proceedings

Proceeding Paper

Enhancing Creative Thinking Through Coding: Preliminary Evidence from an Experimental Study in Primary and Lower-Secondary Education †

Federica Pelizzari *, Marta Giudici, Michele Marangi and Simona Ferrari

CREMIT (Research Centre for Education in Media, Innovation and Technology), Department of Pedagogy, Catholic University of the Sacred Heart, Largo Gemelli 1, 20123 Milan, Italy
* Correspondence: federica.pelizzari@unicatt.it
† Presented at the 1st International Online Conference on Education Sciences (IOCES 2026), Online, 15–17 June 2026.

Abstract

Coding and educational robotics are increasingly promoted in schools as catalysts for creative thinking, yet recent reviews suggest that this link is not automatic and depends critically on pedagogical design. Empirical evidence from Italian compulsory education remains scarce. We report preliminary findings from the Coding&Learning project, a quasi-experimental study involving 291 enrolled students (138 primary, 153 lower-secondary) and 23 trainers across four Italian schools. Verbal divergent thinking was assessed pre- and post-intervention with the Divergent Association Task (DAT); complete pre–post data were available for 217 students. All four groups improved between pre- and post-test, and within-group gains were statistically significant in both experimental groups (primary: +3.27 points, $p = 0.031$; lower-secondary: +2.61 points, $p = 0.007$) and in the lower-secondary control group. When experimental and control classes were compared directly—through gain-score *t*-tests and ANCOVA controlling for pre-test—the between-condition differences did not reach significance, and effect sizes were small (primary $g = 0.18$; lower-secondary $g \approx 0$), as is not unexpected for a brief, statistically underpowered preliminary study. At the primary level, a larger share of experimental students improved (66.7% vs. 52.3%). Trainers reported coherent perceived gains in dispositions adjacent to verbal divergent thinking. Findings are specific to verbal divergent thinking and should not be generalised to creativity as a broader construct.

Keywords: coding; creative thinking; educational robotics; divergent thinking; Divergent Association Task; computational thinking; primary education; lower-secondary education; preliminary study

Check for updates

Academic Editor: Mike Joy

Published: 2 September 2026

1. Introduction

The relationship between coding and creativity has become a central concern of contemporary educational research. On the one hand, coding and educational robotics are increasingly introduced in school curricula worldwide as core 21st-century competences, often presented as natural vehicles to foster creative thinking, problem solving and computational thinking (CT) [1–3]. On the other hand, recent empirical and theoretical work questions the assumption that coding per se automatically enhances creative dispositions. Woo and Falloon [4], in an exploratory study of middle-school students engaged in creative coding tasks, found that novice learners rarely employed systematic, computationally oriented strategies to solve coding problems; rather, they tended either to bypass problems by simplifying their stories or to rely on low-level trial-and-error strategies. The authors conclude that the popular notion of coding as a "natural vehicle" for computational thinking and creativity is empirically fragile and depends heavily on the pedagogical scaffolding provided.

A complementary picture emerges from the scoping review by Wang, Yang and Yeung [5], who synthesised 26 empirical studies on the interplay between CT and creativity published between 2006 and 2024. They identify two main relational patterns—creativity emerging within CT-oriented tasks, and CT emerging within creativity-oriented tasks—and stress that the cognitive mechanisms underlying both pathways are still poorly understood. In particular, they highlight the role of executive functions (working memory, cognitive flexibility, and inhibitory control), the importance of age-related differences, and the central role of pedagogical context in moderating learning effects. Similar conclusions emerge from recent meta-analytic work: Montuori et al. [6] found that programming activities produce moderate-to-large effects on creative thinking ($g \approx 0.73$), but with substantial heterogeneity across studies; Israel-Fishelson and Hershkovitz [7] show that cultivating creativity within CT tasks can, in turn, improve CT skills. For educational robotics specifically, the review and meta-analysis by Sapounidis et al. [8] reaches convergent conclusions, reporting positive effects on problem-solving and STEM learning in primary education while emphasising that outcomes depend strongly on the quality of pedagogical implementation. All these contributions converge on a common point: educational design matters as much as—or more than—the technological artefact itself.

In this study, creative thinking is understood, following Runco and Acar [9] and operationalised as in Olson et al. [10], as the capacity to generate ideas that are both original and appropriate to the task or context. The construct is here measured through verbal divergent thinking, a sub-component of creativity that captures the ability to produce semantically distant lexical associations. This dimension has the advantage of being measurable through a brief, language-comparable, and automatically scored instrument: the Divergent Association Task (DAT) [10]. The DAT has been applied to a growing range of educational populations, including elementary-school children [11], and adapted to multiple languages through cross-lingual semantic-distance models [12]. As with any single-method measure, however, the DAT addresses only one facet of creativity and should not be equated with creative ability as a whole.

The Coding&Learning project, promoted by CREMIT (Università Cattolica del Sacro Cuore) in collaboration with the ST Foundation and the NGO Acra, was designed precisely to investigate whether and how a structured coding and educational robotics programme—explicitly conceived as a meaningful, open-ended and collaborative learning environment—can foster creative thinking in primary and lower-secondary students. The project draws on three theoretical traditions: (i) constructionist and "lifelong kindergarten" approaches to creative computing [2,13]; (ii) the Creative Computational Problem Solving (CCPS) framework [14]; and (iii) socio-cultural and media-educational perspectives on technology in education [15]. These traditions converge on the view that coding fosters creativity not because of the technology itself, but because of the type of activity it makes possible—provided the pedagogical design supports imagination, iteration, collaboration, and reflection [16].

While the literature reviewed above identifies what is needed (design-based, open-ended tasks) and why it should work (cognitive mechanisms), empirical evidence from real-world classroom interventions explicitly grounded in constructionist principles—and especially from Italian compulsory education—remains scarce. Recent reviews of educational robotics in primary education [8] still report a predominance of short, lab-based interventions, and few studies have combined a standardised divergent-thinking measure with trainer-perception data in a single design. The present paper addresses this gap by reporting preliminary descriptive evidence from a four-week design-based intervention delivered in regular Italian classrooms, in which verbal divergent thinking was measured before and after the activity in both experimental and control classes.

Three research questions guided the study:

- RQ1. Does a structured coding and educational robotics programme affect students' verbal divergent thinking, as measured by the DAT, more than ordinary curriculum?
- RQ2. What differences emerge between primary and lower-secondary school students?
- RQ3. What elements of transformation in students' creative thinking are identified by the trainers who designed and delivered the programme?

The aim of the study is to provide preliminary empirical evidence on the conditions under which coding and educational robotics may support the development of creative thinking in compulsory education, and to contribute to the ongoing pedagogical debate on the design of "creative coding" interventions.

2. Materials and Methods

2.1. Research Design

The study adopted a quasi-experimental mixed-methods design with pre- and post-test data collection. Four schools (two primary and two lower-secondary), located in northern Italy, were involved. In each school, one experimental class and one control class participated; the experimental classes received the Coding&Learning intervention, while the control classes followed their ordinary curriculum during the same period. Class allocation to the experimental or control condition was made by the school in agreement with the research team, on the basis of teachers' availability and timetable constraints rather than randomisation; this is acknowledged as a methodological limitation (Section 4.5). Quantitative data from students were complemented by qualitative and quantitative data from trainers, gathered through pre- and post-course questionnaires.

2.2. Participants

A total of 291 students participated: 138 primary-school students (third grade, aged 8–9) and 153 lower-secondary students (second grade, aged 12–13), distributed across experimental and control classes as reported in Table 1. Twenty-three trainers—educators with prior experience in educational robotics and digital education—designed, delivered, and reflected on the intervention. Written informed consent was obtained from the legal guardians of all participating children, and age-appropriate verbal assent was sought from the children themselves before each DAT administration. The study was conducted in accordance with the Declaration of Helsinki and received institutional ethical clearance from the Università Cattolica del Sacro Cuore.

2.3. The Educational Intervention

The intervention consisted of an 8 h programme delivered in four 2 h sessions over four consecutive weeks. Each session followed the same four-phase structure to support reproducibility: (1) a short framing of the design challenge and its real-world context; (2) exploration in small collaborative groups, beginning with a more closed, guided task and moving progressively toward an open-ended one; (3) prototyping and iterative testing, in which groups built, ran and revised their solutions; and (4) a whole-class debrief in which groups presented and justified their choices. Within each session, students worked in small collaborative groups on open-ended design challenges that combined unplugged coding activities, block-based programming and educational robotics. Two parallel curricula were designed, one for each school level, sharing the same pedagogical principles but differing in content and complexity:

- Primary school (third grade). The intervention focused on sustainability, group work, and creative problem-solving. Students were invited to design an "ideal playground" for their community, using unplugged coding sequences and the Minecraft Education environment as design tools. The challenge required them to combine spatial reasoning, narrative imagination, and basic algorithmic thinking.
- Lower-secondary school (second grade). The intervention took the form of a mission on global issues (e.g., environmental degradation, accessibility). After analysing the strengths and weaknesses of a chosen context, students used an educational robot (programmable through a block-based environment) to prototype possible improvements. The challenge encouraged systemic thinking and the iterative refinement of solutions.

Both curricula were anchored in three pedagogical dimensions characteristic of the Coding&Learning model—a design framework developed at CREMIT and grounded in the Italian media-education tradition [15]—namely: a civic and media-educational dimension, a didactic-operational use of coding and robotics, and a pedagogical co-responsibility between students, teachers and trainers. Crucially, students were not asked to reproduce given solutions, but to imagine, prototype, test, and iterate their own—following the design-based logic identified by Woo and Falloon [4] and Resnick [2] as a precondition for creativity to emerge.

Across both curricula, scaffolding was deliberately faded from closed to open-ended tasks: early activities provided worked examples and step-by-step prompts, while later activities required students to define sub-goals, choose strategies, and evaluate their own solutions. Trainers acted as facilitators rather than solution-providers—following the Creative Computational Problem-Solving stance [14], they responded to requests for help with questions ("what have you tried?", "how could you test that?") and peer-referral rather than direct answers, precisely to discourage the bypass and copy-and-paste strategies documented by Woo and Falloon [4]. The classroom teacher was co-present throughout, providing contextual knowledge of the class, managing group dynamics and linking the activity to ongoing curricular work, while the trainer led the coding-specific content. As a concrete example, in the primary "ideal playground" challenge, groups specified requirements, expressed a layout as an unplugged sequence and then realised and iterated on it in Minecraft Education; at the lower-secondary level, groups analysed a real accessibility or environmental problem and programmed an educational robot to prototype and test a candidate improvement.

2.4. Measures

2.4.1. The Divergent Association Task (DAT)

Verbal divergent thinking was assessed through the Divergent Association Task (DAT) [10]. The DAT asks participants to produce ten nouns that are as semantically distant as possible. A score (range 0–100) is automatically computed by measuring the average semantic distance between the words, using pre-trained word-embedding models; higher scores indicate greater semantic flexibility. The DAT was validated as a brief, objective, and language-comparable indicator of verbal divergent thinking, and it correlates moderately with established creativity measures such as the Alternative Uses Task and the Bridge-the-Associative-Gap Task [10].

The original DAT was developed for adult English speakers. Two methodological considerations therefore apply to its use in the present study. First, recent work has shown that the DAT is applicable to elementary-school children: Ding et al. [11], in a sample of 348 Chinese elementary-school children, reported that DAT scores correlated positively with both the Alternative Uses Task and the Bridge-the-Associative-Gap Task, supporting the suitability of the instrument for this age range. Second, multilingual extensions of the DAT have been developed to make semantic-distance scoring comparable across languages [12]. In the present study, the Italian version of the task was administered, with semantic distance computed on an Italian word-embedding model; the same procedure was applied identically in the pre- and post-tests, so that any administration-related variance affected both measurements equally. To the best of the authors' knowledge, however, no formal psychometric validation of the DAT for Italian primary and lower-secondary populations was published. Cross-language psychometric equivalence cannot therefore be assumed, and this is acknowledged as a methodological limitation (Section 4.5).

It is important to acknowledge two intrinsic limitations of the DAT. First, the DAT captures only one component of creativity—verbal divergent thinking—and does not address figural creativity, originality of products, elaboration, or other facets of creative ability. Findings should be interpreted as specific to verbal-semantic flexibility. Second, recent validation work has questioned the strength of the DAT's convergent validity, reporting only weak-to-moderate correlations with other divergent and convergent thinking measures [12]. The DAT is therefore best understood as a complementary, not substitutive, indicator of divergent thinking.

2.4.2. Trainer Questionnaires

Trainers completed a pre- and post-course questionnaire investigating their perceptions of students along a set of learning dimensions. Each dimension was rated on a six-point Likert scale (1 = "not at all", 6 = "completely"). The pre-course form asked trainers what they expected students to learn through the coding experiences, while the post-course form used the same items to ask what trainers judged students had actually learned; the two administrations therefore share an identical item set and scale, enabling a direct pre–post comparison. The instrument covered sixteen student-learning dimensions, ranging from logical–analytical skills, independence and complex-problem reduction to creativity, critical thinking and collaboration. For each dimension, the aggregate value reported is the relative change in the mean trainer rating from the pre- to the post-course administration, expressed as a percentage; it summarises the direction and magnitude of perceived change across the 23 trainers and is not a student-level outcome. Because the trainers were also the designers and deliverers of the intervention, these ratings are susceptible to expectancy bias; they are therefore reported throughout as exploratory, complementary evidence rather than as inferential evidence of student change (see Sections 3.3 and 4.4).

2.5. Procedure and Data Analysis

The DAT was administered individually to all students at the beginning of the first session (pre-test) and at the end of the fourth session (post-test). Both administrations were supervised by the classroom teacher and one project trainer; conditions were kept as similar as possible across administrations and across groups. Trainer questionnaires were administered immediately before and after the four-session cycle.

For each of the four groups (primary control, primary experimental, lower-secondary control, and lower-secondary experimental), analyses were based on students with valid, complete DAT scores at both pre- and post-test (n = 65, 45, 41 and 66, respectively; 217 of the 291 enrolled students). Incomplete or invalid responses—chiefly protocols with fewer than the seven valid words required for DAT scoring—were excluded listwise. For each group, we computed means and standard deviations at pre- and post-test, the percentage-point gain (post−pre) and the relative gain, together with a paired-samples t-test of the pre–post change and its standardised effect size (dz). To address RQ1, experimental and control classes were compared at each school level in two complementary ways: (i) an independent-samples Welch *t*-test on individual gain scores, with the standardised mean difference reported as Cohen's d and Hedges' g with 95% confidence intervals; (ii) an analysis of covariance (ANCOVA) on post-test scores with pre-test score as covariate and condition as factor, adjusting for baseline differences and mitigating regression-to-the-mean. Baseline equivalence of the conditions was checked with Welch *t*-tests on pre-test scores, and the proportion of students whose individual DAT score increased was computed for all four groups. Students were nested within intact classes and schools; with only two classes per condition per level, this clustering could not be formally modelled (e.g., through multilevel models), and the analyses treat students as the unit of analysis—a recognised limitation of the design (Section 4.5). Analyses were conducted in Python 3.11 with SciPy 1.11; trainer-questionnaire results are reported descriptively as exploratory evidence.

3. Results

3.1. Average DAT Scores: Experimental vs. Control Groups

Sample sizes, means, standard deviations, absolute (percentage-point) gains, and relative gains for the four groups are reported in Table 1. Average DAT scores improved in all four groups between pre- and post-test. Within-group paired-samples *t*-tests showed that the pre–post increase was statistically significant in the primary experimental group (M = 72.65 → 75.92; +3.27 pp; t (44) = 2.23, $p = 0.031$, dz = 0.33), the lower-secondary control group (M = 75.33 → 77.88; +2.55 pp; t (40) = 2.16, $p = 0.037$, dz = 0.34) and the lower-secondary experimental group (M = 74.26 → 76.87; +2.61 pp; t (65) = 2.79, $p = 0.007$, dz = 0.34), but not in the primary control group (M = 73.41 → 74.88; +1.47 pp; t (64) = 1.16, $p = 0.25$). In descriptive terms, the experimental group's gain exceeded its control group's at the primary level (+3.27 vs. +1.47 pp), whereas at the lower-secondary level, the experimental and control gains were of very similar magnitude (+2.61 vs. +2.55 pp).

Table 1. Pre- and post-test Divergent Association Task (DAT) scores by group (analytic sample: students with valid pre- and post-test scores), with standard deviations, absolute (percentage-point) and relative gain, and the paired-samples pre–post test. M = mean; SD = standard deviation.

Group	*n*	Pre-Test M (SD)	Post-Test M (SD)	Gain (pp)	Rel. Gain (%)	Pre–Post t (*p*)
Primary, control	65	73.41 (8.29)	74.88 (5.99)	+1.47	+2.00	1.16 (0.25)
Primary, experimental	45	72.65 (6.97)	75.92 (6.33)	+3.27	+4.50	2.23 (0.031)
Lower-secondary, control	41	75.33 (6.17)	77.88 (5.23)	+2.55	+3.38	2.16 (0.037)
Lower-secondary, experimental	66	74.26 (5.44)	76.87 (5.33)	+2.61	+3.51	2.79 (0.007)

Crucially, the direct experimental-versus-control comparison did not support a differential effect of the intervention on DAT scores. On gain scores, the between-condition difference was not statistically significant at either level, and effect sizes were small to negligible: for primary school we have Welch t = 0.93, $p = 0.36$, and Hedges g = 0.18 (95% CI [−0.20, 0.56]); for lower-secondary school we have Welch t = 0.04, $p = 0.97$, and Hedges g = 0.01 (95% CI [−0.38, 0.40]). The ANCOVA on post-test scores, controlling for pre-test, led to the same conclusion: the adjusted effect of condition was not significant in primary school (b = +1.01, $p = 0.40$; adjusted post-test means 75.91 experimental vs. 74.89 control) or in lower-secondary school (b = −0.96, $p = 0.37$; adjusted means 76.89 vs. 77.85). Experimental and control groups did not differ at baseline (primary: $p = 0.60$; lower-secondary: $p = 0.36$), so the pattern is unlikely to be an artefact of pre-test non-equivalence. These results should be read as those of a small, statistically underpowered preliminary study:

they are consistent with a modest, favourable trend at primary level but do not, on their own, establish an effect of the intervention on verbal divergent thinking. The descriptive pattern is illustrated in Figure 1.

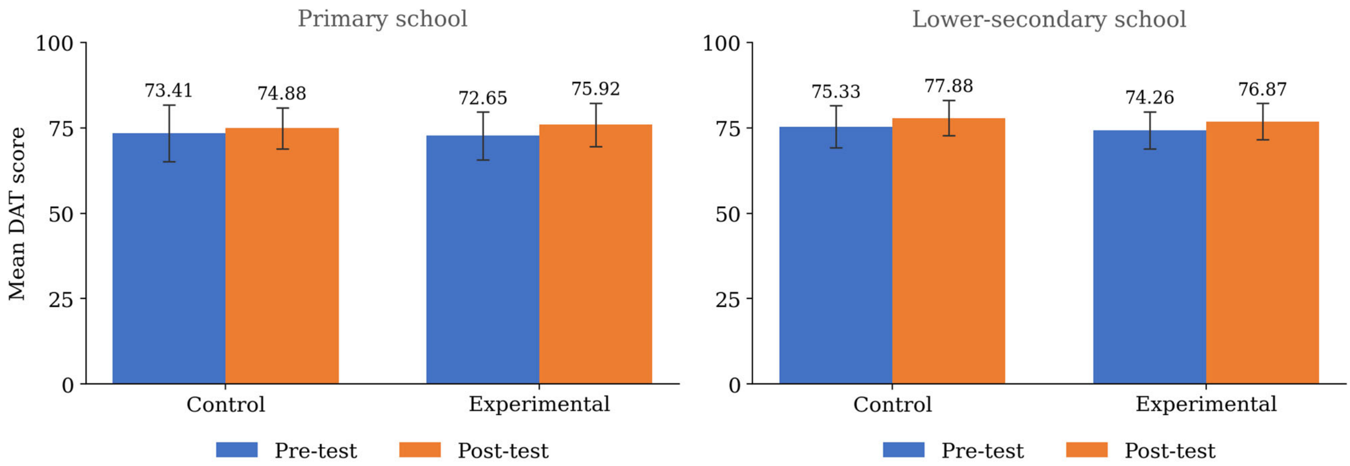

Figure 1. Mean DAT scores at pre- and post-test, by group and school level (error bars: ±1 SD). Analytic sample: primary control $n = 65$, primary experimental $n = 45$; lower-secondary control $n = 41$, lower-secondary experimental $n = 66$.

Two methodological points require explicit comment. First, an earlier version of this manuscript reported identical pre-test means (M = 72.64) for the two experimental groups. On verification against the individual-level data, this was traced to a transcription error: the value 72.64 had been duplicated from the primary experimental group (whose pre-test mean is M = 72.65), whereas the lower-secondary experimental pre-test mean is in fact M = 74.26. All values in Table 1 and in the analyses above are computed directly from the individual-level dataset. Second, because the analytic sample comprises students with valid DAT protocols at both time points, the group sizes ($n = 41$–66) are smaller and more unbalanced than the enrolled cohort; combined with the four-week duration and the clustered design, this limits statistical power and is the most plausible reason why a consistent favourable direction at primary level does not translate into a statistically detectable between-group effect.

3.2. Distribution of Individual Improvements (Primary School)

Beyond group means, we examined the proportion of students whose individual DAT score increased between pre- and post-test, now available for all four groups (Figure 2). At the primary level, the pattern favoured the intervention: 66.7% of experimental students improved, compared with 52.3% of control students—a difference of roughly 14 percentage points, indicating that the primary-level gain was distributed across learners rather than concentrated in a few high performers. At the lower-secondary level, however, the two conditions were comparable, with a similar share improving in the control (65.9%) and experimental (63.6%) groups. This mirrors the DAT results above: a favourable but non-significant signal at the primary level, and no experimental advantage at the lower-secondary level.

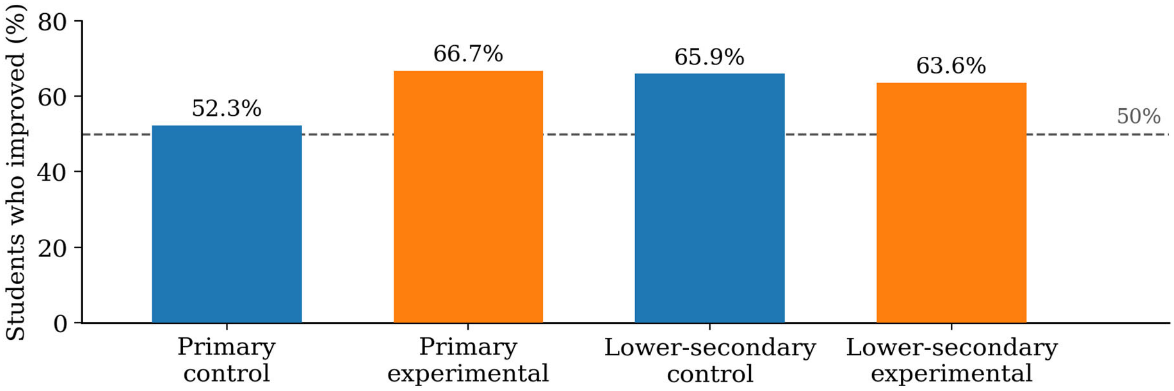

Figure 2. Proportion of students whose individual DAT score increased between pre- and post-test, by group. Blue bars denote control groups and orange bars denote experimental groups; the dashed line marks 50%.

3.3. Trainers' Perceptions of Student Change (Exploratory)

The trainer questionnaires provided a complementary, exploratory perspective on the intervention (Figure 3). On the sixteen-dimension instrument (six-point scale), trainers reported perceived increases across multiple competences, with the largest perceived gains concentrated on dispositions conceptually adjacent to verbal divergent thinking. Aggregate percentage differences—reported here as descriptive summaries of trainer perceptions, not as student-level outcomes—indicated increases in students' perceived capacity to make generalisations from specific problems (+33%), work independently (+24%), reduce complex issues to simpler aspects (+23%), develop logical–analytical skills (+22%), experiment with different strategies (+18%), gain basic programming skills (+18%) and develop creativity (+14%). Smaller positive perceived changes were reported on critical thinking (+10%), being more careful (+11%) and expressing themselves more freely (+5%). A small number of dimensions showed negative perceived changes—most notably producing artefacts (−14%) and, marginally, having fun (−5%) and working in groups (−3%)—which trainers attributed in open-ended answers to the higher cognitive demand of design-based tasks compared with more familiar play-oriented activities.

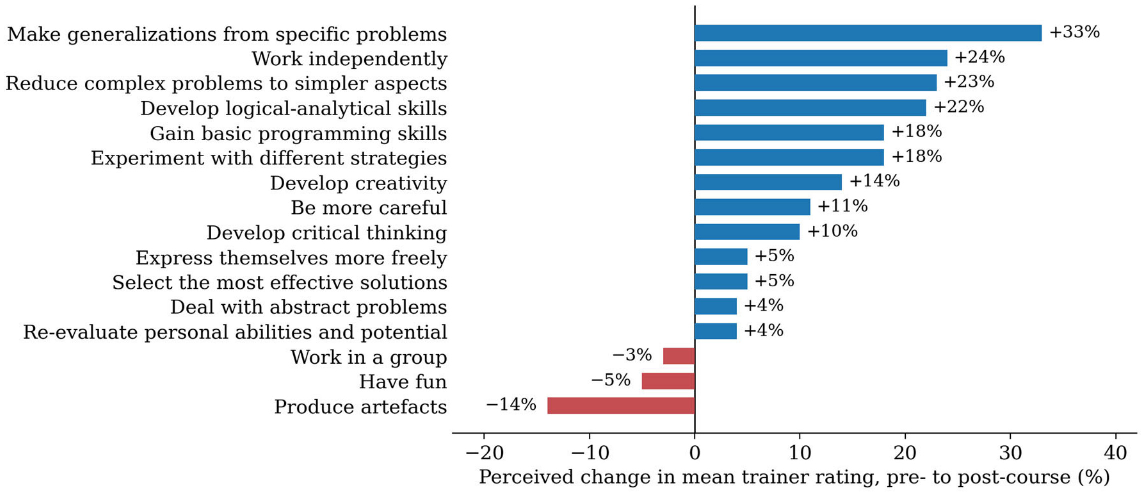

Figure 3. Trainers' perceptions of change in students between pre- and post-course (n = 23 trainers), across 16 learning dimensions. Values are relative changes in mean trainer ratings on a six-point scale and are exploratory (see Section 2.4.2). Blue bars indicate positive perceived changes; red bars indicate negative perceived changes.

https://doi.org/10.3390/proceedings2026145009

Open-ended responses converged on a recurring qualitative theme: trainers described coding not merely as a technical activity, but as a medium through which students could exercise imagination, justify choices to peers, and iteratively refine their ideas. Several trainers explicitly contrasted the project's design-based approach with their previous experiences of coding as a sequence of exercises with pre-defined solutions—a contrast that resonates with the diagnoses of Woo and Falloon [4] and Resnick [13].

4. Discussion

4.1. Coding Can Support Creative Thinking—Under Specific Conditions

The evidence reported here supports a cautiously stated, two-part conclusion. First, verbal divergent thinking, as measured by the DAT, improved significantly from pre- to post-test in both groups that received the structured coding and educational robotics programme, and trainers independently reported substantial perceived gains in several creativity-adjacent dispositions. Second, and equally important, the intervention could not be shown to raise DAT scores more than the ordinary curriculum: the direct experimental-versus-control contrast was non-significant at both school levels, with small effects, and only at primary level did the descriptive pattern (larger mean gain; larger share of improving students) favour the intervention. These findings are exploratory and statistically underpowered; they motivate, rather than confirm, an effect of the programme on verbal divergent thinking, and call for replication in larger, adequately powered and appropriately clustered samples.

These findings should not be read as a confirmation of the "naïve" view that coding per se enhances creativity. Rather, they align with the more cautious position advanced by Woo and Falloon [4], Wang et al. [5] and recent meta-analytic work [6,7]: any positive effect of coding on creative thinking is likely to depend on the way the activity is designed and scaffolded, rather than on the technology itself. The intervention reported here was explicitly built around design-based, open-ended, and collaborative principles, and the observed pattern is plausibly attributable to that pedagogical scaffolding rather than to the specific tools used (Minecraft Education, the educational robot)—a possibility that future analyses with appropriate process measures could begin to test.

4.2. Convergence Between Quantitative and Qualitative Evidence

A relative strength of the study lies in the partial convergence between two independent sources of evidence: DAT scores from students and perceptions from trainers. The DAT, a brief and language-comparable measure of verbal divergent thinking, captured significant within-group gains in the experimental groups, even though these gains did not significantly exceed those of the control groups. The trainer questionnaires, which targeted a much broader set of dispositions, identified a coherent pattern in which the largest perceived gains clustered around generalisation, independence, complexity reduction, experimentation and creativity. The convergence is theoretically meaningful: these are precisely the dispositions that the literature on Creative Computational Problem Solving [14] and the constructionist tradition [2,13] would predict to emerge from design-based coding tasks. The convergence is, however, partial and indirect—the DAT and the trainer questionnaire measure different constructs at different levels of granularity, and the trainer data are exploratory—so the consistency between the two should be read as a heuristic, not as a triangulated confirmation.

4.3. Differences Between Primary and Lower-Secondary School

A second pattern concerns differences between school levels. Once the pre-test transcription error is corrected, the more favourable—though still non-significant—pattern is found at primary rather than lower-secondary level: the primary experimental group showed the largest mean gain (+3.27 pp) and the clearest separation from its control, whereas at lower-secondary level, the experimental and control groups improved to a very similar extent (+2.61 vs. +2.55 pp). Any interpretation must remain tentative given the absence of a significant between-group effect, but several post hoc hypotheses can be considered. The lower-secondary control group itself improved significantly, which is consistent with a test–retest or maturation effect that may be more pronounced at this age and could mask an intervention-specific contribution. Older students may also have a wider vocabulary and denser conceptual associations, potentially compressing the room for measurable change on a semantic-distance task. Finally, as noted by Wang et al. [5] and in line with developmental evidence on executive functions, working memory and cognitive flexibility develop substantially between ages 8 and 13 and may moderate the impact of CT-related interventions. These remain speculative hypotheses to be tested in future research with more sensitive measures and adequate statistical power.

4.4. The Role of Pedagogical Design

A central pedagogical implication of the study—even at this preliminary stage—is that the quality of the activity appears to matter more than its mere presence on the timetable. Woo and Falloon [4] showed that, in the absence of explicit scaffolding, novice students tend to bypass coding problems—reshaping the task to avoid the cognitive cost of solving them—or to rely on inefficient strategies such as trial and error and copy-and-paste. The intervention reported here was structured precisely to discourage such bypass strategies: each session presented students with a non-trivial design challenge, and trainers were instructed not to provide pre-packaged solutions but to support students through questioning and peer reflection. The pattern of trainer-reported perceived gains—especially the increases in independence (+24%) and experimentation (+18%)—is consistent with the interpretation that this pedagogical stance had some effect. It must be noted, however, that the trainers who reported these perceived gains were also the designers and deliverers of the intervention; the convergence between pedagogical intention and reported outcome should therefore be interpreted with appropriate caution, and the inferential strength of this conclusion is limited by the design.

4.5. Limitations

Several limitations should be acknowledged. First, although inferential analyses are now reported, the study is small and statistically underpowered: the analytic groups ($n = 41$–66) are modest and unbalanced, and the four-week intervention is brief, so a genuine but small effect could plausibly go undetected. Second, students were nested within a small number of intact classes and schools (two classes per condition per level); this clustering could not be formally modelled, and treating students as independent may understate standard errors—multilevel modelling on a larger sample of classes is needed. Third, the DAT assesses only verbal divergent thinking and has shown only weak-to-moderate convergent validity with other creativity measures, so findings should be interpreted as specific to verbal-semantic flexibility rather than creativity as a whole; the DAT has also not been formally validated in Italian primary and lower-secondary populations, limiting the interpretation of absolute scores.

The study also used a non-randomised quasi-experimental design with a convenience sample, so pre-existing group differences and limited generalisability cannot be excluded, even though experimental and control groups were statistically equivalent at baseline on the DAT. Trainer data were collected from a small, self-selected group of educators who both designed and delivered the intervention, making expectancy and selection biases plausible; these data are therefore treated as exploratory throughout. Finally, the intervention was brief (8 h over four weeks), and the trainers were experienced specialists, so the persistence and transferability of any effects, and their replicability with non-specialist teachers, remain uncertain.

5. Conclusions and Future Directions

This paper contributes to the empirical and pedagogical debate on coding and creativity in three respects. First, it provides preliminary evidence, based on the DAT, that a structured coding and educational robotics programme designed around constructionist and design-based principles is associated with statistically significant within-group gains in verbal divergent thinking in both primary and lower-secondary students—while, in a transparent test of the stronger causal claim, it finds no statistically significant advantage of the programme over the ordinary curriculum, with a favourable but non-significant trend confined to primary school. Second, it shows that these gains are partially coherent with trainers' (exploratory) perceptions of change in creativity-adjacent dispositions. Third, and most importantly, it offers an empirical illustration—within Italian compulsory education—of a position increasingly visible in the international literature [4–7]: coding may support creative thinking when it is embedded in meaningful, open-ended and collaborative challenges, supported by a pedagogical design that values imagination, iteration and reflection over the reproduction of pre-defined solutions. The present data neither confirm nor refute that hypothesis; they specify what an adequately powered test of it would require.

Several directions for future work follow directly from the limitations identified above. (i) The present analyses will be scaled up to a larger, adequately powered sample with more classes per condition, enabling multilevel (mixed-effects) models that explicitly account for the clustering of students within classes and schools. (ii) The DAT will be complemented with task-based and observational measures of creativity—including figural divergent thinking and the rating of students' artefacts—to provide a more complete picture of creative development. (iii) A longitudinal follow-up will assess the persistence of any observed gains. (iv) The trainer questionnaire instrument will be revised to support psychometric analysis (Cronbach's alpha, factor structure) and to reduce expectancy bias, for example, by complementing designer-trainers' ratings with independent observers. (v) Replication with non-specialist classroom teachers, in line with the concerns raised by Woo and Falloon [4], will be essential to test the transferability of the pedagogical model. (vi) Qualitative analysis of trainers' open-ended responses will be used to refine the pedagogical principles of the Coding&Learning model and to make them transferable to mainstream classrooms.

In a school landscape where coding is increasingly framed as a generic "21st-century competence", the present study offers a modest but specific contribution: a reminder that coding may be most valuable for creative thinking when it is treated not only as structure and sequence, but also as creative construction.

Author Contributions: Conceptualization, F.P.; methodology, F.P.; investigation, F.P. and M.M.; data curation, F.P. and M.G.; formal analysis, F.P.; writing—original draft preparation, F.P.; writing—review and editing, M.G., M.M., and S.F.; supervision, S.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the ST Foundation and the NGO Acra; no specific grant number was assigned. The APC was funded by the authors.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and received institutional ethical clearance from the Università Cattolica del Sacro Cuore. The project was carried out within the established research and teacher-training activities of CREMIT.

Informed Consent Statement: Written informed consent was obtained from the legal guardians of all participating minors. Children were also informed about the study in age-appropriate terms before participation, and verbal assent was sought before each DAT administration.

Data Availability Statement: The anonymised dataset supporting the descriptive findings of this study is available from the corresponding author upon reasonable request, subject to compliance with the privacy and confidentiality conditions set by the participating schools. The full individual-level dataset will be deposited in an open repository alongside the planned follow-up study.

Acknowledgments: The authors thank the four participating schools, their head teachers, classroom teachers and—above all—the students and trainers who contributed to the Coding&Learning project. Special thanks are due to the ST Foundation and to the NGO Acra for their collaboration in designing the intervention.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analysis, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Wing, J.M. Computational thinking. *Commun. ACM* **2006**, *49*, 33–35. [CrossRef]
2. Resnick, M. *Lifelong Kindergarten: Cultivating Creativity Through Projects, Passion, Peers, and Play*; MIT Press: Cambridge, MA, USA, 2017.
3. Bocconi, S.; Chioccariello, A.; Kampylis, P.; Dagienė, V.; Wastiau, P.; Engelhardt, K.; Earp, J.; Stupurienė, G. *Reviewing Computational Thinking in Compulsory Education: State of Play and Practices from Computing Education*; Publications Office of the European Union: Luxembourg, 2022. [CrossRef] [PubMed]
4. Woo, K.; Falloon, G. Problem solved, but how? An exploratory study into students' problem solving processes in creative coding tasks. *Think. Ski. Creat.* **2022**, *46*, 101193. [CrossRef]
5. Wang, J.; Yang, W.; Yeung, M.K. Cognitive foundations in the interplay between computational thinking and creativity: A scoping review. *Think. Ski. Creat.* **2025**, *55*, 101729. [CrossRef]
6. Montuori, C.; Gambarota, F.; Arfé, B.; Tzuriel, D.; Pozzoli, T.; Altoè, G.; Mammarella, I.C. The cognitive effects of computational thinking: A systematic review and meta-analytic study. *Comput. Educ.* **2024**, *210*, 104961. [CrossRef]
7. Israel-Fishelson, R.; Hershkovitz, A. Cultivating creativity improves middle school students' computational thinking skills. *Interact. Learn. Environ.* **2024**, *32*, 431–446. [CrossRef]
8. Sapounidis, T.; Rodríguez-Aguilar, J.A.; Stamovlasis, D.; Demetriadis, S. Educational robotics and STEM in primary education: A review and a meta-analysis. *J. Res. Technol. Educ.* **2024**, *56*, 462–490. [CrossRef]
9. Runco, M.A.; Acar, S. Divergent thinking as an indicator of creative potential. *Creat. Res. J.* **2012**, *24*, 66–75. [CrossRef]
10. Olson, J.A.; Nahas, J.; Chmoulevitch, D.; Cropper, S.J.; Webb, M.E. Naming unrelated words predicts creativity. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2022340118. [CrossRef] [PubMed]
11. Ding, G.; He, Y.; Yi, K.; Li, S. Using the Divergent Association Task to measure divergent thinking in Chinese elementary school students. *Think. Ski. Creat.* **2024**, *52*, 101503. [CrossRef]
12. Beaty, R.E.; Johnson, D.R. Multilingual semantic distance: Automatic verbal creativity assessment in many languages. *Psychol. Aesthet. Creat. Arts* **2023**, *17*, 668–679. [CrossRef]
13. Resnick, M.; Rusk, N. Coding at a crossroads. *Commun. ACM* **2020**, *63*, 120–127. [CrossRef]
14. Chevalier, M.; Giang, C.; Piatti, A.; Mondada, F. Fostering computational thinking through educational robotics: A model for creative computational problem-solving. *Int. J. STEM Educ.* **2020**, *7*, 39. [CrossRef]
15. Ferrari, S.; Pelizzari, F. Coding and creativity. In *Pedagogy, Learning, and Creativity*; IntechOpen: London, UK, 2023. [CrossRef]
16. Pelizzari, F.; Ferrari, S. Exploring Coding and Educational Robotics in Primary Schools: Results and Perspectives from an Action Research Approach to Teaching Innovation. In *ATEE Spring Conference 2024—Teacher Education Research in Europe: Trends, Challenges, Practices and Perspectives*; Università degli Studi di Bergamo: Bergamo, Italy, 2025; pp. 278–292. [CrossRef]