Enhancing the Quality of Thinking:
Evidence from a School-Based Cognitive Expansion Program

Sule Alan
Cornell University

Merve Derebasoglu
University of Göttingen

NBER Summer Institute 2026
Children and Families

About Higher-Order Thinking Skills

HOTS are are advanced cognitive abilities: sophisticated reasoning unique to humans

- Nonverbal HOTS (perceptual reasoning)
  - Fluid intelligence (learning speed, abstract reasoning)
  - Visual analogical reasoning
  - Pattern recognition and mental rotation
- Verbal HOTS (reading comprehension and verbal reasoning)
  - Critical thinking (evidence evaluation, logical inference, detecting logical inconsistencies, argument justification)
  - Cognitive empathy (Theory of Mind)

Why care about higher-order thinking skills (HOTS)?

While core cognition (executive functions), such as attention and working memory, are gateway to learning (knowledge accumulation),

- HOTS enable one to use knowledge to connect ideas, reason, solve problems, and create new ideas (innovate).
- HOTS, including socio-cognitive skills are essential for rational human activities

These skills are fundamental for learning further, academic achievement, career development, and social success.

How Higher-Order Thinking Skills (HOTS) Develop

- Naturally:
  - HOTS gradually emerge as children's brain mature
  - Development is shaped by executive function growth, especially working memory and attention
  - These skills rely on neural networks that continue to develop well into adolescence, with especially rapid growth in childhood due to high neural plasticity.
- Schooling:
  - Schools provide structured opportunities for reasoning and learning
  - Exposure to rich language, guided problem-solving, collaborative learning
  - Quality of instructional practices and the **cognitive demands placed on children** directly influence how quickly these skills develop.

SES Gradient

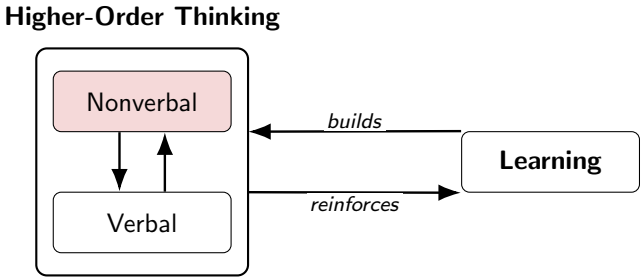
- SES differences shape early environments that support or limit HOTS growth (Johnson et al., 2016; Farah, 2017)
- Lower-SES settings often provide:
  - fewer cognitively stimulating interactions,
  - less complex language exposure,
  - more stress and instability,
  - fewer opportunities to practice complex reasoning or self-regulation
- These disparities translate into slower trajectories of HOTS development, persisting after school entry and often extending into adolescence and adulthood (Last et al., 2018)

What we do

Can placing high cognitive demand on socioeconomically disadvantaged primary school children accelerate the development of higher-order thinking skills?

- Helped design a neuroscience-informed intervention to facilitate the faster development of HOTS
- A two-year intensive program implemented in Turkish primary schools
- Cognitively demanding (both for children and teachers)
- Implemented in grade 2 and grade 3 in two cities in Turkey
- As part of daily teaching so crowding out possible
  - Grade 2 has very low official curricular load
  - Grade 3 and 4 high official curricular load

The Idea and the Theory of Change



- A subset of nonverbal skills are targeted through activities with minimal reading and writing
- While crowding out some other activities and content in the curriculum, aimed at speeding up learning and overall skills development, including FLN
- Development of verbal skills might be affected differently in L1 and L2 environments

Program Content

Developing Brain

- **Year 1**
  - Categorical Reasoning
  - Pattern Recognition
- **Year 2**
  - Analogical Reasoning
  - Logical reasoning & Pattern-matching

Teacher Workshops

- Two rounds of training at the beginning of the academic year: Grade 2 and Grade 3
- Introduced the student materials and classroom implementation plan
- Hands-on workshop format: teachers completed the activities themselves

Classroom Applications

- Delivered by classroom teachers during school hours
- Three to four class hours per week for two academic years
- Implemented using project-provided program materials

Evaluation Design: Sample

- Two southeastern cities in Türkiye:
  - Gaziantep (pop. 2.2M), predominantly Turkish-speaking
  - Şırnak (pop. 550k), bilingual; Kurdish widely spoken
  - Primary school is four years
- Instruction language: Turkish in all schools (no exceptions)

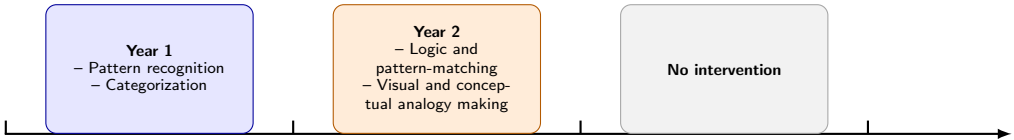
Figure 1: Location of study provinces in Türkiye

Evaluation Design: Sample

- **Sample size**
  - 132 public primary schools in disadvantaged neighborhoods
  - 567 classrooms, >14,000 students
- **Geographical coverage**
  - 84 schools in Gaziantep
  - 48 schools in Şırnak
- **Treatment assignment**
  - Randomisation stratified by district (9 districts)
  - 67 schools assigned to the program
  - 43 in Gaziantep, 24 in Şırnak

Appendix: Baseline Balance

Timeline of the Program



Baseline

Beginning of Grade 2

September 2023

- Fluid intelligence
- Analytical categorization
- Cognitive empathy
- Literacy and numeracy

Midline

End of Grade 2

May 2024

- Fluid intelligence
- Analytical categorization
- Cognitive empathy
- Literacy and numeracy
- Analogical reasoning
- Working memory
- Attentional control

Endline

End of Grade 3

May 2025

- Fluid intelligence
- Analytical categorization
- Cognitive empathy
- Literacy and numeracy
- Analogical reasoning
- Working memory
- Attentional control
- Critical thinking task

Follow-up

End of Grade 4

May 2026

- Fluid intelligence
- Analytical categorization
- Cognitive empathy
- Literacy and numeracy
- Analogical reasoning
- Working memory
- Attentional control
- Critical thinking task
- Inconsistency Detection

Evaluation of the Program

We evaluate the effect of the program on four outcome domains:

Nonverbal HOTS

- Abstract reasoning
- Analogical reasoning
- Categorical reasoning

Executive Functions

- Working memory
- Attentional control

Verbal HOTS

- Cognitive empathy
- Critical thinking
  - Evidence evaluation and inference
  - Logical inconsistency detection

Academic Performance

- Literacy
- Numeracy

Nonverbal HOTS: Abstract Reasoning

Raven's Progressive Matrices Test (Raven and Court, 1998):

- Nonverbal test of abstract reasoning
- Score is the number of correct responses

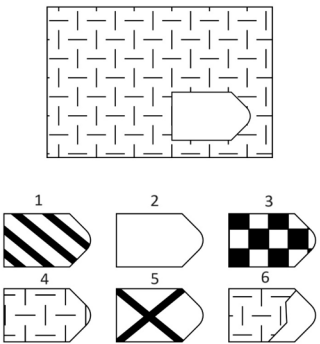


Figure 2: Sample screen in the abstract reasoning test

Nonverbal HOTS: Analytical Categorisation

Triad Task (Kagan et al., 1963):

- Students are presented with images of three objects
- They are asked to group two of them without giving any context
- Responses are classified as analytical or holistic categorization
- Score is the number of analytical categorizations

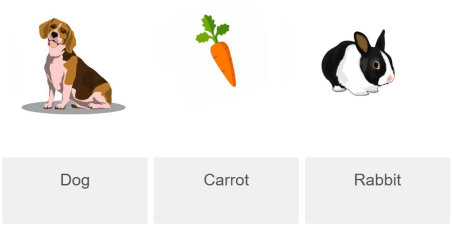


Figure 3: Sample screen in the analytical categorization task

Nonverbal HOTS: Analogical Reasoning

A custom-designed visual analogy test:

- Each item has the form A : B :: C : ? and students choose the option that completes the analogy
- The relations include part-whole, function similarities, sequence (order), and are adapted to children's everyday concepts
- Score is the number of correct responses

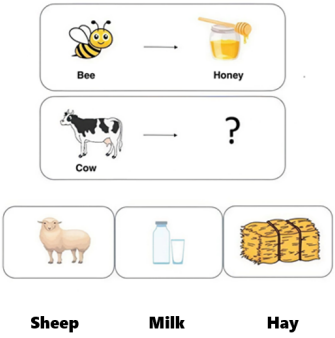


Figure 4: Sample screen in the analogical reasoning task

Executive Functions: Working Memory

A custom-design working memory task:

- Sequences of colored shapes for 10 seconds
- Questions about the location, quantity, or color of the shapes
- Score is the number of correct responses

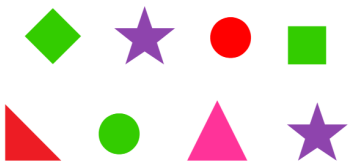
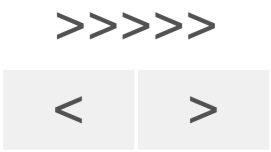


Figure 5: Sample screen in the working memory task

Executive Functions: Attentional Control

Flanker Task (Eriksen and Eriksen, 1974):

- Equal number of congruent and incongruent trials, presented in random order for 2 seconds.
- Attentional control measured as the difference in correct responses between congruent and incongruent trials.



(A.) Congruent trial

(B.) Incongruent trial

Figure 6: Sample screens in the flanker task

Verbal HOTS: Cognitive Empathy

Reading the Mind in the Eyes Test (Baron-Cohen et al., 1997):

- Students are presented with photos showing only the eye region
- They choose the word that best describes the person's emotion
- Captures the ability to infer other people's feelings from subtle facial clues
- Score is the number of correct responses

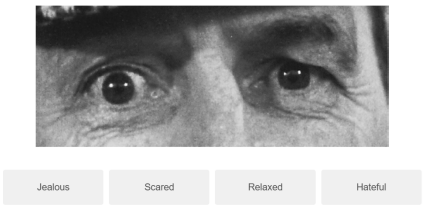


Figure 7: Sample screen in the cognitive empathy test

Verbal HOTS: Critical Thinking

- The ability to analyze information and evaluate evidence to reach reasoned judgments rather than accepting claims at face value
  - Distinguishing relevant from irrelevant information
  - Evaluating the credibility and consistency of evidence
  - Drawing logical inferences from available information
  - Explaining and justifying one's reasoning
- Requires reading comprehension and verbal reasoning
- *No established instruments exist to measure critical thinking in young children*

Critical Thinking Task

We developed a novel task with two components that capture complementary aspects of critical thinking:

- **Incentivized Mystery Solving**
  - Solve a mystery by interpreting sequential narrative clues
  - Eliminate inconsistent hypotheses after each clue
  - Justify each elimination

    → Evidence evaluation, logical inference, argument justification
- **Inconsistency Detection**
  - Read a narrative (30+ sentences) with numbered sentences
  - Identify a subtle logical inconsistency between two sentences

    → Integrating multiple premises, detecting logical inconsistencies

Appendix: Critical Thinking Task Details

Predictive Validity: Critical Thinking Index

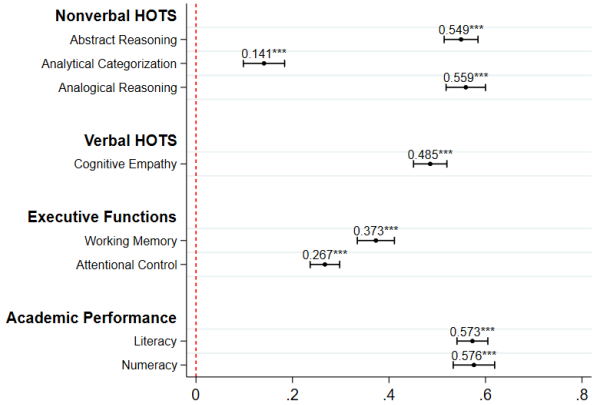


Figure 8: Correlation of the critical thinking index with cognitive skills and academic performance in the control group

Appendix: Predictive Validity

Main Results

Effect on Cognitive Skills and Academic Performance

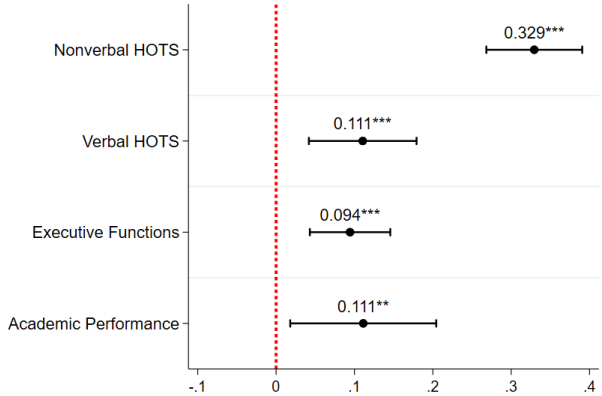


Figure 9: Treatment effect on aggregated cognitive skills and academic performance

Effect on Cognitive Skills and Academic Performance

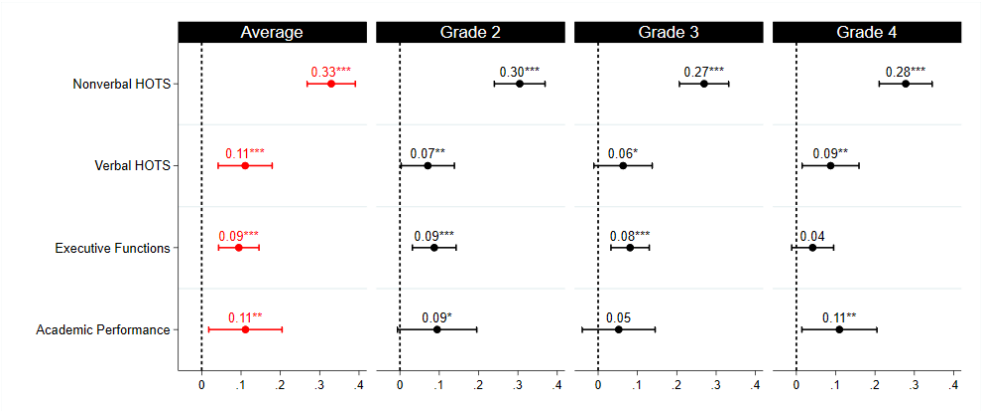


Figure 10: Treatment effect on aggregated cognitive skills and academic performance in each grade

Effect on Non-language-based Outcomes

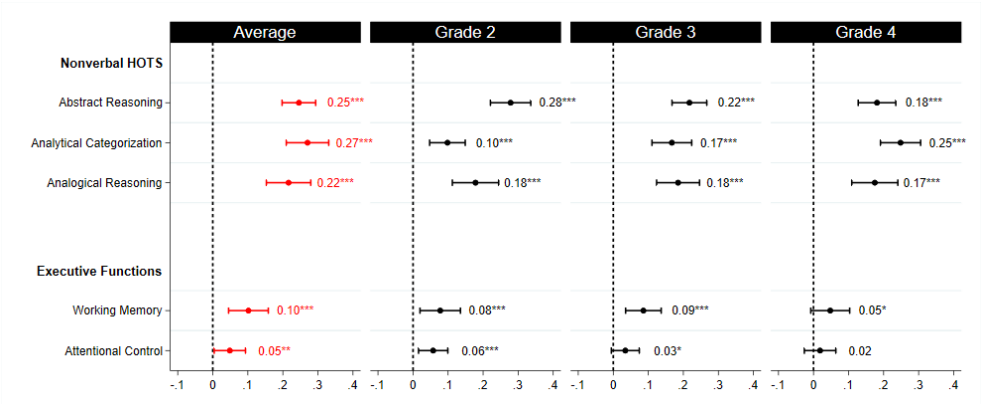


Figure 11: Treatment effect on nonverbal HOTS and executive functions in each grade

Effect on Language-based Outcomes

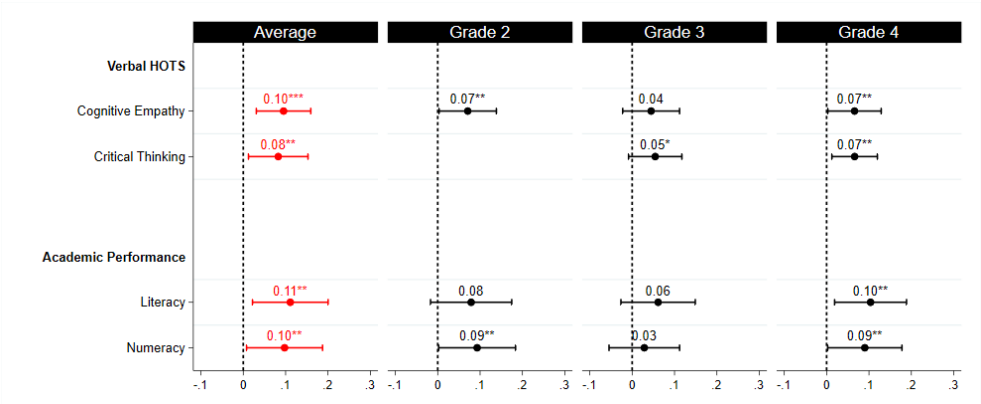


Figure 12: Treatment effect on verbal HOTS and academic performance

Heterogeneity Analysis: Non-language-based Outcomes

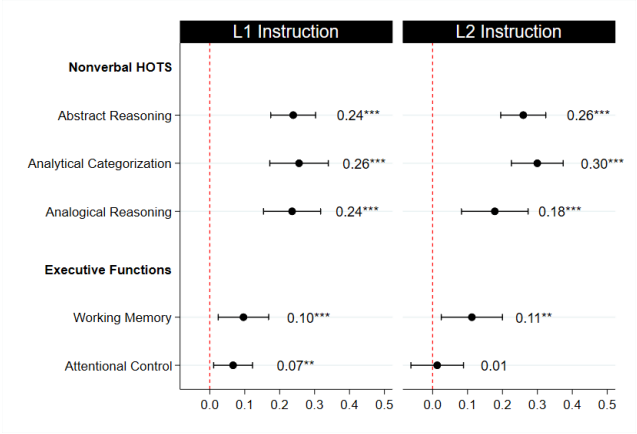


Figure 13: Heterogeneous effects on non-language-based skills by home-school language match.

The significance patterns for the L1–L2 difference test are shown at the far right of the figure.

Appendix: Heterogeneity Analysis

Heterogeneity Analysis: Language-based Outcomes

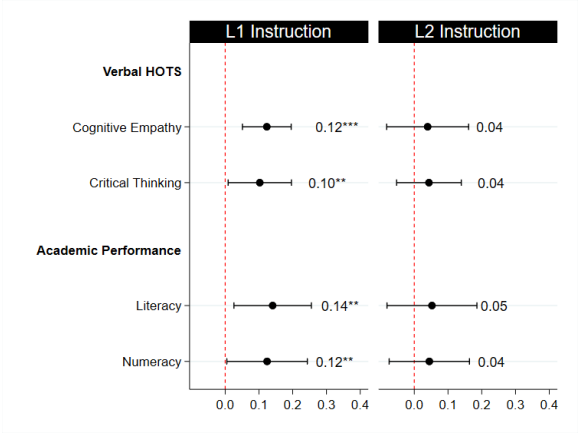


Figure 14: Heterogeneous effects on language based skills by home-school language match.

The significance patterns for the L1–L2 difference test are shown at the far right of the figure.

Why did this work?

The program was effective

- in improving core cognition and non-verbal HOTS for all children
- in improving verbal HOTS only in the L1 context

- Childhood is a period of rapid cognitive development. Placing sustained cognitive demands through challenging exercises appears to be an effective strategy.
- Consistent with evidence from cognitive neuroscience, cognitively demanding activities can expand cognitive capacity.
- Targeting a subset of cognitive skills can generate improvements in other domains (attention, abstract reasoning, cognitive empathy, critical thinking, verbal reasoning, and academic performance were not directly targeted).
- Despite substantial crowding out (especially in Grade 3), academic performance improved without increasing teacher workload.

Why did this program work? Evidence from Teachers

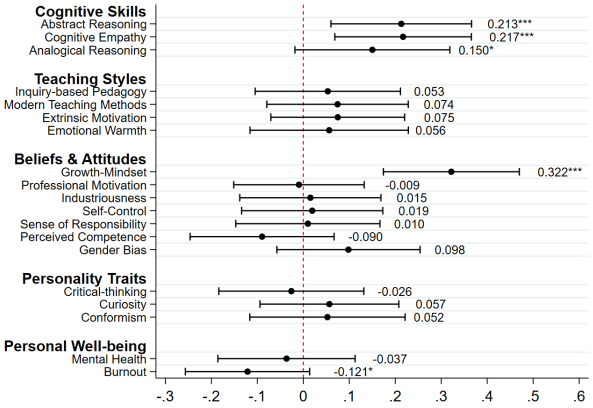


Figure 15: Treatment effect on teaching styles and teachers' cognitive outcomes

Conclusion

- Learning is the central objective of education, yet learning poverty remains widespread. (Mundial et al., 2022)
- Existing approaches focus on improving learning through remedial education, tutoring, SEL, or new pedagogical approaches (e.g., game-based and inquiry-based learning).
- We take a different approach, grounded in cognitive neuroscience. We test whether systematically pushing the limits of children's cognitive capacity can improve learning, thinking, and reasoning (cost: approximately $15 per child).
- The results are encouraging. They suggest that expanding children's cognitive capacity may be a promising new pathway to improving learning and higher-order thinking.
- We are also developing a dynamic human capital model in which higher-order thinking skills and learning reinforce each other over time.

Theory of Change

References

Baron-Cohen, S., Jolliffe, T., Mortimore, C., and Robertson, M. (1997). Another advanced test of theory of mind: Evidence from very high functioning adults with autism or asperger syndrome. Journal of Child psychology and Psychiatry, 38(7):813–822.

Eriksen, B. A. and Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. Perception & psychophysics, 16(1):143–149.

Farah, M. J. (2017). The neuroscience of socioeconomic status: correlates, causes, and consequences. Neuron, 96(1):56–71.

Johnson, S. B., Riis, J. L., and Noble, K. G. (2016). State of the art review: poverty and the developing brain. Pediatrics, 137(4).

Kagan, J., Moss, H., and Siegel, I. (1963). The psychological significance of styles of categorization. Basic cognitive processes in children. Monographs of the Society for Research in Child Development, 28(2).

Last, B. S., Lawson, G. M., Breiner, K., Steinberg, L., and Farah, M. J. (2018). Childhood socioeconomic status and executive function in childhood and beyond. PloS one, 13(8):e0202964.

Mundial, B. et al. (2022). The state of global learning poverty: 2022 update. World Bank Group. https://www. worldbank. org/en/topic/education/publication/state-ofglobal-learning-poverty.

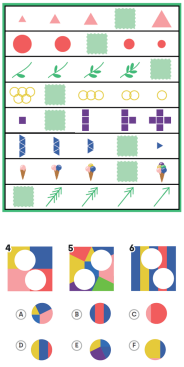
Raven, J. C. and Court, J. H. (1998). Raven's progressive matrices and vocabulary scales. Oxford Psychologists Press.

Appendix

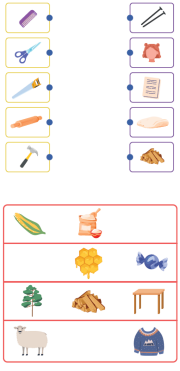
Program Content

Categorical Reasoning

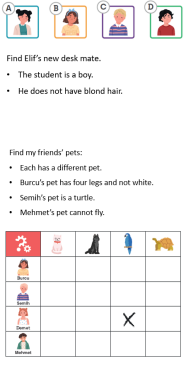
Pattern Recognition



Analogical Reasoning



Logical reasoning & Pattern-matching



Baseline Balance: Student Outcomes

	Count	Control Mean	Treatment Mean	P-value(Difference)
Student demographics				
Age in months	14803	88.20	88.11	0.36
Female (1,0)	14806	0.49	0.50	0.21
Refugee (1,0)	14806	0.08	0.07	0.48
Student cognitive skills				
Fluid IQ score	14806	0.00	0.08	0.21
Analytical categorization score	14806	0.00	-0.03	0.33
Cognitive empathy score	14665	0.00	0.05	0.33
Student academic outcomes				
Numeracy score	14478	-0.00	0.08	0.22
Literacy score	14038	-0.00	0.10	0.22

Back

Baseline Balance: Teacher and Class Characteristics

	Count	Control Mean	Treatment Mean	P-value (Difference)
Class Characteristics				
Class Size	567	34.54	34.33	0.95
Class Refugee Share	567	0.11	0.11	0.73
Teacher Demographics				
Age	567	34.03	35.01	0.25
Male (1,0)	567	0.34	0.34	0.95
Married (1,0)	565	0.56	0.61	0.27
Teacher Cognitive Skills				
Fluid IQ	566	-0.00	0.02	0.95
Cognitive Empathy	564	0.00	-0.07	0.46
Teaching Styles				
Inquiry-based Pedagogy	562	-0.00	-0.01	0.94
Modern Teaching Methods	561	-0.00	0.05	0.59
Extrinsic Motivation	559	-0.00	0.04	0.46
Emotional Warmth	558	-0.00	-0.02	0.89

Back

A Critical Thinking Task

A novel incentivized mystery-solving task:

- Solve a case (e.g., finding a lost item) by interpreting a sequence of short narrative clues and selecting the correct solution from several options.
- Designed to capture
  - Desire to find out the truth (willingness to pay)
  - ability to evaluate evidence/inference
  - ability to explain own reasoning

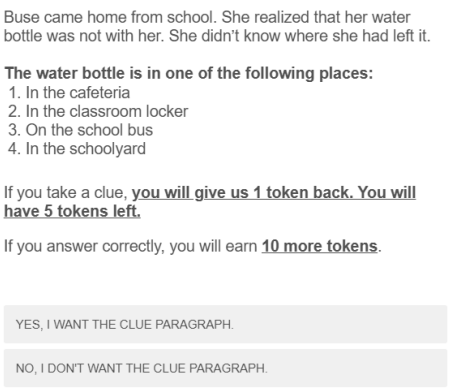


Figure A.1: Desire to find out the truth (willingness to pay to come close to the truth)

Back: Critical Thinking Task

A Critical Thinking Task

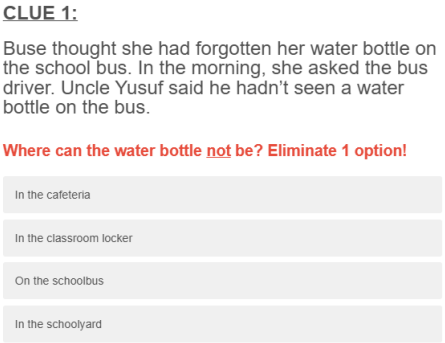


Figure A.2: Ability to evaluate evidence/inference

A novel incentivized mystery-solving task:

- Solve a case (e.g., finding a lost item) by interpreting a sequence of short narrative clues and selecting the correct solution from several options.
- Designed to capture
  - Desire to find out the truth (willingness to pay)
  - ability to evaluate evidence/inference
  - ability to explain own reasoning

Back: Critical Thinking Task

A Critical Thinking Task

A novel incentivized mystery-solving task:

- Solve a case (e.g., finding a lost item) by interpreting a sequence of short narrative clues and selecting the correct solution from several options.
- Designed to capture
  - Desire to find out the truth (willingness to pay)
  - ability to evaluate evidence/inference
  - ability to explain own reasoning

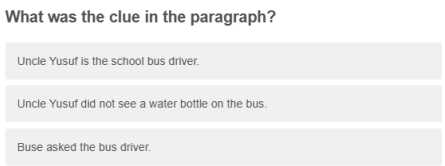


Figure A.3: Ability to justify reasoning

Back: Critical Thinking Task

Predictive Validity: Willingness to Take Clues

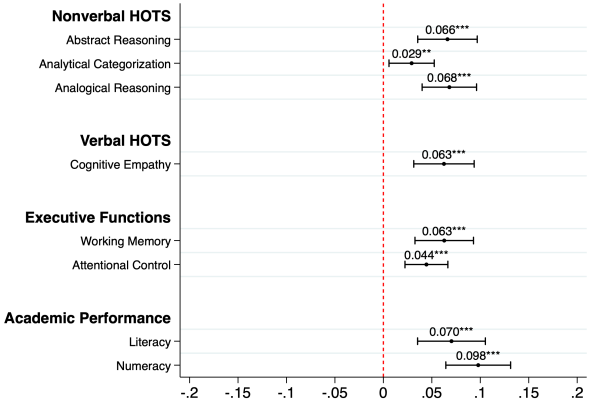


Figure A.4: Correlation of the number of clues taken with cognitive skills in the control group

Back: Predictive Validity

Predictive Validity: Inference and Reasoning

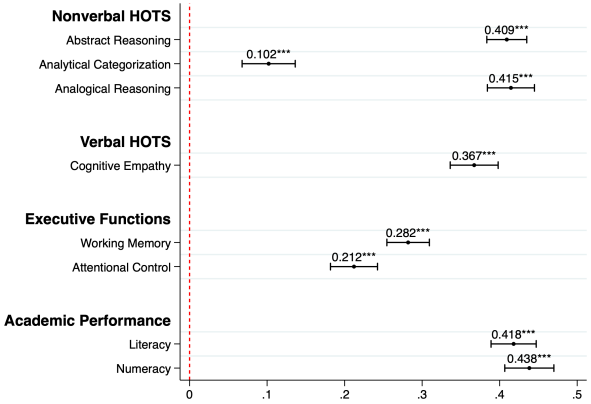


Figure A.5: Correlation of the number of correct elimination and justification with cognitive skills in the control group

Back: Predictive Validity

Predictive Validity: Inconsistency Detection

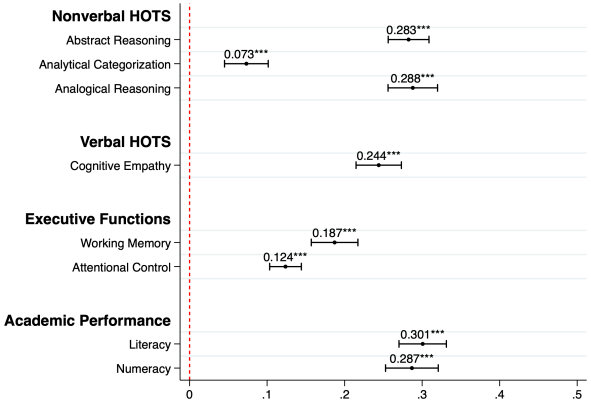


Figure A.6: Correlation of the number of correctly detected inconsistencies with cognitive skills in the control group

Back: Predictive Validity

Heterogeneity by Gender: Cognitive Skills and Academic Performance

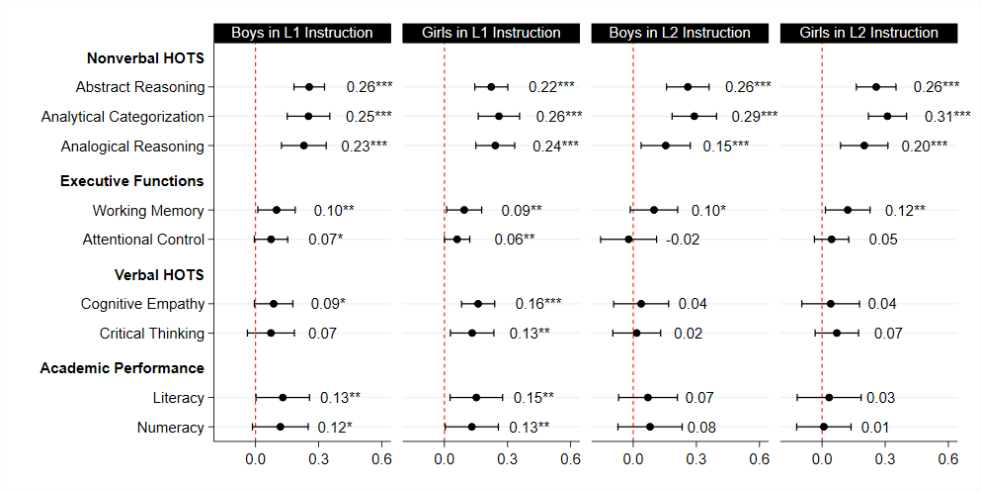


Figure A.7: Heterogeneous treatment effects by gender and mother-tongue

Cognitive Outcomes of Boys: Heterogeneity by Mother-tongue

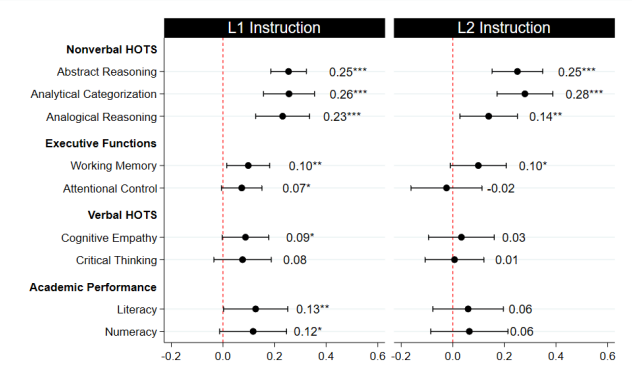


Figure A.8: Heterogeneous treatment effects by mother-tongue for boys

The significance patterns for the L1–L2 difference test are shown at the far right of the figure.

Back to Heterogeneity

Cognitive Outcomes of Girls: Heterogeneity by Mother-tongue

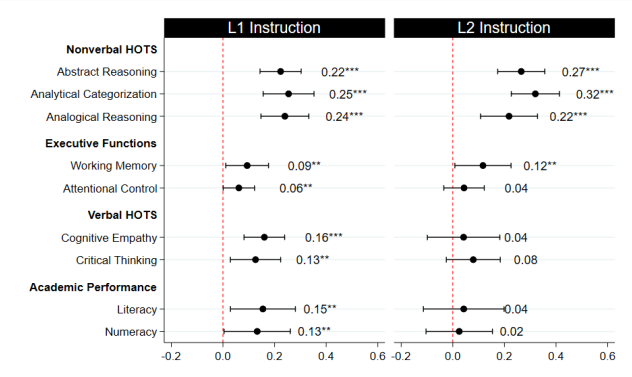


Figure A.9: Heterogeneous treatment effects by mother-tongue for girls

The significance patterns for the L1–L2 difference test are shown at the far right of the figure.

Back to Heterogeneity

Willingness to Pay for Clues: Heterogeneity by Gender and Mother-tongue

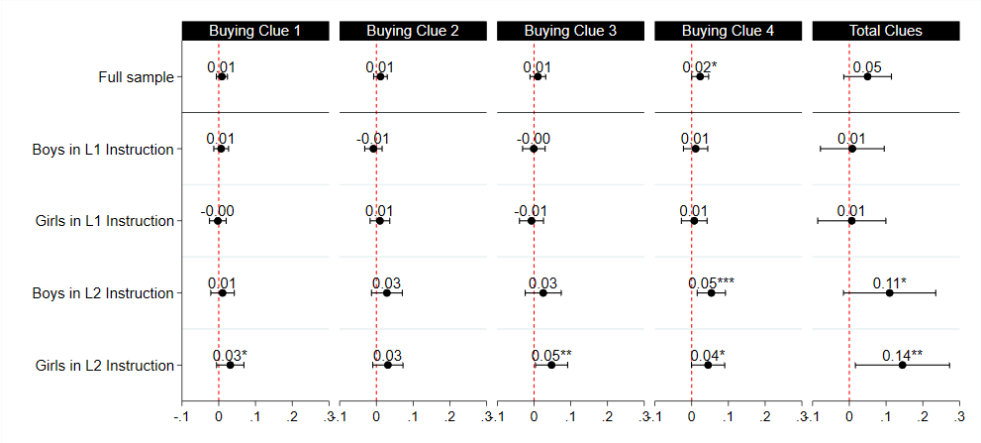


Figure A.10: Heterogeneous treatment effects by gender and mother-tongue

Inference and Reasoning: Heterogeneity by Gender and Mother-tongue

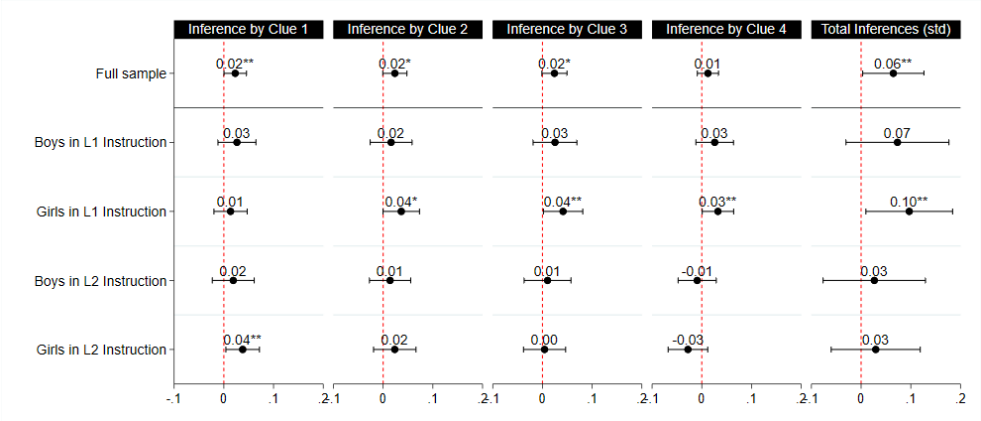


Figure A.11: Heterogeneous treatment effects by gender and mother-tongue

Back to Critical Thinking Game

Teacher Outcomes: Heterogeneity by age

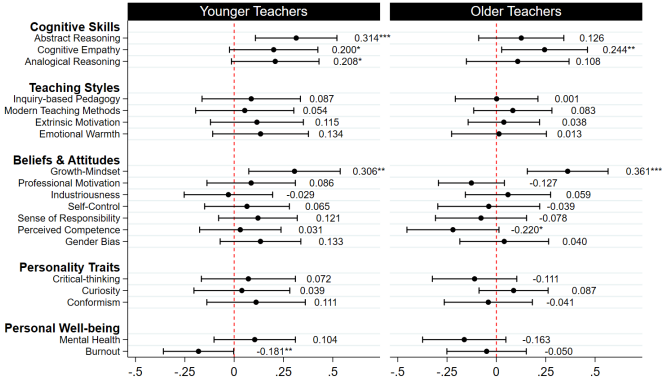


Figure A.12: Heterogeneous effects on teaching styles and teachers' cognitive outcomes by age

Back: Teacher Outcomes

Teachers' cognitive outcomes at the baseline

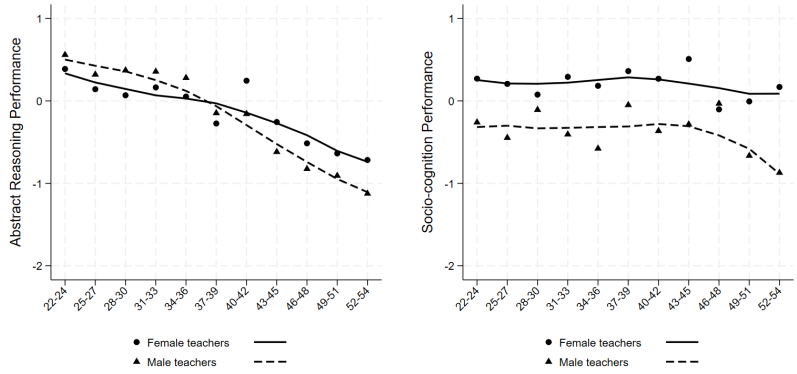


Figure A.13: Teachers' abstract reasoning and socio-cognition outcomes at the baseline by age: Local linear fit

Back: Teacher Outcomes

Teachers Outcomes: L1 Instruction

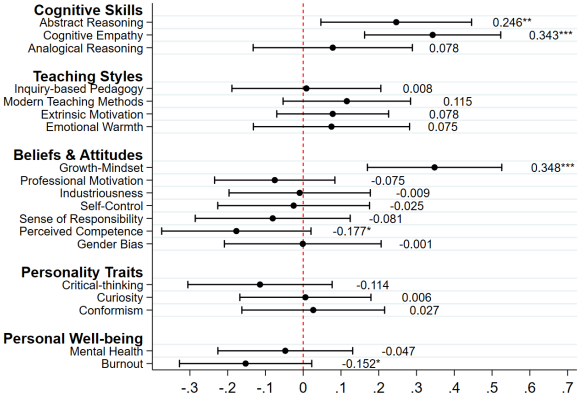


Figure A.14: Effect on teaching styles and teachers' cognitive outcomes in Gaziantep

Back: Teacher Outcomes

Teachers Outcomes: L2 Instruction

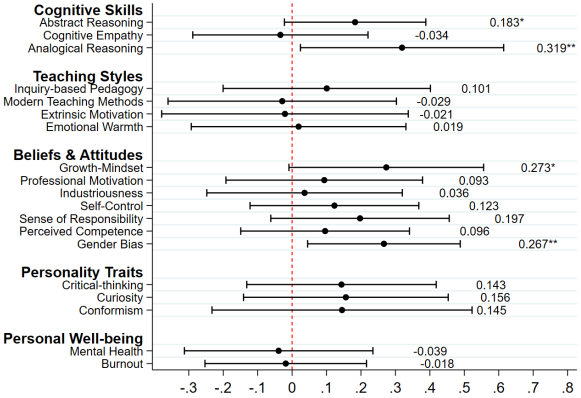


Figure A.15: Effect on teaching styles and teachers' cognitive outcomes in Sirnak

Back: Teacher Outcomes

Robustness Checks

- Checked the robustness of our results:
  - Without covariates Go to results
  - Multiple hypothesis testing Go to results

Robustness Check: Covariates

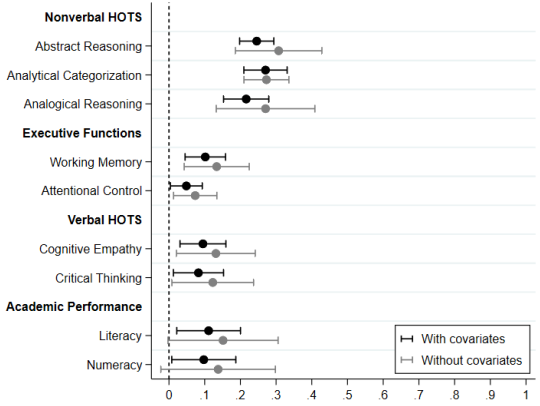


Figure A.16: Treatment effects with and without covariates

Back: Robustness Checks

Multiple Hypothesis Testing

	Original	Romano–Wolf
Panel A: Full Sample		
Nonverbal HOTS	0.0000	0.0010
Verbal HOTS	0.0018	0.0018
Executive Functions	0.0004	0.0010
Academic Performance	0.0197	0.0197

Table A.1: Original p-values and p-values corrected for multiple hypothesis testing using the Romano-Wolf algorithm

Back: Robustness Checks

Student Outcomes: Heterogeneity by Teacher Age

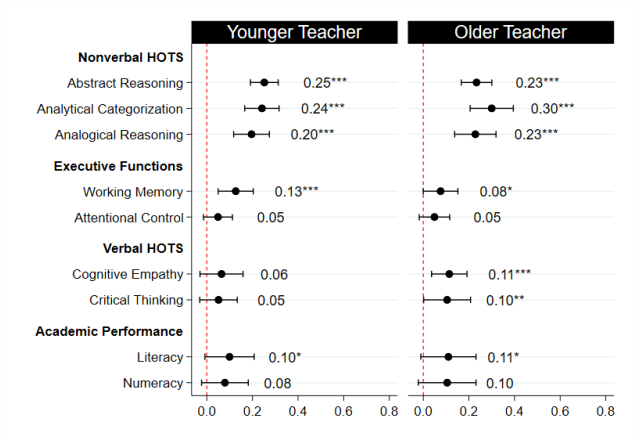


Figure A.17: Heterogeneous effects on non-language based skills by teacher's age

Student Outcomes: Heterogeneity by Baseline Ability

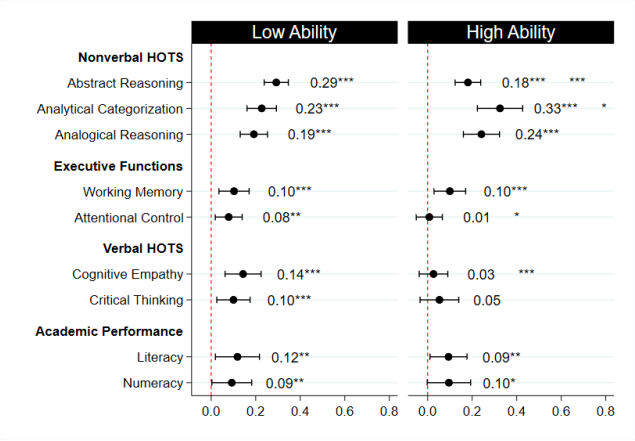


Figure A.18: Heterogeneous effects by baseline ability