasks was manipulated in reference to the relationships between the classes involved. Syllogisms addressed logical relations of increasing complexity (i.e., modus ponens, transitivity, modus tollens, negation, disjunction).

1.2.2.3. Visuospatial reasoning. Three sets of tasks addressed visuospatial reasoning: Paper folding examined manipulation of rather familiar mental images (3 items). Mental rotation was examined by matching objects rotated to various degrees (4 items). A rotating clock where the two hands would come over each other at various degrees so that pictures on them would merge examined integration of mental images and rotation (3 items). Difficulty was manipulated in reference to the number of dimensions involved and the complexity of rotation.

1.2.2.4. Scientific reasoning. To address scientific reasoning, isolation of variables and hypothesis testing were examined. For isolation of variables, combinations between at least two levels of two variables (plant and light) were given (i.e., someone planted beans at a sunny place and wide beans at a sunny place) and participants were asked to specify which variable is tested (4 items). For hypothesis testing, participants were asked to choose one of several alternative experiments best testing a specific hypothesis (3 items).

Each item was scored on a fail (0) – pass (1) basis. The battery was also very reliable (Cronbach's alpha was .76, .79, and .83 for the three testing waves, respectively).

Analogies addressed to familiar relations, simple patterns varying along a single dimension, mental rotations along a specific dimension, and simple modus ponens and disjunction arguments are solved by early rule-based reasoning. Explicitly constructing relations and mapping then on other relations, identifying interchanging patterns, and mental rotation along two coordinated dimensions require late rule-based reasoning. Finally, verbal analogies and Raven-like matrices based on third-order relations between multiply varying dimensions also require principle-based reasoning. Also the processes addressed by the scientific reasoning tasks are attained in the cycle of principle-based abilities, because they require coordination of general principles, such as a hypothesis and experimentation by isolation of variables.

1.2.3. Processing efficiency tasks

A series of Stroop-like tasks measured speed and attention control under three symbol systems, namely, verbal, numeric, and figural. Specifically, there were 36 stimuli for each symbol system, 18 congruent stimuli addressed to speed and 18 stimuli incongruent addressed to attention control.

For verbal speed of processing, participants read a number of words denoting a color written in the same ink-color (e.g., the word "red" written in red). For verbal control, participants recognized the ink-color of color words denoting another color (e.g., the word "red" written in blue ink). Both tasks employed the following three color names, which, in Greek, have the same number of letters: κόκκινο (red), πράσινο (green), κίτρινο (yellow).

To measure speed and attention control in the number domain, several "large" number digits, which were composed of "small" digits, were prepared. This task involved the numbers 4, 7, and 9. In the compatible condition, the large digit (e.g., 7) was composed of the same "small" digit (i.e., 7). In the incompatible condition, the large digit (e.g., 7) was composed of one of the other digits (e.g., 4). To measure speed, participants recognized the large number digits of congruent digits. To measure attention control, participants recognized the component digit of incongruent digits.

To measure speed and attention control in the visual domain, several geometrical figures were composed like the number digits above. That is, large geometrical figures (circles, triangles, and squares) were made up of the same (congruent) or a different (incongruent) figure. To measure speed in this domain, participants recognized the large geometrical figure of congruent conditions; to measure attention control, the recognized the small figure of incongruent conditions.

Reaction times in ms were used. Reliability was again very high (Cronbach's alpha was .88, .92, and .88 for the three testing waves, respectively).

1.2.4. Short-term and working memory

Both short-term and working memory were examined. Two computer-administered tasks addressed short-term storage. The verbal and the numerical tasks addressed forward word span and 2-digit number span, respectively. There were six levels (2–7 units) with two sets in each level in each system. To test working memory, participants were required to recall from 2 to 7 words following execution of numerical operations to check if simple expressions were right or wrong (e.g., $(4 \div 2) + 1 = 3)$). Each set of words appeared when the participant pressed a key to indicate that she completed her evaluation of the mathematical expression presented. To control for the effect of mathematical facility on working memory, following the task above, children were then asked to mentally execute numerical operations on 1-digit numbers and type their answers. The task addressed to visuospatial storage required to recall the geometrical figures included in sets of figures (e.g., a circle, a triangle, and a square). Set size varied from 2 to 7 figures. Presentation time for each set was proportional to number of figures involved (2 s/figure). The set to be memorized was presented and was removed at the specified time. Then four sets were presented simultaneously and arranged side by side. Participants were asked to choose the one of the four sets fully matching the target set.

Scores indicated the upper level attained on each test. The reliability of these tasks was below optimum, especially at first testing (Cronbach's alpha was .32, 69, and .59 for the three testing waves, respectively). This was caused by random variation of performance on the visuospatial task. Dropping this task resulted in a very large increase of reliability across all three testing waves (Cronbach's alpha was .75, 86, and .83 for the three testing waves, respectively). Thus, the visuospatial task was not used in the analyses to be presented below.

1.3. The intervention program

Our training program aimed to enable students to identify the various dimensions underlying the various mathematical reasoning tasks described above, explicitly conceive of their various groupings, and build the problem solving skills associated with each. Specifically, students were taught to look for and abstract properties and relations, based on similarities and differences between tasks and task types, align them according to a specific goal, conceptualize problems, and build problem specific problem solving strategies. Students were instructed to identify different problem types based on their mathematical and inferential requirements, explicitly represent each structure, and specify similarities and differences between problem types. Thus, they were required to explicitly metarepresent both problem structures and processes as well as their associations. The emphasis was on formative concepts like "attributes", "relations", "similarity", "dissimilarity or difference" and their instantiation in the various problem types.

The intervention comprised twelve 40-min lessons, organized in three phases. The first phase involved the first three lessons. These lessons aimed to enable children to recognize the conceptual and procedural similarities of the various training problems presented. In this phase children were guided to search for and identify relevant attributes or relationships involved in a problem and explicitly represent them into conceptual maps of similarities and differences between tasks (see the problems included in the battery addressed to mathematics—mathematical reasoning tasks; for instance, children were instructed to specify the relation underlying various patterns of numbers and separate patterns into those ruled by the same relation and those differing). The second phase involved six lessons. Each of the six lessons focused on a different type of problems (e.g., increase, decrease, relations between whole numbers, relations between fractions) and instructed children how to solve them (e.g., first specify the relation between the two numbers of a fraction and then specify the relations between fractions) and practice on other problems. Children were guided to construct procedural diagrams explicitly representing the sequence of steps involved in the solution of a problem. Finally, the last phase involved three lessons. These lessons focused on the encoding of relations into rules (e.g., fractions are relations where the number below the line denotes how an entity is divided and the number above the line denotes how many parts of those specified by the other number are taken), the specification of relations between rules (e.g., all fractions can be reduced into a number specifying how an entity is divided), the transfer of problem solving strategies to new problems, the combination of strategies in complex problems requiring more than one strategy, the evaluation of solutions, and their explicit metarepresentation. In sake of metarepresentation, in this phase, students were also required to recall strategies from memory according to problem prompts standing for different problem types and explicitly describe solution processes in detail and explicate why each is appropriate for each problem type. The general scheme guiding actions in phases 2 and 3 involved three steps: (i) search, specify, and classify problem; (ii) compare problem with other problems; (iii) solve problem choosing the best strategy available. Feedback was provided to children about the appropriateness of their answers.

The content of problems was taken from the mathematics curriculum of 5th and 6th grades. For example, activities involved concepts related to the factorization and the divisibility of natural numbers, algebraic expressions and generalizations about the properties of numbers and numbers' operations (e.g., odd + odd = even, the sum of two consecutive triangular numbers), numerical proportions, number sequences (e.g., Fibonacci number sequence, the sequence of triangular numbers, etc.), and attributes and properties of two-dimensional and three-dimensional figures (e.g., different kinds of parallelograms, properties of parallelograms, analogy tasks with figures, etc.) and geometrical patterns.

In terms of the ReConceP model outlined in the introduction, this instruction program focused on establishing and consolidating processes primarily pertinent to the rule-based cycle. In concern to the other cognitive domains tested in this study, instruction was related to analogical reasoning in addition to mathematics which was its primary aim. The other domains addressed by the batteries above were only minimally and indirectly related.

1.3.1. Sessions

All testing took place at schools during regular school hours. There was a separate session for each of the three batteries above, each lasting for about 1 h. Examination followed regular school brakes.

Each of the 12 intervention lessons lasted for 40 min, which is the regular duration of a school period. The 12 lessons spread over nine weeks. Specifically, one or two lessons were weekly delivered until the program was completed. Slight variations in the rate of delivering the lessons was necessary to tune the intervention program with everyday school activities. Obviously, this intervention might be shorter or longer than the 12 sessions delivered here. According to our design, this number of sessions was enough to meet the targets of instruction as presented above. Moreover, it was decided that the density of sessions was appropriate to sustain the necessary continuity across sessions, given the constraints of the school program.

Pretest sessions took place in the last semester of fifth grade. The intervention took place at the first semester of sixth grade. Posttest sessions took place at the second semester of sixth grade.

The intervention was delivered to each class by only one person, the first author of this paper, who is a secondary school mathematics teacher.

2. Results

Four approaches were adopted to specify the nature of change across the three testing waves, the possible effects of instruction, and the interaction between processes. First, a series of Rasch analyses were applied on the performance attained on the mathematical reasoning battery and the cognitive battery. These analyses aimed to construct systematic ability dimensions that would reflect the developmental/difficulty structure of the batteries. These would then be used to specify the effect of intervention on the underlying constructs represented. Three sets of Rasch analyses were run. The first set was applied on the performance attained on the mathematical reasoning battery at each testing wave. The scales abstracted showed that items requiring to identify an explicitly present rule (e.g., numbers double) reside at the lower end of the scale; items requiring to identify multiply varying patterns, match relations vis-à-vis a general principle and explicitly state a principle reside at the higher end of the scale. The second analysis was applied on all of the items included in the cognitive battery. Items requiring simple mental rotations of familiar objects (e.g., paper folding along the diagonal, simple modus ponens or disjunction syllogisms, and verbal analogies involving familiar objects (e.g., Nicosia is for Cyprus what London is for Britain) scaled at the lower end of this scale; mental rotations along multiple dimensions, higher order analogies, negation syllogisms and all scientific reasoning items requiring scaled at the higher end of this scale. Obviously, this scale is a powerful index of Gf as it stands for a wide variety of cognitive processes. The third analysis was applied only on the six verbal analogies and the 11 Raven-like matrices. This scale, narrower as it is than the scale standing for performance on all cognitive tasks, allows differentiating relational inferential processes per se from other domain-specific skills, such as experimental or mental rotation. Taken together, these two scales would allow capturing possible differences in transfer between relational thought as such and its implementation to domains possibly requiring additional processes. All scales were very reliable as indicated by their high item (all >.6) and person reliability indices (all >.9).

Second, the scores of each participant on each of the Rasch scales above were subjected to various ANOVAs used to specify the general effects of training on the various factors of interest.

Third, growth modeling was used to precisely specify and model the patterns of change across abilities. Finally, structured means analysis was used to pinpoint the exact magnitude of change across abilities and specify their structural relations.

2.1. Capturing instruction effects

A series of analyses were applied on the logit scores attained by each participant on each of the three Rasch scales described above. Specifically, the first ANOVA compared the two experimental groups on the mathematical logit scores across the three testing waves. The main effect of experimental group was non-significant, F(1114) = 1.06, p > .05, η = .01. However, the main effect of testing wave, F(2113) = 24.28, p < .001, η = .30 and the experimental group x testing wave interaction, F(2114) = 9.31, p < .001, η = .14, were significant. The trends uncovered by this analysis are shown in Fig. 1A (the corresponding raw mean scores are shown in Supplementary Table 1). It can be seen that performance improved across testing waves in both the control (1.51, 1.80, 1.53, for the three testing waves, respectively) and the experimental group (1.29, 2.59, 1.79, for the three testing waves, respectively). Pairwise comparisons indicated that the difference between the first and the second testing and between the second and the third testing were significant, (Wilk's lambda = .70, p < .001). Univariate comparisons showed that the two groups did not differ at pretest, F(1114) = 1.417, p > .05, η = .01; however, the experimental group significantly outperformed the control group at immediate pretest, F(1114) = 6.81, p < .01, η = .06, but not at delayed posttest, F(1114) = .63, p > .05, η = .01. Within groups comparisons for differences between testing sessions showed that, in the experimental group, performance at first, (57) = 8.68, p < .001, and second pretest, t(57) = 3.90, p < .001, was significantly higher than performance at pretest. Performance at second posttest was significantly lower that performance at first posttest, t(57) = 3.68, p < .001. In the control group only the first of these three differences reached significance, t = 2.12 (57), p < .04. These results suggest that instruction was effective to improve mathematical reasoning but this effect weakened with time.

To test the possibility of transfer to non-trained processes, a second ANOVA compared the two groups on the logit scores estimated on the basis of performance on all of the tasks included in the cognitive battery. The main effect of experimental group was not significant, F(1114) = .06, p > .05, η = .00. However, both the main effect of testing wave, F (2113) = 9.58, p < .001, η = .14, and the testing wave x experimental group interaction were significant, F(1113) = 3.84, p < .03, η = .06. The effects reflected the fact that performance improved across testing in both the control (.81, .90, and .87) and the experimental group (.64, 1.02, and .82). Pairwise comparisons indicated that the difference between the first and the second and the first and the third testing was significant, (Wilk's lambda = .86, p < .001). Univariate comparisons of the two groups indicated that their difference at the first, F(1114) = 1.60, p > .05, η = .01, the second, F(1114) = .42, p > .05, η = .01, and the third testing, F(1114) = .08, p > .05, η = .00, was not significant. The within groups comparisons for differences between testing sessions showed that, in the experimental group, performance at first pretest was better than performance at pretest, t(57) = 4.93, p < .001; performance at the second posttest was only marginally better than at pretest, t(57) = 1.72, p < .09; performance at the second posttest was lower than at the first pretest, t = 2.11 (57), p < .04. None of these differences approached significance in the control group. Therefore, it seems that the experience of taking the cognitive test did influence all participants positively; instruction provided only a slight but not stable advantage to the trained participants (see Fig. 1B and Supplementary Table 1).

One might object that it may be too much to expect training transfer to an index of so disparate abilities including processes totally unrelated to the training provided here (i.e., spatial and scientific reasoning). To examine this possibility, a third ANOVA compared the two groups on the logit scores standing for performance on analogical reasoning (verbal analogies and Raven-like matrices). The results here were very different from the results presented so far. Specifically, the main effects of experimental group F(1114) = 7.77, p < .006, η = .06, and testing wave were highly significant, F(2113) = 19.19, p < .001, η = .25, as well as their interaction, F(1114) = 12.33, p < .001, η = .18. These results indicated that performance of both the control (1.26, 1.38, and 1.25 at the three testing waves, respectively) and the experimental group (1.22, 2.36, and 1.79 at the three testing waves, respectively) improved across testing. Pairwise comparisons suggested that the difference between the first and the second and the first and the third testing was significant (Wilk's lambda = .75, p < .001). Univariate comparisons showed that the two groups did not differ at pretest, F (1114) = .046, >.05, η = .00. However, the experimental group significantly outperformed the control group at both the second, F(1114) = 27.86, p < .001, η = .15, and the third testing, F(1114) = 5.26, p < .02, η = .04. Pairwise within groups comparisons for possible session differences showed that, in the experimental group, performance at first, t = 7.27 (57), p < .001, and second posttest, t = 3.36 (57), p < .001, significantly exceeded performance at pretest; performance at the second posttest was lower than at the first posttest, t = 3.58 (57), p < .001. None of these differences was significant in the control group. It is clear, therefore, that there was strong transfer of training to relational inferential reasoning which was sustained over time, despite a tendency to weaken (see Fig. 1C and Supplementary Table 1).

To explore the possible effects of training on working memory and processing efficiency three ANOVAs were run. The first was a 2 (experimental groups) × 3 (mean working memory performance on the working memory, the word, and the digit span tasks at the three testing

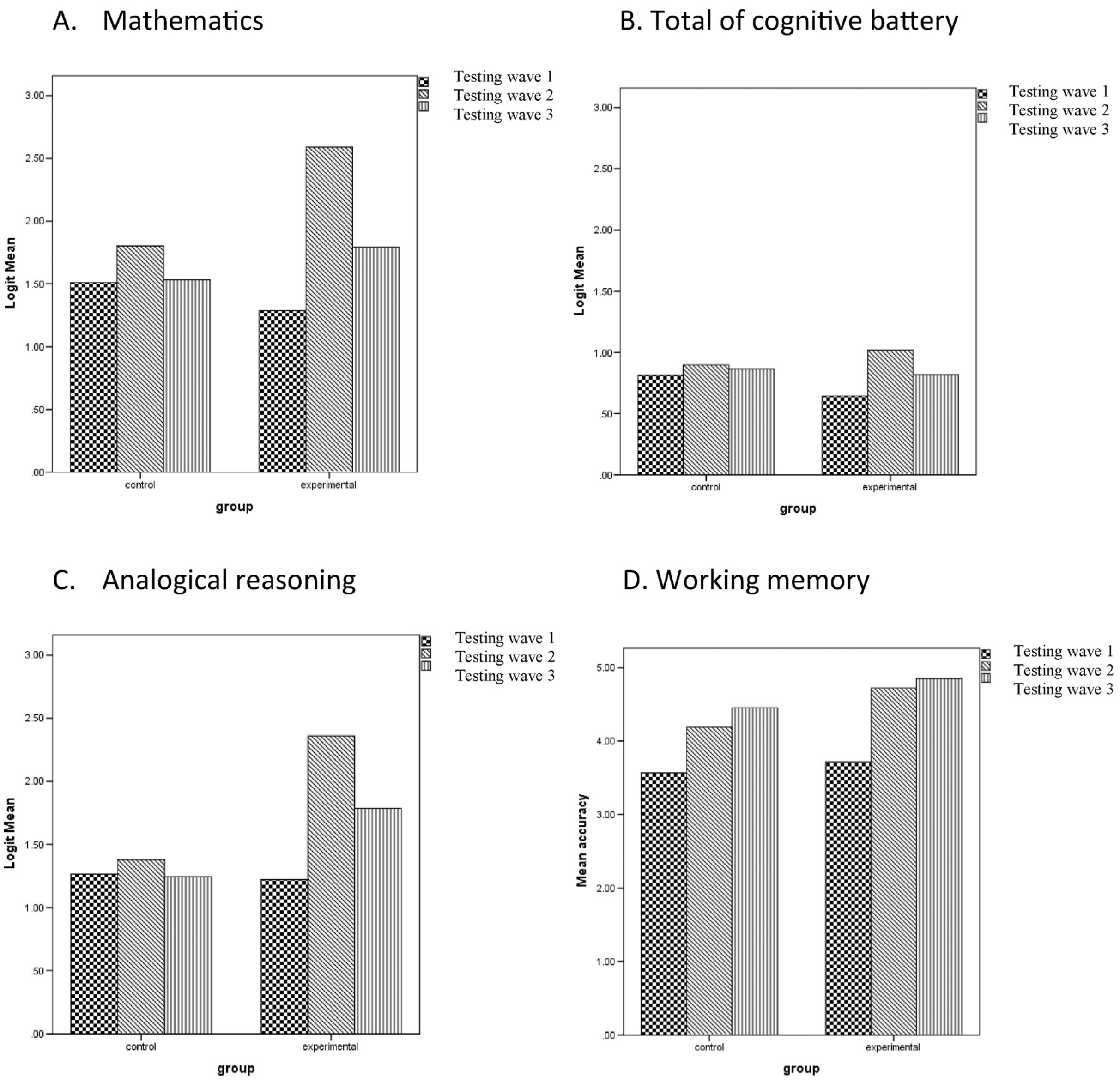


Fig. 1. Mean logit scores as a function of intervention, process, and testing wave.

waves) ANOVA with repeated measures on the last factor. The main effect of group was significant, F(1114) = 13.73, p < .001, η = .11. The effect of wave, F(2113) = 124.85, p < .001, η = .69 was very powerful. The wave × experimental group, F(1114) = 3.93, p < .03, η = .06 was also significant. Individual comparisons between the experimental groups revealed no significant difference at first testing, F(1114) = 1.68, p > .05, η = .02, but significant differences in favor of the experimental group at both the second, F(1114) = 18.98, p < .001, η = .14, and the third testing, F(1114) = 10.05, p < .002, η = .08. The pairwise within groups comparisons for possible session differences showed that, in the experimental group, performance at first, t = 8.46 (57), p < .001, and second posttest, t = 10.63 (57), p < .001, significantly exceeded performance at pretest; also, performance at the second posttest was higher than at the first posttest, t = 1.99 (57), p < .05. However, all of these differences, although smaller were also significant and in same direction in the control group, (t = 5.64, 8.16, and 3.57, p < .001 in all cases, respectively). Obviously, training did positively influence working memory, although the testing experience was also influential (see Fig. 1D and Supplementary Table 2).

To explore possible effects on speed and attention control two analyses were run. The first included mean performance on the compatible tasks at each testing wave. Only the wave effect was significant, F (1113) = 30.40, p > .001, η = .35, indicating systematic decrease of processing speed across the three waves in both groups (mean RT was 1263 ms, 1109 ms, and 1072 ms at the three waves, respectively). The second analysis was applied on mean performance attained on the incompatible tasks. In this analysis, the wave x experimental group interaction was significant, F(1113) = 3.74, p > .03, η = .06, in addition to the main effect of testing wave, F(1113) = 28.46, p > .001, η = .34. These effects suggested that the experimental group improved more (1467 ms, 1193 ms, and 1191 ms for the three waves, respectively) than the control group (1357 ms, 1230 ms, and 1225 ms for the three waves, respectively) across testing waves. It is noted, however, that none of the between-group comparisons within waves ever reached significance. The pairwise within groups comparisons for possible session differences showed that, in the experimental group, performance at first, t = 6.61 (57), p < .001, and second posttest, t = 2.77 (57), p < .008, significantly exceeded performance at pretest; performance at the second posttest did not differ from the first posttest, t = .06 (57), p > .05, indicating a leveling off. Similar but weaker trends were observed in the control group, (t = 3.66, p < .001, 2.63, p < .01, and .12, p > .05, for the three comparisons, respectively). Therefore, there seems to be a weak trend for training to positively affect attention control but not pure speed (see Supplementary Table 3).

One might ask how the training effect above may relate to general intelligence rather than separate processes. To answer this question, factor scores were obtained for the first principal component of each testing wave that resulted from a factor analysis applied on the basic dimensions tested at each testing wave (mathematics, analogical, Raven-like, syllogistic reasoning and class inclusion, isolation of variables and hypothesis testing, mental rotation, speed and attention control, and short-term and working memory). These factor scores may be regarded as measures of g at each testing wave. These scores were subjected to a 2 (the two groups) × 3 (the three testing waves) ANOVA. The main effect of group was marginally significant, F(1113) = 3.24, $p < .07$, $\eta = .03$. The main effect of testing wave was nonsignificant, F(2112) = .001, $p > .05$, $\eta = .00$. However, the group × testing wave interaction was highly significant, F(2112) = 23.06, $p < .001$, $\eta = .29$. These results reflected the lack of difference between the two groups at first testing (.06 vs. −.06 for control and experimental group, respectively, F (1114) = .48, $p > .05$, $\eta = .00$), their large difference favoring the experimental group at first posttest (−.36 vs. .37 for control and experimental group, respectively, F(1114) = 16.68, $p < .001$, $\eta = .13$), and a marginally significant superiority of the experimental group at the second posttest (−.15 vs. .15 for control and experimental group, respectively, F (1114) = 2.61, $p < .10$, $\eta = .02$). Therefore, our intervention did somehow change “true g”, to the extent these scores reflect this construct.

2.2. Growth and generalization

The analyses above suggested clearly that the patterns of change across testing waves differed between domains. Growth modeling is the method of choice for mapping growth patterns when there are multiple testing waves. Moreover, growth modeling is more appropriate than other methods to reveal possible differences in the form of change caused by intervention in different process. In sake of this aim several growth models were separately applied on the scores attained by children in the control and the experimental group in each of the five cognitive domains: mathematical reasoning (i.e., number series, number analogies, and explicit grasp of principles underlying mathematical analogies), logical reasoning (i.e., deductive and class reasoning), analogical reasoning (i.e., verbal analogies and Raven-like matrices), spatial reasoning (i.e., mental rotation and image integration), and causal-scientific reasoning (i.e., isolation of variables and hypothesis testing). To ensure comparability across domains mean scores on each of these domains were transformed into z scores. Control and working memory were also included, after being transformed into z scores (speed was not used here because its strong covariation with attention control might cause collinearity problems to model estimation). Specifically, the mean z score for each of these processes was related to the intercept related to all processes to capture the possible influence of these processes on growth patterns. All models were tested in a 2-group set up to examine possible differences in the form of growth between the control and the experimental group that might be ascribed to intervention. The correlations between the variables used in these models are presented in Supplementary Tables 1–6.

The first model assumed complete similarity between the two groups. This model assumed linear growth across testing occasions and across domains. Specifically, in this model, there was one intercept factor set to 1 for all three testing waves across the five domains; there was also one slope factor set to 1, 2, and 3 for each testing wave, respectively, across the five domains. The fit of this model was poor, χ^2 (270) = 633.15, $p < .001$, CFI = .65, RMSEA = .12, model AIC = 93.15, suggesting that the assumption of linear growth across domains and groups was not tenable. The second model assumed that there was no growth in the control group and linear growth, as above, in the experimental group. Technically, the only difference between this and the first model was the dropping of the slope factor in the control group. The fit of this model, although slightly better than the first model, was also poor, χ^2 (271) = 624.81, $p < .001$, CFI = .67, RMSEA = .12, model AIC = 80.81. These two models suggest strongly that change is much more complicated than any simple model that would assume no change in the control group and linear change in the experimental group.

A more realistic model would assume some change in the control group, to reflect the influence of testing experience and a variable pattern of change in the experimental group to reflect the differential impact of training on the domain trained and the other domains, according to their similarity to the trained domain. A first approach in implementing this model would be to assume, first, that there is systematic change in the control group in mathematics, to reflect the fact that mathematics is an object of education where there is learning independent of the experiment. Second, change in the other domains in the control group would be limited, most expressed at third testing, to reflect the influence of repeated testing. To implement these assumptions in the model, the intercept factor of all measures was set to 1 in the control group. The slope was set to 1, 2, and 3 for the three testing waves in mathematics and to 0, 0, and 1 for the three waves in all other domains. Third, in the experimental group, there should be systematic change in the domain of mathematics to reflect both the effect of teaching at school and our training and also a relative drop of performance from first to second posttest, to reflect a wane out effect that is common in learning experiments. Fourth, there should also be change in the deductive and analogical reasoning in the experimental group to reflect generalization to inferential processes related to training. Finally, there should be limited change in scientific and spatial reasoning to reflect, in the fashion specified in the second model in concern to the control group, the effects of repeated testing. To implement this model, the intercept was set to 1 for all measures as above. The slope for mathematics, deductive, and analogical reasoning was set to 1, 2, and 3, for the three testing waves, respectively; the slope for scientific and spatial reasoning was set to 0, 0, and 1, for the three testing waves, respectively. The fit of this model, although still not acceptable, was better than any of the models above, χ^2 (266) = 566.06, $p < .001$, CFI = .72, RMSEA = .11, model AIC = 34.06, indicating that change is a multifaceted process; its different faces reflect variations in learning experiences and differences in cognitive domains.

To precisely capture the various faces of change, a dampening factor was introduced, in addition to the intercept and the slope factor. In growth modeling, dampening factors are introduced to represent possible changes in the slope factor at different testing intervals. In the present model, this factor was equal to the slope factor for waves 1 and 2, respectively. However, the value for the third wave was the relation between the difference of the first from the third and the first from the second wave. Therefore, this factor aimed to capture the relative drop of performance from the second to the third wave (see Bentler, 2006; Stoolmiller, 1995). In the control group, a dampening factor was assumed only for mathematics. In the experimental group a dampening factor was assumed for mathematics, analogical, and spatial reasoning. The error terms of the first testing across all domains were constrained to be equal across groups, assuming that measurements behaved the same in the two groups before intervention. The fit of this model was good, χ^2 (264) = 447.40, $p < .001$, CFI = .99, RMSEA = .08, model AIC = −80.60.

It is noted that the correlation between the intercept and slope was high and positive in the control group (.93); in the experimental group it was lower and negative (−.42). This indicated that, under conditions of spontaneous development, initial higher ability was associated with higher gain at later measures. Under conditions of guided development as attempted in the experimental group, initially lower performing children gained more from instruction. Interestingly, the relation between intercept and the dampening factor was positive and significant in the control group (.48) but negative and significant in the experimental group (−.59), indicating that the higher the gain caused by intervention the higher the drop at a later testing time. The various parameters and relations generated by this model are shown in Table 1. It is clear, therefore, that (i) growth is different between the control and the training group; (ii) it is stronger in the trained ability than others; (iii) it generalizes to procedurally similar abilities; (iv) it relatively weakens at delayed posttest. The models to be presented below will further specify these trends and explore the relations between processes.

2.3. Developmental and functional interactions

Structured means analysis is complementary to growth modeling. It enables one to specify the possible interactions between abilities, in addition to specifying the possible effects of learning. In addition, it enables one to specify the possible transfer of effects from specific to general abilities. Three models were tested. These models explored the effects of training from (i) pretest to the immediate posttest, (ii) pretest to the delayed posttest, and (iii) the immediate to the delayed posttest, respectively. The first model included the following scores: two mean scores for speed and two mean scores for control of processing, respectively, to stand for each of these two dimensions of processing efficiency at pretest; two mean scores to stand for verbal and numerical working memory at pretest, respectively; five scores, one for each thought domain, to stand for performance in the five thought domains at pretest; these three sets of scores were regressed on three separate factors, for speed or control, working memory, and Gf, respectively. There were also three sets of standardized gain scores (i.e., the difference between a later and an earlier testing in each of the scores above divided by the SD of the later score), to stand for change from pretest to immediate posttest and from immediate to delayed posttest in each of the eight dimensions above. These scores were regressed on factors corresponding to the factors above. Thus, these factors represent change from the one testing to the next in each of the various abilities. It is noted that speed and control were not used simultaneously in the same model. The one or the other was included in the model involving all other factors. This manipulation was considered necessary to specify the role of each of these two dimensions of processing efficiency, without causing co-linearity effects which unnecessarily burden model estimation. For space considerations, here we only present results from the models involving control. Finally, the error variances of the same measures across testing waves were allowed to correlate to control for possible systematicity in error variation across testing waves. The correlations between the variables used in these models are presented in Supplementary Tables 4–6. Means and standard deviations are presented in Supplementary Tables 1–3.

To specify structural relations, the working memory factor was regressed on the processing efficiency factor and the Gf factor was regressed on both of these factors. Also, each change factor was regressed on its corresponding performance factor. The efficiency change factor was also regressed on the working memory change factor; the working memory change factor was also regressed on the Gf change factor. Finally, all ability-specific factors were regressed on the performance intercept. The factor loadings and the variable-intercept relations were constrained to be equal between the control and the experimental group. The change intercept for all three change factors was set to 0 in the control group and it was left free to be estimated in the experimental group. In the second model all ability scores were the same as above. However, the change scores represented change from pretest to the delayed posttest. In the third model, the ability scores represented performance at the immediate posttest and the change scores represented change from the immediate to the delayed posttest. These change intercepts can be interpreted as effect sizes in Cohen's d terms. The structural relations in these models were the same as above (differences in df across models are due to the fact that correlations between error variances were dropped when non-significant). The fit of all three models was good (see fit indexes in Table 2). The main findings of these analyses are summarized in Table 2. They are as follows.

In concern to change, from pretest to immediate posttest, there were significant differences between the control and the experimental group in all domains but scientific reasoning. This was reflected in the fact that the difference between the two groups in Gf (.38) was also significant. Notably, the difference between the control and the experimental group in attention control (−.10) and working memory was also

Table 1
Growth model for change across abilities and the three testing waves.

	Control group						Experimental group					
Ability	Inter	Slope	Damp	T1	T2	T3	Inter	Slope	Damp	T1	T2	T3
Functions	1.92*	.00	.16*				1.55*	.04+	.32*			
	.04	.02	.05				.08	.03	.03			
Maths				.02	−.42	−.21				−.07	.39	.20
				.93	.95	.88				1.05	.88	1.08
Deductive				.07	.06	.05				−.11	−.07	−.07
				.94	1.05	1.03				1.05	.96	.99
Analogical				.02	−.35	−.13				−.03	.31	.10
				.93	.97	1.01				1.08	.92	.99
Scientific				.02	−.07	−.03				−.05	.02	.00
				1.02	1.01	.93				.99	.98	1.06
Spatial				.12	−.018	.01				−.12	.18	−.04
				.98	1.12	.86				1.04	.81	1.13
Control	1.36			-.17	.06	.05	1.46			.20	−.07	−.06
				.83	1.04	1.23				1.12	.96	.72
WM	1.92			.00	−.15	−.08	4.16			−.01	.14	.07
				.31	.48	.40				.47	.45	.58
Inter var	.04*						.26*					
	.02						.08					
Slope var	.02*						.03*					
	.01						.01					
Damp var	.13*						.01					
	.03						.01					
Inter R		.93*	.48					−.42	−.59			
Slope R			−.27						−.90			

Note 1: z scores (and SD below each z score) used for modeling at Times 1, 2, and 3. Raw scores and standard deviations are presented in Supplementary Tables 1-3. The correlations between the variables used in these models are presented in Supplementary Tables 4-6.

Table 2
Structural relations between change in WM, Gf, and the various domains as a function of training, time, and process.

Ability	Control group				Experimental group					
	Contr	WM	Gf	Gf ch	Contr	WM	Gf	WM ch	Gf ch	Intercept
T1→T2: χ^2 (238) = 290.39, p = .01, CFI = 1.0, RMSEA = .06										
WM	−.28*				−.12					−.03
Gf	–	.26*	–		–	.18*				−.12
Change										
Control	−.39*			–	−.72*					−.10*
WM	–	−.83*	–	–	–	.67*	–	–	.72*	.93*
Gf	–		.77*		–	.10	−.61*			.38*
Maths										.38*
Deduct										.12*
Analog										.20*
Scient										.06
Spatial										.08*
T1→T3:, χ^2 (237) = 323.27, p = .01, CFI = .94, RMSEA = .08										
WM	−.27				−.28*					.09*
Gf	−.37	.20			–	.31*				−.19
Change										
Control	−.23				−.41*	–		–	−.10	−.11*
WM		−.89*	.40		–	−.99*			.36*	.60*
Gf		–	1.00			.30*	−.42*			.20*
Maths										.20*
Deduct										.13*
Analog										.05
Scient										.10
Spatial										.01
T2→T3: χ^2 (234) = 317.39, p = .01, CFI = .99, RMSEA = .08										
WM	−.05				−.16					1.24*
Gf	–	.49*			–	.36*				.52*
Change										
Control	−.25			.28	−.58*			.20	–	.02
WM		−.65*	.44*			−.57*	.40*		.33*	−.01
Gf		−1.0	–			.27	−.36*			−.11
Maths										−.11
Deduct										.09*
Analog										−.08
Scient										.06
Spatial										−.03

Note 1: Intercept for control group was set to 0.
Note 2: The correlations between the variables used in these models are presented in Supplementary Tables 1–6.

significant (.93). In the second model capturing relations from pretest to the delayed posttest two effects dropped below significance: analogical (.05) and spatial thought (.01). The effect on mathematics (.20), deductive reasoning (.20), Gf (.20), and attention control (−.11), and WM (.60) were still significant. In the third model all effects were negative but one (deductive reasoning) indicating performance drop. The drop was significant only in analogical (−.08) and spatial thought (−.03). Interestingly, the difference in deductive reasoning (.09) indicated that the experimental group continued to rise. This is in line with the fact that in this model there was still a significant difference between the control and the experimental group in working memory (1.24) and Gf (.52).

How much change can an intervention bring about? Interestingly, this study suggested strongly that there was a developmental limit to how much change can occur. This was the upper level of ability associated with a particular developmental level. The closer an individual was to this level the less this individual gained from instruction directed to the attainment of this level. Thus, the limit marker of change was the individual accomplishment before the intervention. The higher the accomplishment the less the room left for change as a result of the intervention. This was suggested by the systematic but negative relation between Gf at a prior testing and change in Gf, in the experimental group (−.61, −.42, and −.36 in the three models, respectively).

How is change in various abilities mediated by other abilities? The answer to this question lies in the relations between change in each ability and change or prior state of others. Change in working memory was strongly mediated by change in Gf (.72, .36, and .33 in the three models in the experimental group, respectively). Change in Gf was mediated by the prior state of working memory (.10, .30, and .27, in the three models in the experimental group, respectively), although these relations were weaker. It is notable that change in Gf or in working memory was not related to change in control.

3. Discussion

In the introduction section we asked if learning to use general cognitive processes in a specific domain (i) augments fluid intelligence, (ii) transfers to domain-free representational and processing capacities, (iii) varies over time, and (iv) is constrained by developmental level. The answer is "yes" to all of these questions. Specifically, this study showed an interesting combination of changes associated with our intervention: Change in the domain of mathematical reasoning, which was the focus of intervention, was considerable at the immediate post-test but it was not sustainable in time. However, in line with findings presented by other researchers, the gains did transfer to domain-free analogical reasoning tasks and they proved sustainable (e.g., Klauer & Phye, 2008). Also, structured means analysis showed that these gains did generalize to other domains, such as deductive and spatial reasoning, that differ from the processes trained. Interestingly, gains in deductive reasoning continued to improve from second to third testing, when they dropped in other domains. Naturally, these gains were clearly expressed at the level of the latent construct standing for Gf.

Special attention is drawn to the transfer of effects to domain general processes reflecting processing and representational efficiency. This finding runs contrary to the first prediction derived from psychometric (Jensen, 1998; te Nijenhuis et al., 2007; te Nijenhuis et al., 2014) and neo-Piagetian developmental theory (Case, 1985; Pascual-Leone, 1970), which assume that the direction of causality runs from processing efficiency to reasoning. The transfer of effects to all PREM factors, working memory in particular, suggests that learning did go through from domain-geared inferential processes down to domain free indexes of g. It is stressed that the magnitude of change in working memory caused by change in Gf (52%, 13%, and 11% of variance of working memory change accounted for by Gf change in the three models presented above) was much larger than the magnitude of change in Gf accounted for by change in working memory (1%, 9%, and 7%, respectively). In agreement with the third prediction, these effects suggest that learning to reason tightened the whole system up, modifying indexes of g in the way spontaneous development does. These findings are, to our knowledge, novel in this field and align with the predictions of the ReConceP model. Notably, this model presumes that at the end of the rule-based concepts cycle the Gf-working memory relations are much stronger than the Gf-speed relations, and that when rule-based concepts are consolidated at the end of this cycle, speed improves, opening transition to the next cycle of principle-based concepts (Demetriou et al., 2013; Demetriou, Spanoudis & Shayer, 2014; Demetriou, Spanoudis, Shayer, van der Ven, Brydges, Kroesbergen, Podjarny & Swanson, 2014). This is precisely what the present intervention generated.

At the same time, the impact of the program, however respectable it appears if expressed in these terms, was not enough to modify thought processes that belong to a next cycle of development, namely the principle-based cycle. This was suggested by the fact that scientific reasoning remained impervious to learning experiences provided here. We would ascribe this finding to the fact that the aspects of scientific reasoning examined here (hypothesis formation and testing) belong to the principle-based cycle. The results discussed above about the impact of learning on PREM indexes of g suggest that the transition processes to this cycle might have been activated but the impact was not large enough to be expressed into actual cycle-specific reasoning patterns. This pattern of effects, in both its positive and its negative side, bears an important educational implication. Learning programs must cycle along the cycles of development themselves. That is, they must be tailored to successive developmental cycles through the end, each time boosting the processes that relate to the emergence and consolidation of each cycle. Affecting an earlier cycle would not necessarily transfer to the next cycle, even if it raises its level of readiness. This may render observed gains developmentally-specific to a large extent, suggesting that intelligence and related cognitive processes are constrained by powerful developmental cycles that set strong limits to learning. Thus, instruction-based change in various aspects of these processes may be temporary, as shown here. Sustainability and transfer of cognitive change to another cycle may also be constrained by brain-dependent developmental dynamics that may be more powerful than instruction based learning (Wendelken, Ferrer, Whitaker, & Bunge, 2015).

This interpretation may explain the distressing fade out effect of intelligence research, suggesting that interventions aiming to boost intelligence wane out in 2–3 years after the end of intervention. This interpretation expands rather than contradicts Protzko's (2015) interpretation that sustainability of learning gains require that the environment is continuously as demanding as the intervention environment. It suggests that in addition to the need to be continuously available, environmental demands must adapt to changing developmental needs until gains are locked into the system as habitual ways of dealing with problems. In fact, the results of the various growth models presented here suggest that the size and forms of transfer of gains from an intervention vary as a function of proximity between the process trained and the other processes. Therefore, interventions that may be relevant to the real life of education or clinical practice must accommodate the variable terrain of cognitive processes.

One might object here that these differences were caused by random variation in each of the two groups rather than the intervention addressed to the experimental group. Redic (2015) showed recently that training effects in studies aiming to raise cognitive processes such as working memory may be caused by a relative drop of performance in the control group rather than true change in the experimental group. This is clearly not the case here. We showed above that change in the control group, if any, was in the direction of increase rather than decrease. Moreover, the present study involved a rather large sample in both groups and comparisons were effected on the level of latent constructs built on multiple measures rather than on a few observed measures. Thus, the change observed here was genuine.

There is a message here for researchers looking for the holy grail of learning and intelligence boosting in a particular process, however central it might be, such as training executive control or working memory: However useful and effective these training programs may be (e.g., Barnett, 2011; Diamond, 2013), they should expand to include other processes as well, such as awareness of mental processes and how they connect to a particular problem domains (Christoforides et al., in press) if they would have a permanent and adulthood-important effect (e.g., Greiff et al., 2014). In fact, the present study showed that the influence going from relational thought to working memory was much stronger than the other way around. At the same time, even if targeting powerful inferential processes as attempted here, the results of learning may often be shaky and unstable. This may reflect, on the one hand, the fact that sustainability of gains requires sustainability of support for high level intellectual functioning. On the other hand, it may indicate that learning gains are developmentally-specific. That is, they may change a process at the level targeted, but they do not fully consolidate and automate unless they are embedded in the supportive frame of operating at a next higher level developmental cycle. The present study showed clearly that transfer to processes specific to the next cycle, such as scientific thinking, was not attained by our intervention. One might assume that the learning gains in the domain-trained might be more stable if interventions would also target these higher level abilities than the abilities of a particular level. Finally, this study also showed that the fade out effect may be more apparent than real if it co-exists with gains in other domains of functioning. That is, interventions may have beneficial effects discernible in other domains, even if the measures of the trained dimension eventually settles down to levels not very different from those taken before these interventions. Obviously, this study would have to be validated with training addressed to other processes, other developmental cycles, and longer durations that would ensure greater transfer and permanence of gains. Moreover, we would need to use methods of diagnosing and modeling gains that would be sensitive enough to locate them and precisely map them.

Acknowledgment

Special thanks are due to Roberto Colom, John Protsko, Michael Shayer and three anonymous reviewers for their constructive comments on earlier versions of this paper. This paper is based on the doctoral dissertation of the first author which was implemented under the supervision of the other authors. Statistical modeling reported here and the writing of this paper was the responsibility of the last author.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.intell.2016.02.005.

References

Barnett, W.S. (2011). Effectiveness of early educational intervention. *Science*, *333*, 975–978. http://dx.doi.org/10.1126/science.1204534.

Bentler, P.M. (2006). *EQS 6 structural equations program manual*. Encino, CA: Multivariate Software, Inc.

Brainerd, C.J. (1977). Cognitive development and concept learning: An interpretative review. *Psychological Bulletin*, *84*, 919–939.

Carroll, J.B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press.

Case, R. (1985). *Intellectual development: Birth to adulthood*. New York: Academic Press.

Cattell, R.B. (1963). Theory of fluid and crystallized intelligence: a critical experiment. *Journal of Educational Psychology*, *54*, 1–22.

Christoforides, M., Spanoudis, G., & Demetriou, A. Coping with logical fallacies: A developmental training program for learning to reason. Child Development. (in press)

Csapó, B. (1999). Improving thinking through the content of teaching. In J.H.M. Hamers, J.E.H. van Luit, & B. Csapó (Eds.), *Teaching and learning thinking skills* (pp. 37–62). Lisse: Swets & Zeitlinger.

Ceci, S. (1991). How much does schooling influence general intelligence and its cognitive components? *Developmental Psychology*, *27*, 703–722.

Demetriou, A., Christou, C., Spanoudis, G., & Platsidou, M. (2002). The development of mental processing: Efficiency, working memory, and thinking. *Monographs of the Society of Research in Child Development*, *67*(Serial Number 268).

Demetriou, A., & Kyriakides, L. (2006). A Rasch-measurement model analysis of cognitive developmental sequences: Validating a comprehensive theory of cognitive development. *British Journal of Educational Psychology*, *76*, 209–242.

Demetriou, A., Mouyi, A., & Spanoudis, G. (2008). Modeling the structure and development of g. *Intelligence*, *5*, 437–454.

Demetriou, A., Spanoudis, G., & Shayer, M., (2014a). Inference, reconceptualization, insight, and efficiency along intellectual growth: A general theory. *Enfance*, *issue 3*, *365–396*, http://dx.doi.org/10.4074/S0013754514003097

Demetriou, A., Spanoudis, G., Shayer, M., Mouyi, A., Kazi, S., & Platsidou, M. (2013). Cycles in speed-working memory-G relations: Towards a developmental-differential theory of mind. *Intelligence*, *41*, 34–50.

Demetriou, A., Spanoudis, G., Shayer, M., van der Ven, S., Brydges, C.R., Kroesbergen, E., ... Swanson, L.H. (2014b). Relations between speed, working memory, and intelligence from preschool to adulthood: Structural equation modeling of 14 studies. *Intelligence*, *46*, 107–121.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*, 135–168.

Efklides, A., Demetriou, A., & Gustafsson, J. -E. (1991). Training, cognitive change, and individual differences. In A. Demetriou, A. Efklides, & M. Shayer (Eds.), *Neo-Piagetian theories of cognitive development: Implications and applications for education* (pp. 122–143). London: Routledge.

Fry, A.F., & Hale, S. (1996). Processing speed, working memory, and fluid intelligence: Evidence for a developmental cascade. *Psychological Science*, *7*, 237–241.

Fry, A.F., & Hale, S. (2000). Relationships among processing speed, working memory and fluid intelligence in children. *Biological Psychology*, *54*, 1–34.

Greiff, S., Wüstenberg, S., Csapo, B., Demetriou, A., Hautamaki, J., Graesser, A., & Martin, R. (2014). Domain-general problem solving skills and education in the 21st century. *Educational Psychology Review*, *13*, 74–83.

Gustafsson, J. -E. (2008). Schooling and intelligence: Effects of track of study on level and profile of cognitive abilities. In P.C. Kyllonen, R.D. Roberts, & L. Stankov (Eds.), *Extending intelligence: Enhancement and new constructs* (pp. 37–59). New York: Lawrence Erlbaum Associates.

Gustafsson, J.E., & Undheim, J.O. (1996). Individual differences in cognitive functions. In D.C. Berliner, & R.C. Calfee (Eds.), *Handbook of educational psychology* (pp. 186–242). New York: Macmillan.

Hunt, E. (2011). *Human intelligence*. Cambridge: Cambridge University Press.

Inhelder, B., Sinclair, H., & Bovet, M. (1974). *Learning and the development of cognition*. London: Routledge & Kegan Paul.

Jaeggi, S.M., Buschkuehl, M., Jonides, J., & Perrig, W.J. (2008). Improving fluid intelligence with training on working memory. *PNAS*, *105*, 6829–6833.

Jensen, A.R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger.

Kail, R.V. (1991). Developmental change in speed of processing during childhood and adolescence. *Psychological Bulletin*, *109*, 490–501.

Kail, R.V. (2007). Longitudinal evidence that increases in processing speed and working memory enhance children's reasoning. *Psychological Science*, *18*, 312–313.

Klauer, K.J. (1998). Inductive reasoning and fluid intelligence. A training approach. In W. Tomic, & J. Kingma (Eds.), *Reflections on the concept of intelligence* (pp. 261–289). Greenwich, CT: JAI Press.

Klauer, K.J. (2014). Training des induktiven Denkens–Fortschreibung der Metaanalyse von 2008. *Zeitschrift für Pädagogische Psychologie*, *28*, 5–19.

Klauer, K.J., & Phye, G. (1994). *Cognitive training for children: A developmental program of inductive reasoning and problem solving*. Seattle: Hogrefe & Huber.

Klauer, K.J., & Phye, G. (2008). Inductive reasoning: A training approach. *Review of Educational Research*, *78*, 85–123.

Klauer, K.J., Willmes, K., & Phye, G.D. (2002). Inducing inductive reasoning: Does it transfer to fluid intelligence? *Contemporary Educational Psychology*, *27*, 1–25.

Kyllonen, P., & Christal, R.E. (1990). Reasoning ability is (little more than) working-memory capacity? *Intelligence*, *14*, 389–433.

Melby-Lervag, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology*, *49*, 270–291.

Nutley, S.B., Soderqvist, S., Bryde, S., Thorell, L.B., Humphreys, K., & Klingberg, T. (2011). Gains in fluid intelligence after training non-verbal reasoning in 4-year-old children: A controlled, randomized study. *Developmental Science*, *14*, 591–601. http://dx.doi.org/10.1111/j.1467-7687.2010.01022.x.

Pascual-Leone, J. (1970). A mathematical model for the transition rule in Piaget's developmental stages. *Acta Psychologica*, *63*, 301–345.

Pascual-Leone, J., & Johnson, J. (2011). A developmental theory of mental attention: Its applications to measurement and task analysis. In P. Barrouillet, & V. Gaillard (Eds.), *Cognitive development and working memory: A dialogue between neo-Piagetian and cognitive approaches (13–46)*. New York: Psychology Press.

Piaget, J. (1964). Development and learning. In R.E. Ripple, & V.N. Rockcastle (Eds.), *Piaget rediscovered (pp. 7–20)*. Cornell University Press.

Protzko, J. (2015). The environment in raising early intelligence: A meta-analysis of the fadeout effect. *Intelligence*, *53*, 202–210.

Redic, T.S. (2015). Working memory training and interpreting interactions in intelligence interventions. *Intelligence*, *50*, 14–20.

Rueda, M.R., Checa, P., & Combita, L.M. (2012). Enhanced efficiency of the executive attention network after training in preschool children: Immediate changes and effects after two months. *Developmental Cognitive Neuroscience*, *2S*, S192–S204.

Shayer, M., & Adey, P. (2002). *Learning intelligence: Cognitive acceleration across the curriculum from 5 to15 years*. Milton Keynes: Open University Press.

Shipstead, Z., Redic, T.S., & Engle, R. (2012). Is working memory training effective? *Psychological Bulletin*, *138*, 628–654. http://dx.doi.org/10.1037/a0027473.

Spearman, C. (1904). "General intelligence" objectively determined and measured. *The American Journal of Psychology*, *15*, 201–293.

Strauss, S. (1972). Inducing cognitive development and learning: A review of short-term training experiments I. The organismic developmental approach. *Cognition*, *1*, 329–357.

Stoolmiller, M. (1995). Using latent growth curves to study developmental processes. In J.M. Gottman (Ed.), *The analysis of change* (pp. 103–138). Mahwah, NJ: Erlbaum.

te Nijenhuis, J., Jongeneel-Grimen, B., & Kirkegaard, E.O.W. (2014). Are headstart gains on the g factor? A meta-analysis. *Intelligence*, *46*, 209–215.

te Nijenhuis, J., van Viane, A.E.M., & van der Flier, H. (2007). Score gains on g-loaded tests: No g. *Intelligence*, *35*, 283–300.

Wendelken, C., Ferrer, E., Whitaker, K.J., & Bunge, S.A. (2015). Fronto-parietal network reconfiguration supports the development of reasoning ability. *Cerebral Cortex*, *1-13*. http://dx.doi.org/10.1093/cercor/bhv050.