

The Effect of Philosophy for Children (P4C) Activities on the Caring Thinking Skills of Middle School Students

Çocuklar İçin Felsefe (P4C) Etkinliklerinin Ortaokul Öğrencilerinin Özenli Düşünme Becerilerine Etkisi

Esra ÇAKAR ÖZKAN^{1*}  Aysel Ceren ÖZDEMİR² 

¹ Burdur Mehmet Akif Ersoy Üniversitesi, Eğitim Fakültesi, Burdur, Türkiye

² Burdur Mehmet Akif Ersoy Üniversitesi, Eğitim Bilimleri Enstitüsü, Burdur, Türkiye

ABSTRACT

This study examined the effect of Philosophy for Children (P4C) activities on the caring thinking skills of sixth-grade middle school students. The study was conducted using a pretest-posttest control group quasi-experimental design. The study group consisted of 65 sixth-grade students enrolled in two classes at a public middle school in Burdur during the 2024-2025 academic year, with one class assigned as the experimental group (n = 32) and the other as the control group (n = 33). Over eight weeks, the experimental group received a P4C curriculum designed by the researchers to develop the five dimensions of caring thinking (appreciative, affective, active, normative, and empathic), while the control group continued with the existing national curriculum. Data were collected using the Caring Thinking Skills Inventory and analyzed through one-way ANCOVA; for the sub-dimensions in which the homogeneity of regression slopes assumption was violated, models including the Group $\times$ Pretest interaction were estimated, and the Mann-Whitney U test on gain scores was reported as a supplementary sensitivity analysis. The findings revealed a statistically significant difference favoring the experimental group in the total caring thinking score, with a large group effect, and in the normative and empathic sub-dimensions. In the appreciative, affective, and active sub-dimensions, the experimental group scored higher at average pretest levels, but the group difference diminished as pretest scores increased. The results indicate that P4C activities can meaningfully develop the affective dimension of thinking, and recommendations were offered regarding the applicability of the caring thinking dimension of the 4C model at the middle school level.

Keywords Philosophy for Children (P4C), caring thinking, thinking skills, middle school students

ÖZET

Bu çalışmada, Çocuklar için Felsefe (P4C) etkinliklerinin ortaokul altıncı sınıf öğrencilerinin özenli düşünme becerileri üzerindeki etkisi incelenmiştir. Araştırma, ön test-son test kontrol gruplu yarı deneysel desen kullanılarak yürütülmüştür. Araştırmanın çalışma grubunu, 2024-2025 eğitim-öğretim yılında Burdur ilindeki bir devlet ortaokulunun iki şubesinde öğrenim gören 65 altıncı sınıf öğrencisi oluşturmakta olup şubelerden biri deney (n = 32), diğeri kontrol (n = 33) grubu olarak belirlenmiştir. Sekiz hafta boyunca deney grubuna, özenli düşünmenin beş alt boyutunu (takdir edici, duyuşsal, aktif, normatif ve empatik) geliştirmeyi amaçlayan ve araştırmacılar tarafından geliştirilen P4C program tasarısı uygulanırken kontrol grubu mevcut resmi öğretim programına devam etmiştir. Veriler Özenli Düşünme Becerileri Envanteri ile toplanmış ve tek yönlü ANCOVA ile analiz edilmiştir. Regresyon eğimlerinin homojenliği varsayımının karşılanmadığı alt boyutlarda Grup $\times$ Ön Test etkileşimini içeren modeller kestirilmiş, ek bir duyarlılık analizi olarak da kazanım puanları üzerinde Mann-Whitney U testi raporlanmıştır. Bulgular, özenli düşünme toplam puanında büyük bir grup etkisiyle, normatif ve empatik alt boyutlarda ise deney grubu lehine istatistiksel olarak anlamlı bir fark olduğunu ortaya koymuştur. Takdir edici, duyuşsal ve aktif alt boyutlarında deney grubu ortalama ön test düzeylerinde daha yüksek puan almış, ancak ön test puanları arttıkça gruplar arasındaki fark azalmıştır. Sonuçlar, P4C etkinliklerinin düşünmenin duyuşsal boyutunu anlamlı biçimde geliştirebileceğini göstermekte olup 4C modelinin özenli düşünme boyutunun ortaokul düzeyinde uygulanabilirliğine ilişkin önerilerde bulunulmuştur.

Anahtar Kelimeler Çocuklar için felsefe, özenli düşünme, düşünme becerileri, ortaokul öğrencileri

Received: [21.06.2026] Revised: [30.09.2026] Accepted: [14.10.2026]

Corresponding author: Esra ÇAKAR ÖZKAN · esracakar@mehmetakif.edu.tr

Article Note: This study was presented as an oral presentation at the 5th International 21st Century Educational Research Congress (INER Congress 2026).

1. Introduction

Philosophy is an expression of the human search to comprehend both the self and the world in which one exists. Throughout history, the human being has sought to understand the reasons on which life is grounded, how it is sustained, and according to which criteria it is shaped. By virtue of its continually changing and developing nature, this process of inquiry persists as a dynamic intellectual activity that continues throughout the lifespan (Lafcı-Tor, 2023). Philosophy makes it possible to conduct this process of inquiry in an orderly and coherent manner; however, for the capacity for philosophical thinking to develop in individuals, it is of great importance that this education be imparted at an early age. Indeed, children are, by their very nature, individuals who are curious, who ask questions, and who strive to make sense of their surroundings (Taşdelen, 2014).

In this context, philosophy assumes an important role in the development of individuals' higher-order thinking skills (Bloom, 1956). Higher-order thinking skills have long occupied a significant place in educational programs in both theoretical and applied contexts (Lipman, 1995, 2003). Nevertheless, the tendency to address these skills solely at the cognitive level has led to the neglect of the individual's affective dimensions. Yet emotions, values, and beliefs direct the individual's thinking processes; in this sense, the integration of affective characteristics into thinking skills is of considerable importance (Lipman, 2003; Sharp, 2007).

As a thinking-education program that equips children with diverse perspectives, Philosophy for/with Children (P4C/PwC) also guides children toward learning and thinking more effectively (Lam, 2021). This program helps children derive meaning from their lived experiences and provides them with tools to become aware of, and to reflect upon, both their own thoughts and those of others (Fisher, 2008). First implemented in the United States in the early 1970s by Matthew Lipman and Ann Margaret Sharp, P4C education is today widely used across many fields and countries (Boyacı, Karadağ, & Gülenç, 2018).

The fundamental point of departure for Lipman's (1976) approach lies in his observations, made in the late 1960s, regarding the inadequacy of students' critical thinking skills within the American education system. Having noticed that the students he taught at Columbia University possessed weak thinking and inquiry skills, Lipman argued that the causes of this deficiency should be sought in childhood (Lipman, 1976). In particular, major social and political developments of the period, such as the Vietnam War and the Civil Rights Movement, rendered the absence of critical thinking in education more visible. Within this context, Lipman maintained that, in order to remedy this deficiency, children should be equipped with philosophical thinking skills from an early age. At the core of this approach lies the aim of activating students' thinking processes through stimulating texts, images, and videos; of encouraging their inquiry through guiding questions; and of fostering, from the impetus of curiosity, the asking of questions while simultaneously developing critical, creative, caring, and collaborative thinking skills from a philosophical standpoint (Lipman, 1998). Tracing its origins back to the Socratic dialogues, this understanding was also substantially influenced by John Dewey's experience-based views of education (Taşdelen, 2014).

One of the fundamental components of the approach developed by Lipman, "caring thinking," has been articulated as an affective dimension of education (Leong, 2005). Conceiving of thinking as a form of action that is always situated within particular social and ethical-political contexts, Lipman (2003) regards caring thinking as an indispensable complement to, and a unifying element of, critical and creative thinking (Theodoridis, 2026). According to Lipman

(1995), caring becomes a mode of thinking when it carries out such cognitive operations as scanning for alternatives, discovering or inventing relationships, forging new connections among connections, and appraising differences. Sharp (2007), for her part, defined caring thinking as a combination of emotional and cognitive processes. Lipman (1995) initially held that the concept of caring comprised four dimensions—valuational thinking, affective thinking, active thinking, and normative thinking—and later increased the number of dimensions to five by also incorporating empathic thinking (Lipman, 2003).

The first of these dimensions, appreciative thinking, rests on the individual's relation to the object under evaluation, to the idea under consideration, or to the situation deemed worthy of appreciation, and it manifests as a sensitivity to the quality of objects, ideas, and relationships. Appreciative thinking is connected to the ethical principles that mediate the individual's capacity to enter evaluative judgments (Uluçınar & Arı, 2019).

The second dimension, affective thinking, is a mode of thinking that does not separate emotion from judgment but rather operates through the reciprocal relation between the two. Here, thinking is not isolated as an individual process; on the contrary, it is shaped within relationships and takes the form of a reflectively articulated empathy that makes understanding possible without denying the critical dimension of thought. Nussbaum (2004) clearly demonstrated that emotions are a form of thought and are generally bound up with the individual's beliefs.

The third dimension, active thinking, possesses a structure that translates thinking into action. According to Lipman (2003), this mode of thinking involves a passionate focus on, and active engagement with, a cause or value. In this dimension, rather than succumbing to feelings of helplessness or being overwhelmed in the face of the situations encountered, the individual learns to concentrate on what can be done about the situation in question and, over time, develops a tendency to improve existing conditions (Lipman, 2003).

Normative thinking, in turn, stands out as a kind of sensitive thinking skill applicable in local or global contexts. In this dimension, students concern themselves with universal laws and principles and with global issues such as protecting the environment, animals, and human rights; empathy, moral imagination, and the avoidance of egocentric thought develop markedly in this process. Being sensitive to inconsistencies between ideals and existing states of affairs, conceptualizing the problems society faces, generating solutions, and grappling with profound human and global challenges constitute the foundations of normative thinking (Sharp, 2007).

The final dimension, empathic thinking, was defined by Lipman (2003) as placing oneself in another person's position and feeling that person's emotions as if they were one's own. Lipman (2003) emphasizes that an important function of empathic thinking is the ability to step back from one's own emotions, perspectives, and views in order to imagine the emotions, perspectives, and views of others; for this reason, he regards the use of imagination as an inseparable part of empathic thinking (Ji, 2025).

A review of the literature reveals that, as a holistic pedagogical model supporting students' cognitive and affective development, the P4C approach has in recent years become the subject of numerous studies in Türkiye as well. Studies exist that examine its effects on students' social skills (Kaya, 2020; Okur, 2008), thinking skills (Akkocaoğlu Çayır, 2015; Altunsöz, 2022; Boyraz, 2019; Kabil, 2021; Özcan, 2023; Türksoy, 2020) and creativity skills (Akan, 2022; Gürdal, 2022; Pekkarakaş, 2020; Taş, 2017).

In addition, the literature includes theoretical research addressing the conceptual dimension of P4C (Alabaz, 2022; Budak Durmuş & Çalışkan, 2022; Erdoğan, 2018; Sormaz Öğüt, 2019), as well as studies examining its effects on problem-solving skills (Işıklar, 2019), critical thinking skills (Daşdelen, 2021; Korla Öztürk, 2022; Turhan, 2021), reasoning patterns and use of evidence (Karagöz, 2022) and religious thinking (Su, 2022). Furthermore, studies focusing on teachers' views (Çakmak, 2021; Davan, 2022; Koyuncu, 2020) have evaluated the reflections of these practices in the field. As can be seen, a substantial portion of these studies has focused on the more cognitive processes of P4C, such as critical, creative, and collaborative thinking. However, although caring thinking is defined as a pedagogical and affective dimension within the 4C (Critical, Creative, Caring, Collaborative thinking) framework emphasized in Lipman's (2003) model, it is striking that research focusing on this skill has remained limited in Türkiye.

The present study aims to fill this gap. By revealing the effect of P4C activities developed for middle school students on students' caring thinking skills, it seeks to provide findings that will guide both practitioners and curriculum-development efforts. Accordingly, the study endeavors to demonstrate the potential of P4C practices to develop components of caring thinking such as empathy, respect, and ethical sensitivity. It is anticipated that the findings obtained may raise awareness regarding the consideration of caring thinking skills in the teaching-learning process and may also provide a basis for work on character education, values education, and the development of affective skills. In this respect, it is hoped that the study may make a modest contribution to the applicability of P4C at the middle school level in the Turkish context and may lay the groundwork for evaluations concerning alternative approaches to the development of thinking skills.

The general aim of this research is to examine, by means of a pretest-posttest control group quasi-experimental design, the effect of Philosophy for Children (P4C) activities on the caring thinking skills of sixth-grade middle school students. In line with the aim of the study, an answer was sought to the following question:

Is there a statistically significant difference between the posttest scores on caring thinking skills and its sub-dimensions (appreciative, affective, active, normative, and empathic thinking) of the experimental group, to which P4C activities were applied, and the control group, to which the official curriculum was applied, when the pretest scores are controlled?

2. Methodology

2.1. Research Model and Design

This research is a study grounded in the quantitative research paradigm. The research was conducted in accordance with the pretest-posttest control group quasi-experimental design, which is situated among the experimental designs. The quasi-experimental design is a design in which the researcher cannot determine the participants in the experimental and control groups randomly but instead incorporates pre-existing groups such as schools and classes into the research process within their natural structures (Cohen, Manion, & Morrison, 2007). The difficulty of keeping all variables under control in educational research renders the implementation of true experimental designs impossible in most cases; for this reason, quasi-experimental designs come to the fore as an alternative with high application validity in educational settings (Karasar, 2008). This design is frequently preferred in educational research where the random assignment of classes is not possible (Kerlinger, 1973). For the reasons in question, the existing

classes were designated as the experimental and control groups, and the research was designed in accordance with the quasi-experimental design.

The independent variable of the research is the Philosophy for Children (P4C) activities, and the dependent variable is the students' caring thinking skills. P4C-based activities were applied in the experimental group, while in the control group the lessons were sustained in the manner prescribed by the official curriculum. Before and after the implementation, the Caring Thinking Skills Inventory (CTSI) was administered to both groups, and pretest and posttest scores were obtained. The design of the research is shown symbolically in Table 1.

Table 1

The Experimental Design of the Research

Group	Pretest	Treatment	Posttest
Experimental (G1)	CTSI	P4C activities	CTSI
Control (G2)	CTSI	National curriculum	CTSI

2.2. Participants

The participants of the research consist of students enrolled in the sixth grade at a public school located in the central district of Burdur province during the 2024-2025 academic year. After the necessary permissions had been obtained from the relevant institutions and organizations, two sixth-grade classes in which the Science Applications course was taught by the same teacher were designated as the study group. The distribution of the study group by groups is presented in Table 2.

Table 2

Distribution of the Study Group by Groups

Group	<i>n</i>	%
Experimental	32	49.2
Control	33	50.8
Total	65	100.0

As shown in Table 2, there are 32 students in the experimental group and 33 students in the control group, for a total of 65 participants. Students were not randomly assigned to groups; instead, the two existing classes were used in their natural structure, with one class assigned to the experimental condition and the other to the control condition. The fact that the classes attended the same school and were taught by the same teacher contributed to minimizing the influence of possible confounding variables such as socio-economic structure, academic background, and the teacher variable.

The Science Applications course was selected for the implementation for several reasons. First, as an elective course with a flexible and activity-based content structure, it allowed the P4C sessions to be integrated without reducing the instructional time allocated to core courses. Second, the emphasis of the course on inquiry, discussion, and the ethical and social dimensions of scientific issues is compatible with the dialogic nature of the community of inquiry. Finally, the fact that the course was taught by the same teacher in both classes helped to control the teacher variable.

2.3. Data Collection Instrument

The Caring Thinking Skills Inventory (CTSI), developed by Uluçınar and Arı (2019), was used as the data collection instrument in the research. The inventory consists of 20 problem scenarios, each followed by 5 options, for a total of

100 items. The scenarios were structured so as to encompass the five dimensions of caring thinking - appreciative thinking, affective thinking, active thinking, normative thinking, and empathic thinking- with four scenarios created for each dimension. After each scenario, students mark one of five options by considering how they would behave if they were in the protagonist's position, thereby obtaining a score for each scenario ranging from 1 (uncaring thinking) to 5 (high-level caring thinking). Accordingly, the lowest score obtainable from the inventory is 20 and the highest is 100.

The inventory development process was carried out in accordance with a sequential exploratory mixed-methods design. In the first stage, written responses to the scenarios were obtained from 54 seventh-grade students randomly selected from a middle school of average socio-economic level. The items belonging to the scenarios were structured hierarchically in five stages on the basis of Krathwohl, Bloom, and Masia's (1964) taxonomy of the affective domain (receiving, responding, valuing, organizing, and characterizing): (1) uncaring thinking, (2) awareness/sense-making, (3) valuing, (4) justification/organizing, and (5) caring thinking/behavior. For the content validity of the inventory, the opinions of five experts from the fields of science education, Turkish-language education, and curriculum and instruction were consulted; the appropriateness and clarity of each scenario in relation to the intended outcomes were rated out of 5, and the means were found to range between 3.6 and 5.0. Within the scope of face validity, the 100 items beneath the scenarios were evaluated by three experts for their conformity to the taxonomic stages, and the necessary corrections were made in line with the experts' feedback. For construct validity, in a pilot application conducted with 184 seventh-grade students (103 girls, 81 boys), the differences between the lower 27% and upper 27% groups were tested with an independent-samples t-test, and all scenarios were determined to be discriminating ($p < .05$). The Cronbach's alpha internal-consistency coefficient of the inventory was reported as .673 (Uluçınar & Arı, 2019). In the present research as well, the reliability of the inventory was examined through internal consistency and scenario-total correlations in a manner consistent with the scale's development study. Accordingly, the Cronbach's alpha internal-consistency coefficient for the entire inventory was calculated as .74. In this structure, which is scenario-based and measures each dimension with four scenarios, the approach of Uluçınar and Arı (2019), who developed the scale, was followed, and corrected scenario-total correlations were evaluated instead of separate internal-consistency coefficients for the sub-dimensions. These correlations ranged from .03 to .54, with sixteen of the twenty scenarios lying above the value of .20. The values obtained display a pattern similar to the findings reported in the scale's development study (Cronbach's alpha = .673; scenario-total correlations .08–.38) and indicate that the inventory produced acceptably reliable measurements in the present study sample.

2.4. Experimental Procedure

The experimental implementation was carried out over eight weeks, for an hour (one class period) per week, within the scope of the Science Applications course during the 2024-2025 academic year. Before the implementation began, the Caring Thinking Skills Inventory was administered to both groups as a pretest. The pretest results were compared with an independent-samples t-test in order to determine whether the experimental and control groups were equivalent in terms of caring thinking skills prior to the implementation.

The Philosophy for Children (P4C) activities conducted in the experimental group were planned by the researcher on the basis of Lipman's (2003) critical-creative-caring-collaborative thinking (4C) model and by devising inquiries that specifically targeted the five dimensions of caring thinking (appreciative, affective, active, normative, and empathic

thinking). The flow of the activities generally followed these steps: (i) the presentation of a stimulating text, image, or short video to activate students' sense of curiosity; (ii) students' generation of philosophical questions on the basis of the text and the determination of the inquiry question to be addressed collectively; (iii) the conduct of dialogue within a community of inquiry; (iv) the steering of the discussion toward the relevant dimension of caring thinking; and (v) the evaluation of the inquiry process and students' reflections on their own thinking processes. During the discussion process, rather than transmitting information directly, the teacher facilitated the dialogue through guiding questions and afforded students the opportunity to listen to one another's ideas, to justify them, and to complete one another's reasoning.

In the control group, by contrast, the Science Applications course was sustained by the same teacher in the manner prescribed by the official curriculum, without any additional intervention. After the completion of the eight-week process, the Caring Thinking Skills Inventory was administered to both groups as a posttest. Throughout the data-collection process, students' participation on a voluntary basis was ensured, and it was clearly stated to the participants that the data would remain confidential and would be used solely for research purposes.

2.5. Experimental Treatment Material

In this research, the “Philosophy for Children (P4C) Activities Program,” developed by the researchers for application to the experimental-group students, was used as the experimental treatment material. The program aims to systematically develop the caring thinking skills defined in Lipman's (2003) multidimensional thinking model, and its five sub-dimensions (appreciative, affective, active, normative, and empathic thinking), in sixth-grade middle school students. In the development of the program, the community-of-inquiry approach of Splitter and Sharp (1995) was taken as the basis, and the pedagogical flow was shaped in accordance with the classical community-of-inquiry structure proposed by Cam (2006).

The program was designed to be implemented over eight weeks, in one session per week. Each session was planned as one class hour. The applications were arranged so as to be integrable into the flexible content structure of the Science Applications course. The 11-12 age range selected as the target group is regarded as a developmental window in which, according to Piaget (Inhelder & Piaget, 1958), the transition to the formal operational stage begins and, according to Kohlberg's (1984) theory of moral development, the transition to the conventional level occurs; for this reason, it was accepted as a suitable period in which the capacities for reasoning about values and norms, taking another's perspective, and reading emotions as judgments could be pedagogically supported.

The weekly sequence of the program was structured not randomly but on developmental and psycho-pedagogical grounds. The first week was devoted to establishing the community of inquiry and to building in-class trust; the second week began with the appreciative thinking dimension, which is relatively concrete and carries a lighter emotional load. The third and fourth weeks successively comprise the affective thinking and empathic thinking dimensions, on the assumption that an individual who recognizes his or her own emotion can more readily understand the emotion of another. In the fifth and sixth weeks, normative thinking and active thinking—which encompass higher-order processes of evaluation and translation into action—were addressed. The seventh week was structured to achieve integration by means of a multilayered dilemma requiring the joint deployment of all five dimensions, while the eighth week was devoted to a session of reflective synthesis and closure.

All sessions other than the first, seventh, and eighth weeks follow a standard seven-stage structure: the opening circle, the sharing of the stimulus, silent thinking, question generation, question selection, the inquiry dialogue, and evaluation-closure. In this structure, the teacher assumes not the role of “the bearer of the correct answer” but that of a facilitator who supports the flow of the inquiry and safeguards participation. Among the facilitator's core responsibilities are asking open-ended questions, making the conceptual connections among students visible, requesting justifications, upholding the norm of respectful listening, and, when necessary, proposing conceptual distinctions.

For each week of the program, developmentally appropriate literary and audiovisual stimuli were used. In the second week, the “The Fox and the Rose” section from Saint-Exupéry's *The Little Prince* (1943/2015); in the third week, the opening scene of the Pixar production *Inside Out*; in the fourth week, the day on which Auggie first enters the classroom from the film *Wonder* by R. J. Palacio; in the fifth week, a version of Kohlberg's (1984) classic Heinz dilemma adapted for middle school students; in the sixth week, a news report about middle school students who launched a campaign to renovate a village school in Türkiye; and in the seventh week, the “The Class's Decision” scenario, developed by the researcher and calling the class directly to a decision, were employed as stimuli. In the eighth week, instead of an independent stimulus, the student question sheets retained from the first week were used as a reflective tool, with the aim of having students witness their own intellectual change.

During the development process of the program, expert opinion was consulted prior to the implementation; two field experts from the field of curriculum and instruction examined the program in terms of its aims, weekly flow, the developmental appropriateness of the stimuli, the consistency of the activities with the dimensions of caring thinking, and the evaluation processes. In line with the feedback received from the experts, the necessary adjustments concerning the time allocation of the activities, the wording of the inquiry questions, and the selection of stimuli were made, and the activity plans were finalized. During the implementation process, the researcher assumed the role of facilitator and took care, in each session, to adhere to the principles of the community of inquiry set forth by Lipman (2003).

Sample Activity: “The Emotion Map” (Week 3: Affective Thinking)

In the third-week session, in which the affective thinking dimension was targeted, the aim was for students to recognize that emotions are not merely momentary reactions but also meaningful indicators that direct thinking and evaluation processes. Accordingly, the session was structured around the first four-minute opening scene of the Pixar production *Inside Out*. In this scene, it is visualized how, from the birth of the character named Riley, the emotions of Joy, Sadness, Fear, Anger, and Disgust come into play by means of a “control panel,” and how each emotion possesses a particular function in the individual's experience. The activity can also be implemented using the Turkish-dubbed version of the film. After the viewing, a philosophical dialogue was conducted with the students in a community-of-inquiry setting on the function of emotions and their relationship to thinking processes.

Table 3

Activity Plan for the Third-Week Session (“The Emotion Map”)

Week / Theme	Week 3: “What Do Our Emotions Tell Us?”
Target Dimension	Affective thinking

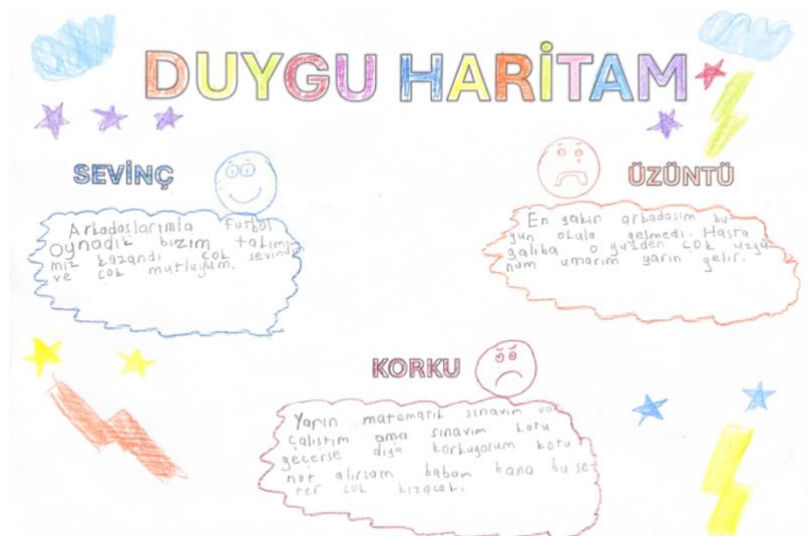
Week / Theme	Week 3: “What Do Our Emotions Tell Us?”
Intended Outcomes	Recognizes that emotions carry judgment and contribute to thinking; discusses how “negative” emotions may also be functional; establishes the link between the emotional reaction given to a situation and the evaluation of that situation.
Stimulus	The first four-minute opening scene of the Pixar production Inside Out.
Materials	Projector or interactive whiteboard; an A4 “Emotion Map” worksheet for each student; colored pencils.
Sample Inquiry Questions	What purpose does fear serve; what would happen if we did not feel fear? Is it correct to say “Sadness is a bad emotion”; what is sadness good for?
Thinking-Triggering Mini-Activity	Students are asked to write, on the Emotion Map, three different emotions they experienced in the past week and what these emotions “told” them. Students share their maps with the class.
Expected Learning Outcome	Students are expected to grasp emotions not merely as momentary reactions but as experiences that carry meaning, involve evaluation, and direct thinking processes.

As presented in Table 3, the intended outcomes of this session were determined as students' (a) recognizing that emotions carry judgment and contribute to thinking, (b) discussing how emotions characterized as “negative” may also be functional, and (c) establishing the link between the emotional reaction given to a situation and the evaluation of that situation. The activity employs a projector or interactive whiteboard, an A4 worksheet titled “Emotion Map” prepared for each student, and colored pencils. After viewing the stimulus, discussion is conducted in the community of inquiry through questions such as the following: What purpose does fear serve; what would happen if we never felt fear? Is it correct to say “Sadness is a bad emotion”; what is sadness good for? Has an emotion ever misled you; how? Are emotion and thought different things, or can emotion too be treated as a form of thinking? Is the statement “I ought to do the right thing even when I do not feel like it” meaningful to you? Through these questions, students are supported in treating emotions not merely as inner states that are experienced but also as intellectual processes that involve judgment and evaluation.

The session also includes a brief thinking-triggering activity. In this activity, students are asked to write, on the Emotion Map, three different emotions they experienced in the past week and what these emotions “told” them. Volunteer students share the maps they have prepared with the class, and through these shared accounts the class discussion concerning how emotions shape the way the individual makes sense of events is deepened. This mini-activity aims to help students grasp emotion not as a reaction entirely independent of judgment but as an experience that carries meaning and involves evaluation about a situation. In this way, the affective thinking dimension is supported through both individual awareness and the philosophical inquiry conducted within the community. An example of a student's work is presented in Figure 1.

Figure 1

An Example of Student Work on “The Emotion Map” Activity



2.6. Data Analysis

The SPSS 26.0 software package was used in the analysis of the data. Before the analysis, the data were examined for outliers and missing values, and no erroneous data entry was encountered. In all statistical analyses, the level of significance was adopted as $\alpha = .05$.

In order to test the equivalence of the experimental and control groups prior to the implementation, an independent-samples t-test was applied to the pretest scores. After the implementation, the significance of the differences in the posttest scores was tested by one-way analysis of covariance (ANCOVA), taking the groups' pretest scores as the covariate. Because ANCOVA enables the statistical control of pretest scores, it is the analysis method preferred in pretest-posttest control group quasi-experimental designs.

The four fundamental assumptions required for conducting ANCOVA were examined in turn: (i) the normal distribution of the dependent variable at the group level, by means of kurtosis and skewness coefficients remaining within the ± 1 limit as well as the Shapiro-Wilk test; (ii) the homogeneity of group variances, by means of Levene's test; (iii) the linearity of the relationship between the covariate (pretest) and the dependent variable (posttest), by means of the Pearson correlation coefficient and the scatterplot; and (iv) the homogeneity of regression slopes, by means of an ANCOVA model including the Group $\times$ Pretest interaction. When this assumption was violated, the model including the Group $\times$ Pretest interaction was retained instead of a standard ANCOVA. In this case, the Group main effect was not interpreted as a single adjusted mean difference; instead, group differences were examined at different pretest levels (the mean and ± 1 SD) through simple-effects analyses.

In addition, as a supplementary sensitivity analysis, the gain scores (posttest – pretest) of the groups were compared with the Mann-Whitney U test. This analysis was not used to correct or validate violations of ANCOVA assumptions; because it is based on gain scores, it does not remove a Group $\times$ Pretest interaction and was therefore reported only as complementary evidence regarding the direction of the group difference. Effect sizes were calculated with partial eta squared (η^2p) for ANCOVA and with the r effect-size value for the Mann-Whitney U test, and the results were interpreted according to the criteria of Cohen (1988) and Field (2018). Cohen (1988) defines a partial eta squared

value of .01 as small, .06 as medium, and .14 and above as a large effect. Field (2018), for his part, evaluates an r effect-size value of .10 as small, .30 as medium, and .50 and above as a large effect.

In line with the research question, in addition to the total score, the effect of the P4C activities on each of the five sub-dimensions of caring thinking (appreciative, affective, active, normative, and empathic thinking) was also examined separately. In order to minimize the risk of inflating the Type I error rate in the multiple comparisons conducted for the five sub-dimensions, the Bonferroni correction was applied and the level of significance for the sub-dimension analyses was set at $p < .01$. This correction was applied to the tests of group differences in the sub-dimensions; the tests of ANCOVA assumptions, including the homogeneity of regression slopes, were evaluated at the conventional $\alpha = .05$ level.

3. Results

The findings are presented in the following order: the descriptive statistics of the students' pretest scores and the results of the independent-samples t -test conducted for the equivalence of the groups; the one-way analysis of covariance (ANCOVA) of the posttest scores and the analyses testing its assumptions; and finally the findings concerning the five sub-dimensions of caring thinking (appreciative, affective, active, normative, and empathic thinking).

3.1. Findings Concerning the Pretest Scores of the Caring Thinking Skills Inventory

An independent-samples t -test was applied to analyze the groups' pretest results on caring thinking skills. Prior to this, the kurtosis and skewness coefficients were examined in order to investigate whether the groups exhibited a normal distribution. The descriptive statistics of the groups' pretest scores on caring thinking skills are given in Table 4.

Table 4

Descriptive Statistics for the Pretest Scores on Caring Thinking Skills

Group	n	M	SD	Min	Max	Skew.	Kurt.
Experimental	32	63.25	3.02	58	68	-0.27	-0.97
Control	33	62.91	3.12	57	69	0.42	-0.70

When Table 4 is examined, the skewness and kurtosis values are seen to lie between (-1) and (+1). The results of the independent-samples t -test conducted to compare the groups' pretest results on caring thinking skills are given in Table 5.

Table 5

Independent-Samples t -Test Results for the Pretest Scores on Caring Thinking Skills

Group	n	M	SD	t	p
Experimental	32	63.25	3.02	0.45	0.656
Control	33	62.91	3.12		

When the results are examined, the experimental group's pretest mean for caring thinking skills ($M = 63.25$; $SD = 3.02$) is seen to be quite close to the control group's pretest mean ($M = 62.91$; $SD = 3.12$). This difference between the two groups was not found to be statistically significant ($t(63) = 0.45$; $p = 0.656 > 0.05$). Accordingly, it can be said that prior to the implementation the groups were equivalent in terms of caring thinking skills.

3.2. Findings Concerning the Posttest Scores of the Caring Thinking Skills Inventory

A one-way ANCOVA was conducted to analyze the posttest scores when the groups' pretest scores on caring thinking skills were controlled. Prior to this, the assumptions required for conducting ANCOVA were verified. The descriptive statistics for the groups' pretest and posttest values on caring thinking skills are given in Table 6.

Table 6

Descriptive Statistics for the Groups' Pretest and Posttest Scores on Caring Thinking Skills

Group	Test	<i>n</i>	<i>M</i>	<i>SD</i>	Min	Max	Skew.	Kurt.
Experimental	Pretest	32	63.25	3.02	58	68	-0.27	-0.97
	Posttest	32	76.44	3.39	70	84	0.29	-0.06
Control	Pretest	33	62.91	3.12	57	69	0.42	-0.70
	Posttest	33	67.42	3.80	60	76	0.18	-0.28

The skewness and kurtosis coefficients for the groups' pretest and posttest scores are seen to lie within the range of (−1) to (+1), which indicates that the distributions satisfy the assumption of normality (Büyüköztürk, 2018; Hair et al., 2019). Nevertheless, the assumption of normality was additionally tested with the Shapiro-Wilk test. According to the Shapiro-Wilk test results, the experimental group's posttest statistic was obtained as $W = 0.964$ ($p = 0.361 > 0.05$) and the control group's posttest statistic as $W = 0.978$ ($p = 0.716 > 0.05$); both groups thus satisfy the assumption of normal distribution.

To determine whether group variances were homogeneous—the other assumption—Levene's test results were examined, and no statistically significant difference was found ($F(1, 63) = 0.24$; $p = 0.623 > 0.05$). According to this result, the groups' variances for the posttest scores on caring thinking skills are homogeneous.

The results of the Pearson correlation analysis conducted to test whether there was a linear relationship between the dependent variable and the covariate are given in Table 7.

Table 7

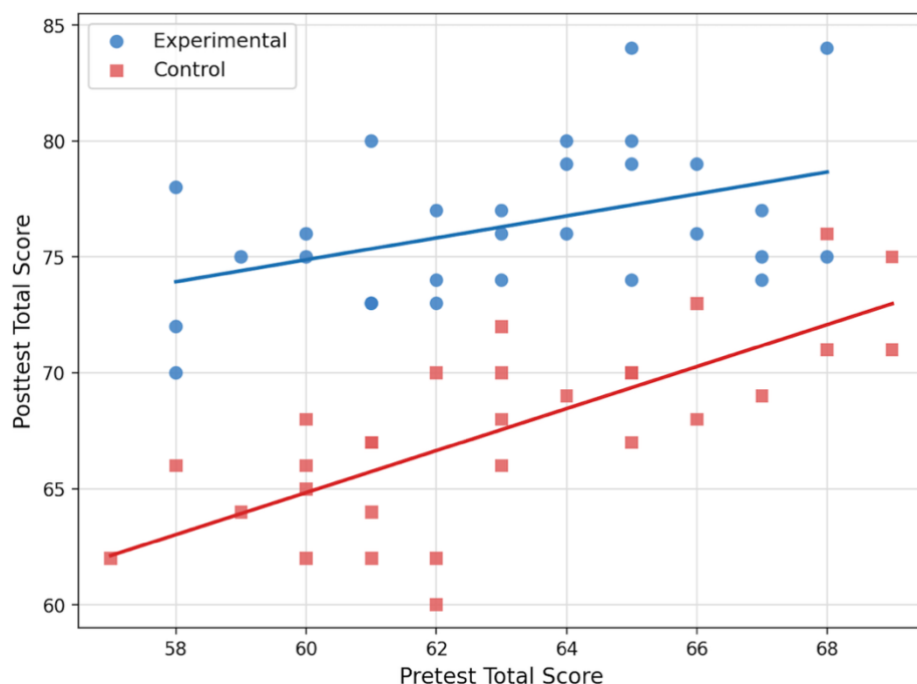
The Relationship Between the Covariate and the Dependent Variable for Caring Thinking Skills

Group	Test	<i>r</i>	<i>p</i>
Experimental	Pretest / Posttest	0.428	0.015
Control	Pretest / Posttest	0.750	<0.001

When Table 7 is examined, positive linear relationships are seen to exist between the pretest and posttest scores in both the experimental and the control groups. While this relationship is at a moderate level in the experimental group ($r = 0.428$; $p = 0.015 < 0.05$), it is significant at a high level in the control group ($r = 0.750$; $p < 0.001$). This result shows that the assumption of the existence of a linear relationship between the covariate (pretest) and the dependent variable (posttest) was satisfied in both groups. The scatterplot showing the relationship between the groups' pretest and posttest caring thinking skills scores is presented in Figure 2.

Figure 2

Scatterplot Showing the Relationship Between the Pretest and Posttest Scores on Caring Thinking Skills



When Figure 2 is examined, the point clusters of the experimental and control groups are seen to diverge markedly along the posttest axis. While the experimental group's posttest scores cluster largely in the range of 70-84, the control group's posttest scores are distributed in the range of 60-76. For both groups, the regression lines are observed to have a positive slope, which supports the linearity between the covariate and the dependent variable. Although the slope appears flatter in the experimental group, the homogeneity of regression slopes analysis reported below (Table 8) showed no statistically significant difference in slopes between the groups.

The results of the ANCOVA analysis conducted to test the assumption regarding the homogeneity of regression slopes are given in Table 8.

Table 8

Homogeneity of Regression Slopes for Caring Thinking Skills

Source of Variance	Sum of Squares	df	Mean Sq.	F	p
Group	1246.09	1	1246.09	154.26	<0.001
Pretest	297.29	1	297.29	36.80	<0.001
Group × Pretest	27.89	1	27.89	3.45	0.068
Error	492.76	61	8.08		
Total	2064.03	64			

When Table 8 is examined, the Group × Pretest interaction effect on students' posttest scores on caring thinking skills is seen not to be significant ($F(1, 61) = 3.45$; $p = 0.068 > 0.05$). According to this result, the assumption of homogeneity of regression slopes required for ANCOVA was satisfied. Considering that the assumptions of normal distribution, homogeneity of variance, and linearity reported in the previous steps were also satisfied, it can be said that all the fundamental assumptions of the ANCOVA analysis were met. In addition, the Mann-Whitney U test was applied to the gain scores as a supplementary sensitivity analysis.

The results of the ANCOVA analysis conducted to determine whether the difference between the groups' adjusted posttest mean scores on the caring thinking skills test is significant, together with results of the supplementary Mann-Whitney U test applied to the gain scores, are given jointly in Table 9.

Table 9

ANCOVA and Mann-Whitney U Test Results for the Caring Thinking Skills Test

Source of Variance	Sum of Sq.	df	Mean Sq.	F / U	p	$\eta^2 p / r$
ANCOVA						
Pretest (covariate)	297.29	1	297.29	35.40	<0.001	
Group	1246.09	1	1246.09	148.39	<0.001	0.705
Error	520.65	62	8.40			
Mann-Whitney U						
Group				U = 18.0	<0.001	r = 0.83

According to Table 9, a statistically significant difference is found, on the basis of the ANCOVA results, between the experimental and control groups' adjusted posttest mean scores on caring thinking skills, adjusted for their pretest scores ($F(1, 62) = 148.39$; $p < 0.001$). This result shows that the P4C activities significantly increased the caring thinking skills of the experimental-group students. The group effect was large (partial $\eta^2 = .705$) according to Cohen's (1988) criteria.

According to the Mann-Whitney U test results applied to the gain scores (posttest – pretest) on the same data, a statistically significant difference favoring the experimental group was likewise found between the groups ($U = 18.0$; $Z = 6.67$; $p < 0.001$). The effect size was calculated as $r = 0.83$, which is a large effect according to Field's (2018) criteria. The direction of this result was consistent with the ANCOVA findings.

The groups' adjusted posttest mean scores on caring thinking skills, adjusted for their pretest scores, following the experimental treatment, are given in Table 10.

Table 10

Mean Scores on the Caring Thinking Skills Test

Group	n	Pre M	Pre SD	Post M	Post SD	Adj. Post M	SE
Experimental	32	63.25	3.02	76.44	3.39	76.31	0.51
Control	33	62.91	3.12	67.42	3.80	67.54	0.51

When Table 10 is examined, the means of the scores students obtained on the caring thinking skills test following the experimental treatment are seen to be 76.44 for the experimental-group students and 67.42 for the control-group students. The mean of the posttest scores adjusted for the pretest scores on caring thinking skills was calculated as 76.31 in the experimental group and 67.54 in the control group. Accordingly, the experimental group's adjusted posttest mean is 8.77 points higher than the control group's adjusted posttest mean.

3.3. Findings Concerning the Sub-Dimensions of Caring Thinking Skills

In line with the research question, the effect of the P4C activities on each of the five sub-dimensions of caring thinking (appreciative, affective, active, normative, and empathic thinking) was also examined separately. The groups' pretest and posttest descriptive statistics for these five sub-dimensions are given in Table 11.

Table 11*Pretest and Posttest Descriptive Statistics for the Sub-Dimensions*

Dimension	Group	<i>n</i>	<i>Pre M</i>	<i>Pre SD</i>	<i>Post M</i>	<i>Post SD</i>
Appreciative	Experimental	32	13.19	1.42	15.41	1.07
	Control	33	13.24	1.56	13.70	1.53
Affective	Experimental	32	12.44	1.56	15.22	1.45
	Control	33	12.45	1.52	13.64	1.78
Active	Experimental	32	12.97	1.03	15.22	1.26
	Control	33	12.82	1.07	13.82	1.07
Normative	Experimental	32	12.53	1.44	15.38	1.48
	Control	33	12.09	1.40	13.06	1.32
Empathic	Experimental	32	12.12	1.04	15.22	1.64
	Control	33	12.30	1.33	13.21	1.29

When Table 11 is examined, the experimental and control groups' pretest means are seen to be quite close to one another in all five sub-dimensions, whereas marked differences favoring the experimental group emerged among the posttest means. Pretest equivalence for the five sub-dimensions was tested with the independent-samples t-test, and no significant difference was found between the groups in any sub-dimension (all $p > 0.21$; all $|\text{Cohen's } d| \leq 0.31$). The ANCOVA assumptions were tested separately for each sub-dimension. Accordingly, the assumption of homogeneity of regression slopes was violated in the appreciative thinking sub-dimension ($p < 0.001$), in the affective thinking sub-dimension ($p = 0.014$), and in the active thinking sub-dimension ($p = 0.012$), while it was satisfied in the normative and empathic thinking sub-dimensions ($p = 0.072$ and $p = 0.121$, respectively). According to Levene's test results, variance homogeneity was satisfied in all sub-dimensions (all $p > 0.10$). Therefore, for the appreciative, affective, and active thinking sub-dimensions, a model including Group, Pretest, and the Group $\times$ Pretest interaction was estimated instead of a standard ANCOVA (Table 12). For the normative and empathic thinking sub-dimensions, a standard ANCOVA was conducted.

Table 12*Group $\times$ Pretest Models for the Appreciative, Affective, and Active Thinking Sub-Dimensions*

Dimension	Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	η^2p
Appreciative	Group	21.58	1	21.58	22.81	< .001	.272
	Pretest	32.40	1	32.40	34.25	< .001	.360
	Group $\times$ Pretest	15.19	1	15.19	16.06	< .001	.208
	Error	57.70	61	0.95			
Affective	Group	17.12	1	17.12	9.66	.003	.137
	Pretest	47.98	1	47.98	27.06	< .001	.307
	Group $\times$ Pretest	11.46	1	11.46	6.46	.014	.096
	Error	108.18	61	1.77			
Active	Group	10.52	1	10.52	8.91	.004	.127
	Pretest	5.59	1	5.59	4.73	.033	.072
	Group $\times$ Pretest	7.93	1	7.93	6.71	.012	.099
	Error	72.06	61	1.18			

Note. Pretest scores were not centered; the Group effect therefore refers to a pretest score of 0, which lies outside the possible score range (4-20), and is not interpreted. Group differences at the mean and ± 1 SD are presented in Table 13. Type III sums of squares.

As shown in Table 12, the Group $\times$ Pretest interaction was significant for appreciative thinking ($F(1, 61) = 16.06, p < .001$, partial $\eta^2 = .208$), affective thinking ($F(1, 61) = 6.46, p = .014$, partial $\eta^2 = .096$), and active thinking ($F(1, 61) = 6.71, p = .012$, partial $\eta^2 = .099$), indicating that the relationship between pretest and posttest scores differed between the experimental and control groups. In all three dimensions, posttest scores were strongly related to pretest scores in the control group (unstandardized slopes = 0.81, 0.84, and 0.62, respectively), whereas this relationship was weak or absent in the experimental group (slopes = 0.15, 0.29, and -0.05). Accordingly, the Group main effect in these dimensions was not interpreted as a single adjusted mean difference; instead, group differences were examined at different pretest levels through simple-effects analyses (Table 13).

Table 13

Estimated Group Differences (Experimental - Control) at Different Pretest Levels

Dimension	Low pretest (M - 1 SD)	Mean pretest	High pretest (M + 1 SD)
Appreciative	2.72 ($p < .001$)	1.74 ($p < .001$)	0.76 ($p = .032$)
Affective	2.44 ($p < .001$)	1.59 ($p < .001$)	0.75 ($p = .117$)
Active	2.07 ($p < .001$)	1.36 ($p < .001$)	0.65 ($p = .094$)

Note. Pretest values corresponding to M - 1 SD and M + 1 SD were 11.73 and 14.70 for appreciative, 10.91 and 13.98 for affective, and 11.84 and 13.94 for active thinking.

According to Table 13, at the mean pretest level the experimental group scored significantly higher than the control group in all three dimensions (1.74, 1.59, and 1.36 points for appreciative, affective, and active thinking, respectively; all $p < .001$). The advantage of the experimental group was largest among students with low pretest scores (2.72, 2.44, and 2.07 points, respectively; all $p < .001$) and diminished among students with high pretest scores (0.76, 0.75, and 0.65 points; $p = .032, .117$, and $.094$, respectively), none of which reached the Bonferroni-adjusted threshold ($p < .01$). In other words, in these three dimensions the P4C activities made a clearer difference for students who started with lower levels of caring thinking.

Table 14

Standard ANCOVA Results (Normative and Empathic Thinking) and Supplementary Mann-Whitney U Results on Gain Scores

Dimension	<i>F</i>	<i>df</i>	<i>p</i> (ANCOVA)	η^2p	<i>U</i>	<i>Z</i>	<i>p</i> (MWU)	<i>r</i>
Appreciative	*	—	—	—	168.5	4.72	<0.001	0.59
Affective	*	—	—	—	246.5	3.69	<0.001	0.46
Active	*	—	—	—	291.0	3.11	0.001	0.39
Normative	41.44	1, 62	<0.001	0.401	245.5	3.71	<0.001	0.46
Empathic	37.58	1, 62	<0.001	0.377	171.0	4.68	<0.001	0.58

Note. * Standard ANCOVA results are not reported for these sub-dimensions because the homogeneity of regression slopes assumption was violated (see Tables 12 and 13).

According to Table 14, the standard ANCOVA conducted for the normative and empathic thinking sub-dimensions, in which the homogeneity of regression slopes assumption was met, revealed significant differences favoring the experimental group (normative: $F(1, 62) = 41.44, p < .001, \eta^2p = .401$; empathic: $F(1, 62) = 37.58, p < .001, \eta^2p = .377$). These effects are large according to Cohen's (1988) criteria and meet the Bonferroni-corrected significance threshold ($p < .01$). As a supplementary sensitivity analysis, the Mann-Whitney U test on gain scores also showed differences favoring the experimental group in all five sub-dimensions, with effect sizes ranging from $r = 0.39$ to 0.59 . However, because this analysis is based on gain scores, it does not remove or resolve the Group $\times$ Pretest interaction observed

in the appreciative, affective, and active thinking sub-dimensions; it is therefore regarded only as complementary evidence regarding the direction of the group difference.

The adjusted posttest means for the normative and empathic thinking sub-dimensions, for which the standard ANCOVA was appropriate, are given in Table 15.

Table 15

Adjusted Posttest Means for the Normative and Empathic Thinking Sub-Dimensions

Dimension	Experimental (Adj. M)	Control (Adj. M)	Difference
Normative	15.35	13.08	2.27
Empathic	15.26	13.17	2.09

When Table 15 is examined, the differences between the experimental and control groups are seen to be largely preserved even after the pretest differences have been controlled. The adjusted mean differences were calculated as 2.27 points for normative thinking and 2.09 points for empathic thinking. Both differences favor the experimental group.

In summary, while no significant difference emerged between the groups in pretest caring thinking skills scores before the experimental implementation, a significant difference favoring the experimental group emerged in the posttest scores obtained at the end of the experimental implementation. This difference was observed in the total score and in the normative and empathic thinking sub-dimensions. In the appreciative, affective, and active thinking sub-dimensions, the experimental group scored higher than the control group at average and low pretest levels, whereas the difference diminished and was not statistically significant among students with high pretest scores.

4. Discussion and Conclusion

The P4C activities produced a significant and large effect on caring thinking skills ($F(1, 62) = 148.39$; $p < .001$; $\eta^2 p = .705$). The adjusted mean difference was 8.77 points, and the supplementary Mann-Whitney U analysis on gain scores yielded a result in the same direction ($r = .83$). Beyond statistical significance, this difference may also be said to be noteworthy from an educational standpoint. Considering that the scores obtainable from the CTSI range from 20 to 100, an adjusted mean difference of 8.77 points corresponds to approximately 11% of the scale's range; this indicates that the experimental-group students achieved an observable improvement in their levels of caring thinking over the eight-week process. The large group effect observed in the total score (partial $\eta^2 = .705$) suggests that a structured P4C program, in which the stimuli and inquiry questions directly target the dimensions of caring thinking, may be a promising approach for values education in line with the Türkiye Century Education Model (MoNE, 2024), although this effect should be verified in larger samples. Nevertheless, it is considered that, in order to evaluate the behavioral correlate of this difference in the actual classroom environment, it needs to be supported with qualitative and longitudinal data.

Moreover, this effect size lies above the average values reported in similar studies in the P4C literature. In the systematic review by Trickey and Topping (2004), a general effect size of $d = .43$ was reported; in the meta-analysis by García-Moriyón, Rebollo, and Colom (2005), an effect of $d = .58$ for reasoning skills; and in the meta-analysis by Yan et al. (2018), an effect of $d = 1.06$ for reasoning. The recent three-level meta-analysis by Wei and Chen (2025) likewise found the general effect of P4C on cognitive development to be at a medium-to-large level ($g = .59$) and additionally

showed that studies with small samples and short-term interventions (≤ 20 weeks) produced comparatively higher effects. However, the meta-analyses in question are largely concerned with the cognitive outcomes of P4C (reasoning, critical thinking, creativity), whereas the present research addresses caring thinking, which is affective-ethical in nature. For this reason, a direct comparison is limited; moreover, the partial eta squared (η^2p) reported here and the standardized mean difference (d/g) in the aforementioned studies are different effect-size criteria and have no exact equivalence. The high effect in the present research may be explained by the small sample ($n = 65$) inflating the effect size, by the activities focusing directly on caring thinking, and by the novelty effect of the intervention.

When considered in the Turkish context, the findings, while consistent with the literature, are noteworthy in that they focus on a comparatively understudied dimension. Studies such as those by Akkocaoglu Çayır (2015), Boyraz (2019) and Türksöy (2020) have demonstrated the positive effects of P4C on thinking skills, while the systematic review by Lafcı-Tor (2023) revealed that the great majority of domestic research has remained limited to the cognitive dimension. By focusing directly on caring thinking, the present research is considered to contribute to reducing the empirical gap, in the Turkish context, concerning the caring dimension of Lipman's (2003) 4C model.

As another result of the research, the pattern of findings differed across the sub-dimensions. In the normative and empathic thinking dimensions, the group effect was consistent across pretest levels and large ($\eta^2p = .401$ and $.377$, respectively). In the appreciative, affective, and active thinking dimensions, however, the effect of the intervention depended on students' initial levels: while posttest scores in the control group largely mirrored pretest scores, students in the experimental group reached relatively similar posttest levels regardless of their starting point. As a result, the benefit of the P4C activities in these dimensions was most pronounced for students with lower initial levels and became smaller, and statistically non-significant, among students who already had high initial levels. This pattern may suggest that the community of inquiry provided a particularly supportive environment for students with less developed caring thinking; however, it may also partly reflect a ceiling effect, since each sub-dimension is measured with only four scenarios and scores cannot exceed 20.

In the appreciative thinking dimension, the group difference was 1.74 points at the average pretest level and reached 2.72 points among students with low pretest scores. According to Lipman (1995), this dimension is related to the process by which the individual ascribes value to objects, ideas, and people and thereby appreciates them; Sharp (2007), for her part, emphasized that appreciative thinking constitutes the foundation of sensitivity to aesthetic and moral principles. According to Splitter and Sharp (1995), this skill can develop in a dialogic environment within the classroom community of inquiry in which students can articulate, with their own justifications, what they find valuable and why.

Literary texts are regarded as favorable pedagogical tools for the development of the dimension in question. Cam (2006) noted that stories facilitate the discussion of abstract value concepts through concrete characters and situations, while Haynes and Murriss (2012) demonstrated that picture and literary texts enable children to approach the world of values and meaning with philosophical curiosity. It may be said that the inquiry conducted on the "The Fox and the Rose" section of *The Little Prince*, used as the second-week stimulus in the program, afforded students the opportunity to discuss the distinction between "ascribing value" and "possessing" through their own life

experiences. This finding is consistent with the theoretical frameworks according to which appreciative thinking can develop through concrete life experiences and literary characters.

In the affective thinking dimension, the experimental group scored higher at average and low pretest levels, but the difference was not significant among students with high pretest scores. According to Lipman (2003), this dimension operates through the reciprocal relation between emotion and judgment; Nussbaum (2004) argued that emotions are a kind of value judgment. The study conducted by Giménez-Dasí, Quintanilla, and Daniel (2013) in early childhood likewise shows that P4C produces positive effects on the comprehension of emotion. The fact that the effect was concentrated among students with lower initial levels may point to the difficulty of capturing changes in emotion education through short-term interventions, particularly among students who already display a high level of affective thinking.

The effect obtained in the active thinking dimension, which was evident among students with average and low pretest scores, is consistent with the pedagogical design shaped around the question “what do I do because I value it.” Sharp (2007) emphasized that this dimension involves translating values into action. In the program, the use of a concrete campaign example from Türkiye as a stimulus in the sixth week and the “My One Step” activity supported this outcome. The finding that the group difference diminished among students with high pretest scores, together with the comparatively low effect size in the supplementary Mann-Whitney U analysis ($r = .39$), suggests that the eight-week period may have remained limited for an outcome of a behavioral nature.

The high effect found in the normative thinking dimension ($\eta^2p = .401$) is consistent with Sharp's (2007) definition of this dimension as sensitivity to universal principles and the conceptualization of the inconsistency between the ideal and the actual. The study by García-Moriyón et al. (2020) likewise shows that P4C makes a noteworthy contribution to moral development. The use of an age-adapted version of Kohlberg's (1984) Heinz dilemma in the program prompted students to justify rules and to reflect on the distinction between “what is” and “what ought to be.”

The large effect obtained in the empathic thinking dimension ($\eta^2p = .377$) is consistent with Lipman's (2003) definition of empathy as imagining another's perspective and with Sharp's (2007) emphasis that empathy is nurtured within the classroom community of inquiry. Schleifer et al. (2003) found the increase in the empathy scores of the early-childhood P4C group to be significant relative to the control group, while Asgari et al. (2023) detected a similarly directed but not statistically significant effect at the middle school level. The attainment of a significant effect in the present research may be associated with the capacity of the scenario-based structure of the scale used (Uluçınar & Arı, 2019) to capture emotional processes more sensitively than self-report scales.

In general, the finding that the effect was consistent across pretest levels in normative thinking (adjusted mean difference = 2.27 points), whereas in active thinking it diminished among students with higher initial levels, points to a meaningful pattern: discussing and justifying values may be an outcome that can be internalized in a comparatively shorter time than translating those values into action. Because the homogeneity of regression slopes assumption was violated in the appreciative, affective, and active thinking sub-dimensions, the findings for these dimensions should be interpreted as conditional on students' initial levels rather than as a single average effect; the supplementary Mann-Whitney U results indicate the same direction but do not eliminate this conditionality.

4.1. Limitations of the Research

The findings of this research should be evaluated within the framework of certain limitations. First, the study was conducted at a single public school, in two sixth-grade classes taught by a single teacher, and with 65 students. The relatively small sample and its restriction to a single context may account for a portion of the high effect size obtained; indeed, it is known that in small samples effect-size values may be overestimated. For this reason, caution should be exercised in generalizing the findings to different schools, grade levels, and socio-economic contexts. Moreover, because only one experimental and one control class were involved and assignment was made at the class level, the effect of the intervention could not be fully disentangled from class-specific characteristics such as classroom climate and peer dynamics. In addition, the analyses were conducted at the student level, which should be taken into account when interpreting the results.

Second, the fact that the researcher personally assumed the role of facilitator who planned and implemented the P4C activities makes it difficult to entirely exclude the effect of experimenter expectancy (researcher bias). In addition, the sustaining of only the official curriculum in the control group means that the effects of the additional attention, interaction, and intervention novelty afforded to the experimental group (the Hawthorne effect) may have partially confounded the observed difference. The fact that an active comparison activity (active control) involving equivalent time and attention was not provided to the control group is, in this respect, a limitation. Finally, the data were collected only with a scenario-based self-report inventory and immediately after the intervention; qualitative measurements such as observation and interview, as well as follow-up measurements that would test the permanence of the gains, remained outside the scope. These limitations also constitute the basis of the recommendations presented in the following section.

5. Recommendations

5.1. Recommendations for Researchers

Conducting experimental studies with larger samples and based on random assignment will make it possible to test whether the findings are also valid in different contexts. In particular, replicating the P4C program and instructional materials developed in this study in different schools and with larger samples may show whether the large effect observed in the total score can be generalized. Longitudinal studies designed to include follow-up measurements at different time intervals after the implementation may provide information regarding the permanence of the gains obtained. Examining caring thinking with qualitative tools such as in-class observation, focus-group interviews, and student journals, in addition to quantitative measurements, may yield a more holistic picture of the skill's reflections in the actual classroom environment. In the appreciative, affective, and active thinking sub-dimensions, the effect of the intervention varied according to students' initial levels; longer-term applications and designs that examine this conditional effect may contribute to understanding the factors affecting development in these dimensions. Finally, comparative studies conducted at different educational levels (primary school and high school), together with investigation of the relationship of caring thinking to variables such as academic achievement, positive social behaviors, and classroom climate, may contribute to the literature. Moreover, examining the role of caring thinking skills in reducing peer bullying and the violent behaviors encountered in the school environment may both enrich the literature and guide preventive-intervention efforts in schools.

5.2. Recommendations for Practitioners

It is considered that integrating the “caring” dimension of the 4C model into courses such as Turkish, Social Studies, and Religious Culture and Ethics may provide a meaningful basis for values education, character education, and social-emotional learning practices. This approach also overlaps with the fundamental conceptual framework of the Türkiye Century Education Model, which was put into practice in 2024. In the model in question, the “virtue-value-action” cycle, integrated skills, and social-emotional learning domain skills are placed at the center of the curricula (Ministry of National Education [MoNE], 2024). Caring thinking, by virtue of its structure that integrates emotion with judgment and value with action, appears directly aligned with the pedagogical goals of this framework. The study by Lam (2021) reveals that P4C can also bring about transformative effects on teachers' professional identity. Accordingly, it may be recommended that professional-development programs for teachers be designed in order to disseminate P4C practices and to integrate them with the social-emotional learning goals of the Education Model.

In the present research, it was found that devoting the one-class-hour-per-week portion of the Science Applications course to P4C activities provided a significant improvement in the caring thinking skills of the experimental-group students. The study by Topping and Trickey (2007) shows that a one-hour-per-week P4C application can be associated with lasting gains. This indicates that P4C applications can be carried out within existing class hours without imposing an additional burden on the curricula. Accordingly, it may be recommended that instructional materials be developed for integrating P4C activities of a nature supportive of the Education Model's “virtue-value-action” cycle into different courses, and that program-design efforts toward P4C-based social-emotional learning be carried out in cooperation with school counselors.

Declaration of Conflicts of Interest

The authors declare no conflict of interest.

Declaration of Generative AI Use

During the preparation of this manuscript, Grammarly (Grammarly Inc., desktop application, version 1.201.0.0) and Claude (Claude Opus 5.5, Anthropic) were used as needed throughout the writing process, solely to improve the grammar, spelling, punctuation, and fluency of text written by the authors. The tools were used to obtain language-editing suggestions on the authors' own text; each suggestion was reviewed individually, and only changes that improved readability without altering the original content were accepted. The conceptual framework, methodology, data analysis, interpretation of findings, selection of sources, and conclusions were developed entirely by the authors. The authors take full responsibility for the accuracy, integrity, and originality of the manuscript.

Ethical Statement

Ethical approval for this study was obtained from the Burdur Mehmet Akif Ersoy University Non-Interventional Clinical Research Ethics Committee (meeting dated 16.04.2025, No. 2025/3; decision No. GO 2025/1445). The principles of scientific research and publication ethics were observed at all stages of the research process. Participation in the data collection process was voluntary, and participants were clearly informed that their data would be kept confidential and used solely for research purposes. The necessary implementation permissions were obtained from the relevant institutions and organizations.

Author Contributions

Esra Çakar Özkan (corresponding author) made the principal contribution to developing the theoretical framework of the study, determining the research problem and design, developing the experimental intervention materials, collecting and organizing the data, analyzing and interpreting the data, and managing the writing of the manuscript, as well as its critical revision and finalization. Aysel Ceren Özdemir contributed to conducting the literature review, implementing the activities, and preparing the first draft of the manuscript. Both authors read and approved the final version of the manuscript submitted for publication.

References

- Akan, R. (2022). *Hayat bilgisi öğretiminde çocuklar için felsefe (P4C) yaratıcı düşünme becerisi etkinliklerinin öğrencilerin yaratıcı düşünme becerisine etkisi* [Unpublished doctoral dissertation]. Bursa Uludağ University.
- Akkocaoğlu Çayır, N. (2015). *Çocuklar için felsefe eğitimi üzerine nitel bir araştırma* [Unpublished doctoral dissertation]. Hacettepe University.
- Alabaz, B. (2022). *Çocuklar için felsefe eğitimi üzerine* [Unpublished master's thesis]. Muğla Sıtkı Koçman University.
- Altunsöz, F. (2022). *Çocuklar için felsefe programının uygulanmasının önemi* [Unpublished master's thesis]. Pamukkale University.
- Asgari, M., Whitehead, J., Schonert-Reichl, K. A., & Weber, B. (2023). The impact of Philosophy for Children (P4C) on middle school students' empathy, perspective-taking, and autonomy: Preliminary outcomes. *Analytic Teaching and Philosophical Praxis*, 43(1), 26–43.
- Bloom, B. S. (1956). *Taxonomy of educational objectives: The classification of educational goals: Handbook I, cognitive domain*. Longmans.
- Boyacı, N. P., Karadağ, F., & Gülenç, K. (2018). Çocuklar için felsefe / Çocuklarla felsefe: Felsefi metotlar, uygulamalar ve amaçlar. *Kaygı. Bursa Uludağ Üniversitesi Fen-Edebiyat Fakültesi Felsefe Dergisi*, 31, 145-173. <https://doi.org/10.20981/kaygi.474657>
- Boyras, C. (2019). *İlkokulda çocuklarla felsefe uygulamaları: Bir eylem araştırması* [Unpublished doctoral dissertation]. Anadolu University.
- Budak Durmuş, F., & Çalışkan, M. (2022). Çocuklarla felsefe konusunda Türkiye'de yapılmış akademik çalışmaların içerik analizi. *Felsefe Dünyası*, 2(76), 105-134. <https://doi.org/10.58634/felsefedunyasi.1218929>
- Büyüköztürk, Ş. (2018). *Sosyal bilimler için veri analizi el kitabı* (24th ed.). Pegem Akademi.
- Cam, P. (2006). *20 thinking tools: Collaborative inquiry for the classroom*. ACER Press.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education* (6th ed.). Routledge.
- Çakmak, R. Y. (2021). *Okul öncesi öğretmenlerinin çocuklarla felsefe eğitim programına ilişkin görüşleri* [Unpublished master's thesis]. İstanbul Aydın University.
- Daşdelen, K. (2021). *Dördüncü sınıf öğrencilerinde çocuklar için felsefe eğitime yönelik eleştirel düşünme becerilerinin değerlendirilmesi* [Unpublished master's thesis]. Hacettepe University.
- Davan, Z. (2022). *Okul öncesi öğretmenlerinin çocuklarla felsefe hakkındaki görüşlerinin incelenmesi* [Unpublished master's thesis]. Sivas Cumhuriyet University.
- Erdoğan, P. (2018). *Çocuklarla felsefe yaklaşımının düşünsel, tarihi ve sosyal temelleri üzerine bir inceleme* [Unpublished master's thesis]. Ankara University.
- Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). SAGE Publications.
- Fisher, R. (2008). *Teaching thinking: Philosophical enquiry in the classroom* (3rd ed.). Bloomsbury Publishing.

- García-Moriyón, F., González-Lamas, J., Botella, J., González Vela, J., Miranda-Alonso, T., Palacios, A., & Robles-Loro, R. (2020). Research in moral education: The contribution of P4C to the moral growth of students. *Education Sciences*, 10(4), 119. <https://doi.org/10.3390/educsci10040119>
- García-Moriyón, F., Rebollo, I., & Colom, R. (2005). Evaluating Philosophy for Children: A meta-analysis. *Thinking: The Journal of Philosophy for Children*, 17(4), 14–22. <https://doi.org/10.5840/thinking20051743>
- Giménez-Dasí, M., Quintanilla, L., & Daniel, M.-F. (2013). Improving emotion comprehension and social skills in early childhood through philosophy for children. *Childhood & Philosophy*, 9(17), 63–89.
- Gürdal, G. (2022). *Çocuklar için felsefe eğitiminde hipotetik düşünme yönteminin ilkökul BİLSEM öğrencilerinin yaratıcı düşünme becerilerine etkisi* [Unpublished master's thesis]. Muğla Sıtkı Koçman University.
- Hair, J. F., Babin, B. J., Anderson, R. E., & Black, W. C. (2019). *Multivariate data analysis* (8th ed.). Cengage Learning.
- Haynes, J., & Murris, K. (2012). *Picturebooks, pedagogy and philosophy*. Routledge.
- Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking: From childhood to adolescence* (A. Parsons & S. Milgram, Trans.). Basic Books. <https://doi.org/10.1037/10034-000>
- Işıklar, S. (2019). *Çocuklar için felsefe eğitim programının 5-6 yaş çocuklarda felsefi sorgulama yoluyla eleştirel düşünme ve problem çözme becerilerine etkisi* [Unpublished master's thesis]. Çanakkale Onsekiz Mart University.
- Ji, H. I. (2025). Exploring the application of empirical research in philosophy for children: Analysis of two case studies in caring thinking. *Childhood & Philosophy*, 21, 1–42. <https://doi.org/10.12957/childphilo.2025.88956>
- Kabil, G. (2021). *Çocuklarla felsefe eğitici eğitimi alan ve almayan okul öncesi öğretmenlerinin düşünme becerileri eğitime yönelik pedagojik alan bilgilerinin karşılaştırılması* [Unpublished master's thesis]. İstanbul Aydın University.
- Karagöz, S. (2022). *Çocuklarla felsefe eğitiminin 60-72 aylık çocukların akıl yürütme örüntüleri ve kanıt kullanımları üzerindeki etkisinin incelenmesi* [Unpublished master's thesis]. İstanbul Aydın University.
- Karasar, N. (2008). *Bilimsel araştırma yöntemi* (18th ed.). Nobel Yayınevi.
- Kaya, N. B. (2020). *Hayat bilgisi ve sosyal bilgiler derslerinde çocuklar için felsefe: Bir eylem araştırması* [Unpublished doctoral dissertation]. Marmara University.
- Kerlinger, F. N. (1973). *Foundations of behavioral research*. (2nd Ed.) Holt, Rinehart and Winston.
- Kohlberg, L. (1984). *The psychology of moral development: The nature and validity of moral stages*. Harper & Row.
- Korla Öztürk, A. (2022). *Çocuklar için felsefe programının 10. sınıf öğrencilerinde eleştirel düşünme becerisine etkisi* [Unpublished master's thesis]. Düzce University.
- Koyuncu, E. (2020). *Investigating the views of preschool teachers regarding philosophy with children through PwC experience* [Unpublished master's thesis]. Middle East Technical University.
- Krathwohl, D. R., Bloom, B. S., & Masia, B. B. (1964). *Taxonomy of educational objectives: The classification of educational goals. Handbook II: Affective domain*. David McKay.
- Lafcı-Tor, D. (2023). Türkiye’de çocuklar için felsefe (P4C) uygulamalarına yönelik araştırmaların incelenmesi: Sistematiik derleme. *Kaygı. Bursa Uludağ Üniversitesi Fen-Edebiyat Fakültesi Felsefe Dergisi*, 22(3), 703–740. <https://doi.org/10.20981/kaygi.1347085>
- Lam, C.-M. (2021). The impact of philosophy for children on teachers’ professional development. *Teachers and Teaching: Theory and Practice*, 27(7), 642–655. <https://doi.org/10.1080/13540602.2021.1986693>
- Leong, T. C. (2005). *What is caring thinking? Teaching caring thinking: Vision, clarifications, justifications and applications*. Raffles Girls’ School (Secondary). <http://conference.nie.edu.sg/paper/Converted%20Pdf/ab00553.pdf>
- Lipman, M. (1976). Philosophy for children. *Metaphilosophy*, 7(1), 17–33. <https://doi.org/10.1111/j.1467-9973.1976.tb00616.x>
- Lipman, M. (1995). Caring as thinking. *Inquiry: Critical Thinking Across the Disciplines*, 15(1), 1–13. <https://doi.org/10.5840/inquiryctnews199515128>

- Lipman, M. (1998). Teaching students to think reasonably: Some findings of the Philosophy for Children program. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 71(5), 277–280. <https://doi.org/10.1080/00098659809602723>
- Lipman, M. (2003). *Thinking in education* (2nd ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9780511840272>
- Ministry of National Education (2024). *Türkiye Yüzyılı Maarif Modeli: Ortak metin*. MEB Yayınları.
- Nussbaum, M. (2004). Emotions as judgments of value and importance. In R. Solomon (Ed.), *Thinking about feelings: Contemporary philosophers on emotions* (pp. 183–199). Oxford University Press.
- Okur, M. (2008). *Çocuklar için felsefe eğitim programının altı yaş grubu çocuklarının atılabilirlik, işbirliği ve kendini kontrol sosyal becerileri üzerindeki etkisi* [Unpublished master's thesis]. Marmara University.
- Özcan, S. (2023). Özel yeteneklilere Türkçe öğretiminde çocuklar için felsefe (P4C) uygulamaları. *Kaygı. Bursa Uludağ Üniversitesi Fen-Edebiyat Fakültesi Felsefe Dergisi*, 22(3), 423–456. <https://doi.org/10.20981/kaygi.1318661>
- Pekkarakaş, E. (2020). *Okul öncesi eğitim döneminde çocuklar için felsefe eğitiminin öğrencilerin yaratıcılık düzeylerine etkisi* [Unpublished master's thesis]. İzmir Demokrasi University.
- Saint-Exupéry, A. de. (2015). *Küçük Prens* (T. Uyar, Trans.). Can Yayınları. (Original work published 1943)
- Schleifer, M., Daniel, M.-F., Peyronnet, E., & Lecomte, S. (2003). The impact of philosophical discussions on moral autonomy, judgment, empathy and the recognition of emotion in five year olds. *Thinking: The Journal of Philosophy for Children*, 16(4), 4–12. <https://doi.org/10.5840/thinking200316410>
- Sharp, A. M. (2007). Education of the emotions in the classroom community of inquiry. *Gifted Education International*, 22(2–3), 248–257. <https://doi.org/10.1177/026142940702200315>
- Sormaz Ögüt, F. (2019). *Felsefi düşünmenin önemi ve çocuklar için felsefe* [Unpublished doctoral dissertation]. Maltepe University.
- Splitter, L. J., & Sharp, A. M. (1995). *Teaching for better thinking: The classroom community of inquiry*. ACER.
- Su, M. (2022). *Çocuklar için felsefe (P4C) eğitiminin dini düşünce gelişimine etkisi üzerine bir araştırma* [Unpublished doctoral dissertation]. İnönü University.
- Taş, I. (2017). *Çocuklar için felsefe eğitimi programının 48-72 aylık çocukların zihin kuramı ve yaratıcılıklarına etkisi* [Unpublished doctoral dissertation]. Çukurova University.
- Taşdelen, V. (2014). Felsefenin gülümseyen yüzü: Çocuklarla felsefe. *Türk Dili*, 756, 562–568.
- Theodoridis, A. (2026). The cultivation of caring thinking in education for the re-creation of the democratic human being. *Philosophy Study*, 16(1), 1–17. <https://doi.org/10.17265/2159-5313/2026.01.001>
- Topping, K. J., & Trickey, S. (2007). Collaborative philosophical enquiry for school children: Cognitive effects at 10–12 years. *British Journal of Educational Psychology*, 77(2), 271–288. <https://doi.org/10.1348/000709906X105328>
- Trickey, S., & Topping, K. J. (2004). 'Philosophy for children': A systematic review. *Research Papers in Education*, 19(3), 365–380. <https://doi.org/10.1080/0267152042000248016>
- Turhan, A. (2021). *Çocuklar için felsefe eğitimi programı 3. sınıf uygulaması: Bir eylem araştırması* [Unpublished master's thesis]. Marmara University.
- Türksoy, N. (2020). *Çocuklar için felsefe (P4C) eğitiminin ortaokul öğrencilerinin bilimsel sorgulamaya yönelik görüşlerine ve eleştirel düşünme becerilerinin gelişimine katkısı: Bir karma yöntem araştırması* [Unpublished master's thesis]. Alanya Alaaddin Keykubat University.
- Uluçınar, U., & Arı, A. (2019). The development of caring thinking skills inventory based on problem scenarios: A study of validation and reliability. *Universal Journal of Educational Research*, 7(6), 1414–1429. <https://doi.org/10.13189/ujer.2019.070609>
- Wei, C., & Chen, L. (2025). The effects of Philosophy for Children on children's cognitive development: A three-level meta-analysis. *Journal of Intelligence*, 13(10), 130. <https://doi.org/10.3390/jintelligence13100130>
- Yan, S., Walters, L. M., Wang, Z., & Wang, C. C. (2018). Meta-analysis of the effectiveness of Philosophy for Children programs on students' cognitive outcomes. *Analytic Teaching and Philosophical Praxis*, 39(1), 13–33.