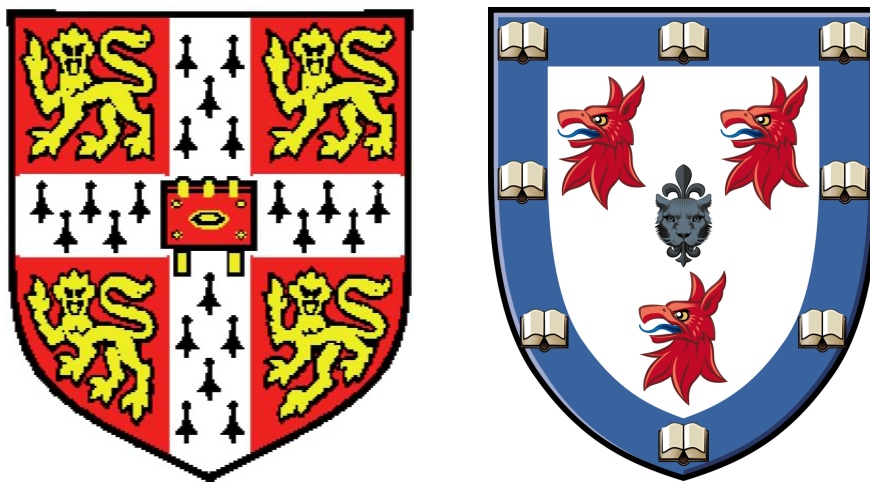


An Investigation of the Relationships Between Executive Functions, General Cognitive Ability, Science Education and Explanatory Coexistence

Rui Wang



Homerton College

University of Cambridge

Supervisor: Prof. Michelle R Ellefson & Dr. Hale Ögel-Balaban

Advisor: Prof. Sara Baker

This thesis is submitted for the degree of

Doctor of Philosophy, Education

June 2024

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the preface and specified in the text. It is not substantially the same as any work that has already been submitted, or, is being concurrently submitted, for any degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the preface and specified in the text. It does not exceed the prescribed word limit for the relevant Degree Committee.

Abstract

From a young age, children develop naïve thinking by observing the world around them (i.e., ‘the Earth is flat because it looks flat’). Correcting such naïve thinking is crucial for effective scientific learning and is a key goal in primary science education, but the process of correction is not easy, as even adults may make mistakes in conflicts between naïve and science thinking. Since the 2010s, empirical studies have suggested that naïve thinking is not simply replaced or changed by, but can coexist with science thinking, in a state of explanatory coexistence. Numerous studies have explored the underlying mechanisms of this explanatory coexistence, yet no consensus conclusion has been reached, although executive functions, general cognitive ability and science education are thought to be its influential contributing factors. This study therefore measures the relative contribution of each of these factors to explanatory coexistence and includes a case study from a non-WEIRD (Western, Educated, Industrialized, Rich, and Democratic) population, to expand the purview of its research findings.

Two studies were conducted to gauge the relative contribution of explanatory coexistence's impact factors. Online computerised cognitive tasks, including four executive functions, one general cognitive ability and three naïve thinking tasks, as well as demographic questionnaires for science education, were used on cross-cultural populations. Study 1 recruited 262 adult participants from the United Kingdom. Results of correlation and path analysis showed that switching and inhibition played the most significant roles in explanatory coexistence. While executive functions and general cognitive ability were contributors; working memory and science education did not contribute to explanatory coexistence at all. Different contributing factors were also involved when task modality varied: while switching contributed to both picture- and text-based tasks, behavioural and interference inhibition contributed to picture-based naïve thinking tasks only, and general cognitive ability contributed to text-based naïve thinking tasks only. Such links were found to be robust across trials,

regardless of naïve and science thinking being congruent or incongruent. Study 2 recruited 234 adult participants from China to conduct cross-cultural research. Using the same research method as in Study 1, results indicated that, unlike participants from the United Kingdom, behavioural inhibition did not predict naïve thinking, a discrepancy that might be caused by differences in response patterns adopted by participants across the two countries, with Chinese participants spending greater time obtaining more accurate results. Regarding the links between other contributing factors and explanatory coexistence, however, findings on modality and task trials were robust across both countries. In sum, results from this study shed light on science education from a cognitive perspective, and indicate the further need to integrate consideration of executive functions into pedagogy design with explanatory coexistence concepts

Conference Talks & Preprint

Wang, R., Tsigardis, G. K., Zheng, J., Feeney, A., & Ellefson, M. R. (2023, August 22-26). *Biology education levels and general cognitive ability influence conceptual learning of biology* [Poster presentation]. European Association for Research on Learning and Instruction Conference 2023, Thessaloniki, Greece.

Wang, R. (2023, April 4). *Investigating the relative contributions of science education, executive functions, and general cognitive ability to the suppression of naïve thinking* [Departmental colloquium]. Department of Psychology, Occidental College, Los Angeles, United States.

Ellefson, M. R., **Wang, R.**, Britton, A., Godfrey, H., & Feeney, A. (2024). *Inhibition and switching make different contributions to counterintuitive science thinking in adults*. PsychArXiv. <https://doi.org/10.31234/osf.io/cm6kg>

Dedication

This dissertation is dedicated to my dearest grandmother, Mrs. Xiaoqin Yuan (苑晓琴).

Thank you for raising me and giving me endless love. You have given me the ability to love others, the courage to face challenges and the power to overcome different adversities.

You always encouraged me to be my best, which has given me endless courage. I would not have started my PhD studies or be the person I am today without your influence. I wish you rest in peace.

Acknowledgement

This PhD would not have been possible without the help and support of numerous people, first among them are the members of the INSTRUCT Lab. I would like to extend my foremost thanks to my supervisor, Prof Michelle R. Ellefson. In the five years working with Michelle, I have received the best guidance on my academic journey, with Prof Ellefson always encouraging me to “keep going despite current difficulties, as these are all good learning experience”. Her supervision and rigorous academic thinking have not only provided me with endless academic encouragement but have helped me maintain a positive attitude in the face of challenges in all areas of my life. I would also like to thank INSTRUCT’s deputy director and my second supervisor, Dr Hale Ögel-Balaban, who has always been available to me when I needed her and provided patient, detailed guidance, not only for my studies but also for my personal needs. Her help was invaluable when organising and structuring my writing, and her perspective has always inspired my work. A third thank you is due to Prof Aidan Feeney, whose insights into this research topic are what inspired this project. Meetings with Prof Feeney helped me interpret my research findings from different, and deeper, perspectives. I would also like to thank Aaron Britton and Harriet Godfrey for sourcing the Animism task materials for this study during its master stage, and my PhD examiners, Profs Sander L. van der Linden and Andrew Shtulman, for their time and consideration in examine my work. Both professors provided critical and insightful feedbacks on the conceptual understanding of this study’s research, data analysis, and interpretation, and I am thankful for having been able to engage them in viva conversation.

I would also like to extend my sincere thanks to other members of the INSTRUCT team, particularly Dr Maria Tsapali, who has always been helpful in resolving my questions; Dr Rina Lai, who has shared her experiences with me unreservedly; and Dr Miriam Yang, who assisted me in overcoming the difficulties we encountered together. Drs Tanya Peas, Gavin Turner, Annie Zonneveld, Tamer

Said, Claudia Chu and Helen Barsham, Nicky Hutson and all the other senior INSTRUCT members were there when I needed help, and I am especially grateful to them and the others I met at the INSTRUCT Lab; all of us learned and grew together. I collected data and prepared a systematic review draft with Lei Huang, to whom I owe thanks for always supporting my ideas. Thank you to Ibtehaj Alrushroud, with whom I exchanged data analysis insights, and Jing Xiao for their detailed aid in checking the results of the study's data analysis. To my most trustworthy collaborator, Dr Konstantinos Tsigaridis, I would like to say that meeting you during my PhD and collaborating with you has been a joy, you are like my older brother. Your insights into science education have brought me fresh academic perspectives, and I will remember the support you have provided for me whenever I face future difficulties.

During my time in Cambridge, I have received invaluable assistance from many senior members and peers. Dr Jillian Lauer provided excellent guidance on my summer research work and patiently guided me in my academic writing, helping me start our new investigation on children's perception of social status. My academic advisor, Prof Sara Baker, offered many valuable insights for my PhD work, while my college tutors, Drs Melanie Keene and Roberto B Sileo, provided support in my daily life, helping me to overcome difficulties. My friends in Cambridge also offered me substantial emotional support, and your friendships are a precious gift I have received in Cambridge. Special thanks to Yanyun Zhou, Wanning Zhang, Kefan Xue, Shengpeng Shi, Changzhu Fu, Zixuan Li, Zehui Yang, Ziqi Lin, Sirui Li, Yichen Wang, Jing Zhang, Shulin Wang, Yutong Wu and Yang Liu. I would like to give particular thanks to two of my friends in particular: Xueni Zhang, you are my best friend in Cambridge and trustworthy collaborator, and your support, both academic and personal, have been the best I could for; and Zhan Shi, your friendship has helped me navigate the toughest moments of my PhD, and sharing my thoughts and experiences with you has been a true joy.

I have also been an active participant of the Cambridge China Education Forum and made many friends in the process. During my final year, when I served as chair, I have been especially grateful to Dr Tyler Denmead, Prof Rupert Wegerif, Prof

Michelle Ellefson and Dr Jillian Lauer for their time and advice. I would like to give thanks to my fantastic Cambridge China Education Forum team, particularly our Secretary-General Shufan Zhang and Deputy Chair Mengxi Tian, who have stood by me in the face of difficulties. Additional thanks to all our committee members, sponsors and volunteers, the advisors from Chinese universities and speakers: without you, our forum could not be where it is today.

I would like to thank two off-campus academic mentors and research collaborators. Prof Robert J. Sternberg from Cornell University has always encouraged me to be myself, providing detailed assistance when needed, and their guidance has made me feel brave enough to be creative during my academic work. Also, Prof Hanbing Zheng from Durham University introduced me to the application of cognition in linguistics; their insights into exploring research questions have been profoundly impactful for my education. Thanks also to my research collaborators, Arezoo Soleimani, Vina Wong, Yumeng Wang, Anran He, Haoshen He: working with you has been a delightful experience. Additionally, to all the senior scholars in my research field who have contributed to my research during conferences or individual meetings: Profs Patrice Potvin, Panayiota Kendeou, Stella Vosniadou, Iro Xenidou-Dervou, Danny Oppenheimer and Christian D. Schunn.

I would like to extend my thanks to my undergraduate professors at Waseda University, Profs Azuma Rena and Adam Guy, whose support has allowed me to embark on my academic journey in Cambridge. My high school teacher, Yi Yang, always encouraged me to keep trying when I was a teenager, and their support has been the warmest motivation for helping me advance further. And to my undergraduate friends Chunxi Pan, Jia Chen, Yaxu Yu, Jiayuan Chen and Kim Pan, your support has been the driving force behind my progress.

Heartfelt thanks go to my family: my parents Jianwen Wang and Li Gao, my grandparents Zhengfu Wang and Xiaoqin Yuan, my grandmother Xiaoping Wu, and my family mentors Jianzhong Wang, Yu Shu, Lin Gao, Yuxin Liu, Lina Wu, Guanlin Song and Chunyang Song; all of you have supported and understood my decisions, and have been proud of me during this process. You are the best family members I

could ever imagine. My lifelong friend Linting Hu is the spiritual companion by my side, while Yushu Shi and Anzhe Li are always there when I need them, their friendship being the most precious thing in my life. Finally, I would like to thank my husband, Tianren Wang, who has encouraged me to pursue my dreams, accompanied me through various phases of life, cheered for any progress I made and has always had my back.

I would like to express my gratitude for the technological advancements provided by the modern world. I used ChatGPT to assist with checking the grammar and typos in parts of this study's literature review, methodology and discussion sections, while Google and YouTube, my best friends, were there to answer all my stupid questions anytime I needed. Finally, I would also like to thank Anya Ciccone to help me to proofread the full draft.

On a final note, thanks very much for the supports from my funders, including Cambridge Trust-CSC Scholarship, Faculty of Education Research Grant and Homerton Research Grant.

Table of Contents

Declaration.....	2
Abstract.....	3
Conference Talks & Preprint	5
Dedication	6
Acknowledgement.....	7
Table of Contents	11
List of Tables.....	16
List of Figures.....	18
Chapter 1: Introduction	20
1.1 Research Motivation	20
1.2 Importance of This Research.....	20
1.3 Outline	22
Chapter 2: The Mechanisms Underlying Explanatory Coexistence	25
2. 1. Conceptual Change and Explanatory Coexistence.....	25
2.1.1 Naïve Thinking and Science Thinking.....	25
2.1.2 Naïve Thinking in Science Education.....	26
2.1.3 Previous Understandings of Naïve Thinking and Science thinking.....	27
2.1.4 Explanatory Coexistence.....	29
2.1.5 The Possible Underlying Mechanism of Explanatory Coexistence.....	31
2.2. Executive Functions and Explanatory Coexistence	33
2.2.1 Overview of Executive Functions	33
2.2.2 Inhibition	34
2.2.3 Switching.....	35
2.2.4 Working Memory	35
2.2.5 The Development and Interrelationship of Executive Functions.....	36
2.2.6 Executive Functions and Explanatory Coexistence	39
2.2.7 Research Gaps on the Link Between Executive Functions and Explanatory	

Coexistence	43
2.3. General Cognitive Ability and Explanatory Coexistence.....	48
2.3.1 General Cognitive Ability, Executive Functions and Education Outcomes.....	48
2.3.2 General Cognitive Ability and Explanatory Coexistence	50
2.3.3 Research Gaps on the Link Between General Cognitive Ability and Explanatory Coexistence	51
2.4. Science Education and Explanatory Coexistence	53
2.4.1 How Science Education Contributes to the Explanatory Coexistence.....	53
2.4.2 Science Education, Executive Functions and General Cognitive Ability	54
2.4.3 Research Gaps on the Link Between Science Education and Explanatory Coexistence	56
2.5. Research Gaps and Research Questions in Study 1	57
Chapter 3: A Cross-Cultural Exploration of Explanatory Coexistence.....	59
3.1 Science Learning and Explanatory Coexistence in Cross-Cultural Contexts	60
3.2 Cognitive Ability and Academic Performance in Cross-Cultural Context.....	63
3.2.1 Executive Functions in a Cross-Cultural Context.....	64
3.2.2 General Cognitive Ability in a Cross-Cultural Contexts.....	65
3.2.3 Robust Links Between Cognitive Ability and Academic Performance in Cross-Cultural Context	66
3.3 The Education System in the United Kingdom and China	68
3.3.1 The United Kingdom Education System.....	68
3.3.2 The Chinese Education System.....	70
3.3.3 A Comparison of the United Kingdom and Chinese Education Systems	71
3.4 Research Gaps and Research Questions in Study 2	71
Chapter 4 Method	74
4.1 Participants	74
4.2 Methods	77
4.2.1 Demographic Information Questionnaire	77
4.2.2 Computerised Behaviour Tasks.....	78
4.3 Procedure.....	84

4.4 Statistical Analyses Procedures	84
4.4.1 Preliminary Analysis	85
4.4.2 Main Analysis.....	89
Chapter 5 Results and Discussion for Study 1: Analyses Using the United Kingdom	
Data	91
5.1 Preliminary Analysis (the United Kingdom).....	92
5.2 Correlation Analysis (the United Kingdom)	93
5.3 Path Analysis (the United Kingdom)	96
5.4 Discussion on Study 1	101
5.4.1 Relative Contributions of Executive Functions, General Cognitive Ability and Science Education on Explanatory Coexistence	103
5.4.2 Different Modalities Require Different Skills.....	107
5.4.3 Congruent and Incongruent Trials Require Different Skills	109
5.4.4 Links Between Study 1 and Study 2	111
Chapter 6 Results and Discussion for Study 2: Analyses Using the United Kingdom and Chinese data	113
6.1 Preliminary Analysis (China).....	114
6.2 Response Patterns Between the United Kingdom and China.....	115
6.3 Preliminary Analysis (the United Kingdom and China)	123
6.4 Multigroup Analysis Between the United Kingdom and China.....	123
6.5 Correlation Analysis (China).....	126
6.6 Path Analysis on Efficiency (China)	128
6.7 Discussion on Study 2	133
6.7.1 Response Patterns of Participants from the United Kingdom and China	136
6.7.2 Relative Contributions of Executive Functions, General Cognitive Ability and Science Education on Naïve Thinking Tasks are Different Between the United Kingdom and China	139
Chapter 7 General Discussion and Conclusion	147
7.1 Key Finding 1: The Underlying Mechanism of Explanatory Coexistence	150
7.1.1 The Involvement of Different Contributing Factors in Explanatory Coexistence	

.....	150
7.1.2 Overview the Underlying Mechanism of Explanatory Coexistence	157
7.2 Key Finding 2: Influences of Modality on the Reasoning of Explanatory Coexistence	160
7.2.1 Links Between Different Contribution Factors and Modality	160
7.2.2 Views from the Modality Perspective	162
7.3 Key Finding 3: Explanatory Coexistence Within a Cross-Cultural Perspective	164
7.3.1 Response Patterns in Cross Cultural Research.....	164
7.3.2 Robust Links Between Cognitive Skills and Explanatory Coexistence.....	166
7.4 Exploratory Finding: Story of Efficiency.....	167
7.5 Limitations and Future Directions.....	169
7.5.1 Limitations on Age Representation	169
7.5.2 Overlaps Between Executive Function Measurements	170
7.5.3 Bias During Cross-Cultural Data Collection.....	171
7.5.4 Limitations of the Form of Task Modality	172
7.5.5 Limitations on Science Educations Measurements.....	172
7.5.6 Limitations on Non-verbal Reasoning Tasks	173
7.6 Practical Implications	174
7.7 Conclusion.....	177
References	179
Appendices.....	234
Appendix A: Research Methodology	234
Appendix B: Revision of Animism and Flanker Tasks	279
Appendix C: Influence of Fixed Button of Animism Task and Science Statement Task	285
Appendix D: Participant Questionnaire	288
Appendix E: Consent Form.....	292
Appendix F: Information Sheet.....	294
Appendix G: Science Statements Task.....	296
Appendix H: Accuracy Analysis (the United Kingdom).....	300
Appendix I: Response Time Analysis (the United Kingdom).....	313
Appendix J: Efficiency Analysis (the United Kingdom).....	326

Appendix K: Accuracy Analysis (China)	336
Appendix L: Response Time Analysis (China)	349
Appendix M: Efficiency Analysis (China)	362
Appendix N: Efficiency Analysis (Combined Data)	375
Appendix O: Accuracy of General Cognitive Ability and Efficiency for All Other Variables	386

List of Tables

Table 1.1	24
<i>Summary of the Chapters</i>	
Table 4.1	75
<i>Demographic information for participants from the United Kingdom</i>	
Table 4.2	76
<i>Demographic information for participants from China</i>	
Table 4.3	86
<i>Details of Steps in Preliminary Analysis</i>	
Table 4.4	90
<i>Details of Steps in Main Analysis</i>	
Table 5.1	95
<i>Correlation Analysis on Efficiency (the United Kingdom)</i>	
Table 5.2	97
<i>Path Analysis Model Fit Indices Summary on Efficiency (the United Kingdom)</i>	
Table 5.3	102
<i>Summary of Study 1 Analysis and Corresponding Discussion</i>	
Table 5.4	106
<i>Summary of Each Task's Modality</i>	
Table 6.1	124
<i>Fit statistics and Significance Tests for Multigroup Analysis on Efficiency</i>	
Table 6.2	126
<i>Correlation Analysis on Efficiency (China)</i>	
Table 6.3	128
<i>Path Analysis Model Fit Indices Summary on Efficiency (China)</i>	
Table 6.4	133
<i>Summary of Analysis and Corresponding Discussion for Study 2</i>	
Table 6.5	141

*Summary of effect size of significant links in United Kingdom and Chinese
Participants*

Table 7.1..... 147

*Summary of Studies, Research Questions, Main Research Results and Key
Research Findings*

List of Figures

Figure 5.1a.	98
<i>Path Analysis for Overall trials on Efficiency (the United Kingdom)</i>	
Figure 5.1b.	99
<i>Path Analysis for Congruent trials on Efficiency (the United Kingdom)</i>	
Figure 5.1c.	100
<i>Path Analysis for Incongruent trials on Efficiency (the United Kingdom)</i>	
Figure 6.1a.	115
<i>Scatterplot of the Link Between Accuracy and Response Time Between China and the United Kingdom: Stop Signal Tasks and Flanker Task</i>	
Figure 6.1b.	116
<i>Scatterplot of the Link Between Accuracy and Response Time Between China and the United Kingdom: Figure Matching Tasks and Corsi Block Task</i>	
Figure 6.1c.	117
<i>Scatterplot of the Link Between Accuracy and Response Time Between China and the United Kingdom: Matrix Reasoning Task and Science Statement Tasks</i>	
Figure 6.1d.	118
<i>Scatterplot of the Link Between Accuracy and Response Time Between China and the United Kingdom: Animism Task and Buoyancy Tasks</i>	
Figure 6.1e.	119
<i>Scatterplot of the Link Between Accuracy and Response Time Between China and the United Kingdom: Congruent and Incongruent Trials of Science Statement Tasks</i>	
Figure 6.1f.	120
<i>Scatterplot of the Link Between Accuracy and Response Time Between China and the United Kingdom: Congruent and Incongruent Trials of Animism Tasks</i>	
Figure 6.1g.	121

*Scatterplot of the Link Between Accuracy and Response Time Between China
and the United Kingdom: Congruent and Incongruent Trials of Animism Tasks*

Figure 6.2a 129

Path Analysis for Overall trials on Efficiency (China)

Figure 6.2b. 130

Path Analysis for Congruent trials on Efficiency (China)

Figure 6.2c. 131

Path Analysis for Incongruent trials on Efficiency (China)

Chapter 1: Introduction

1.1 Research Motivation

Science education is considered to be essential for today's society, as it equips people with knowledge and skills needed to engage with society at large and contribute to future innovations and discoveries (Fugelsang & Mareschal, 2013; Mason et al., 2019; Osborne & Freyberg, 1985). Promoting science learning involves not only setting appropriate pedagogy but teacher training and instilling an understanding of science learning's cognitive processes (Ellefson et al., 2019). Cognitive science, which covers how the human mind processes information and shapes the way individuals experience the world, can therefore be used to help educators better comprehend student thinking patterns and help them design effective teaching methods.

During my masters' study at the University of Cambridge, I participated in a collaborative research project at the intersection of science education and cognitive science, co-led by Prof Michelle R. Ellefson (University of Cambridge) and Prof Aidan Feeney (Queen's University Belfast) and alongside two master's students from Queen's University Belfast, Aaron Britton and Harriet Godfrey. The research investigated links between two components of executive functioning, inhibition and switching, and explanatory coexistence. This research work was turned into a manuscript (Ellefson et al., 2024) whose detailed results are provided in Section 2.2.6. Observing first-hand how cognitive science research was conducted during this project, I grew interested in the vital role played by executive functions in science thinking, and with further research I designed this PhD project to further explore the mechanisms underlying explanatory coexistence, and whether this mechanism was robust across cultures (Aikenhead & Jegede, 1999).

1.2 Importance of This Research

Student difficulties in shifting from naïve to science thinking have long been

observed within science classrooms, and the problem exists among both among children and adults (Hodson, 2003). Research suggests that correcting naïve thinking is a challenging procedure that involves multiple processes and conditions (Chan et al., 2017; Lewandowsky et al., 2012), and several models and theories have been proposed to explain this correction process, how it is conducted and the conflicts it generates (see models put forward by Kendeou et al., 2014 and Nadelson et al., 2018). These frameworks have been tested in existing empirical research, yet few have measured directly the roles played by different cognitive components (i.e., executive functions, general cognitive ability) and academic experience (i.e., science education) in resolving naïve thinking. Moreover, current research on naïve thinking has focused predominantly on WEIRD (Western, Educated, Industrialised, Rich and Democratic) samples, resulting in a scarcity of research on people from various countries and regions and leading to a lack of generalisability in research findings.

This research conducts two studies assessing adult participants in both the United Kingdom and China, in order to provide clearer insights into key contributing factors to selecting scientific answers, and the manner in which these factors are used during the selection process. Through its design of different trial types with varying modalities for naïve thinking tasks, this research will contribute to the understanding of mechanisms underlying scientific reasoning at the cognitive level, whilst also considering individual differences stemming from demographic features. Its results may generate a directive meaning for similar research in children.

From a practical perspective, although this research is unlike other intervention or design-based research studies that might provide direct suggestions for improving pedagogy, it may provide both researchers and educators worldwide with knowledge about how people process counter-intuitive ideas. Therefore, this research may provide ideas for more scientific designs for science classroom pedagogy, thereby enhancing people's later acquisition of learned scientific knowledge. As correcting naïve thinking is common for both children and adults, improvements may not be limited only to adult science classrooms, but may inform children's and adolescents' education (see research conducted by Brookman-Byrne et al., 2018; Potvin et al., 2015; Shtulman & Valcarcel, 2012). Meanwhile, with a deeper understanding of how executive functions, general cognitive ability and science education participate in this process, future intervention programmes may provide useful strategies for enhancing learner skills in solving naïve scientific questions.

1.3 Outline

As mentioned above, this research project included two interrelated Studies. In Study 1, I expanded upon my master's research exploring the possible underlying mechanisms of explanatory coexistence, using participants from the United Kingdom. Study 2 replicated Study 1 with Chinese participants to explore whether the factors influencing explanatory coexistence found in Study 1 are robust roles across cultures. The structure of this research project has been designed with the following considerations in mind: a) the possible underlying mechanisms of explanatory coexistence reviewed in this paper will require a substantial amount of text; b) the research and analysis methods used in Studies 1 and 2 are roughly the same; and c) the analysis and results of Study 2 are based on an integrated interpretation of the analysis and results of Study 1.

Chapters 2 and 3 of this paper provide a literature review for this research. Chapter 2 explains the research background for Study 1, discussing concepts related to explanatory coexistence, possible influence factors put forward in current research and the current understanding of these factors. Chapter 3 reviews the research background for conducting cross-cultural research comparisons and results of all factors mentioned in Chapter 2. Due to the conceptual overlap between Chapters 2 and 3, I do not repeat the explanations for concepts in Chapter 3 that have been provided in Chapter 2, instead emphasising the material necessary for researching cross-cultural contexts. Chapters 4 of this dissertation is the method section. Chapter 4 presents participant information, materials, procedures, ethical considerations and open science, providing the Studies' specific research steps. Chapters 5 and 6 offer the Studies' research results and a brief discussion of their implications, with Chapter 5 focusing on results from Study 1, analysing the data obtained from UK participants, and Chapter 6 corresponding this data to the results of Study 2, including a comparison between the United Kingdom and China following the addition of Chinese data. Finally, Chapter 7 concludes this thesis, providing a general discussion combining the overall results from Study 1 and 2, situating these results against this

research's theoretical frameworks and outlining the study's limitations, implications and future research directions. Table 1.1. summarises these chapters and related studies in the current research.

Table 1.1*Summary of the Chapters*

Chapter	Contents	Related Study
1	Introduction	Study 1&2
2	Literature Review	Study 1
3	Literature Review	Study 2
4	Method	Study 1&2
5	Results and Discussion	Study 1
6	Results and Discussion	Study 2
7	General Discussion and Conclusion	Study 1&2

Chapter 2: The Mechanisms Underlying Explanatory Coexistence

2. 1. Conceptual Change and Explanatory Coexistence

2.1.1 Naïve Thinking and Science Thinking

Learning involves an intricate interplay between a learner's pre-existing knowledge and newly acquired information (Dykstra et al., 1992; Posner et al., 1982; Vosniadou, 2013). When grasping a new concept, individuals diverge on how to integrate this concept with their prior knowledge (Bartlett, 1995), an issue critical for science thinking as compared to prior knowledge, scientific knowledge is learned later and constantly updated by experts in the scientific field (Chakravartty, 2022). Prior knowledge may be consistent with current scientific knowledge, but may also conflict (Ackerman, 2007; Brod, 2021; Hailikari et al., 2008; Staub & Stern, 2002): for example, although current scientific knowledge does not associate gravity with weight, people may believe that a heavier ball falls more quickly than a lighter ball when dropped from the same height in a vacuum (Potvin et al., 2023).

While correct prior knowledge is critical for the acquisition of science knowledge, serving as a foundation for future learning, leading to the proper grasp of a subject and creating lasting connections with incoming information (Pintrich et al., 1993; P. L. Ackerman, 2007; Brod, 2021; Hailikari et al., 2008; Staub & Stern, 2002), incorrect prior knowledge hinders learning, conflicting with incoming scientific knowledge and making learning more challenging (Alvermann & Hague, 1989; Chen et al., 2020; Menz et al., 2021). These conflicts result in misconceptions (Allen, 2014; Braasch et al., 2013; Tippet, 2010) such as childhood animism, when young children associate movements with the status of being alive (Inagaki & Hatano, 1996; Bird-David, 1999). In this case, children find it challenging to conceptualise clearly the notion of being alive, and are likely to conflate living with other concepts such as reality, visibility, presence or existence (Carey, 1985). Thus, childhood animism may

be considered a form of naïve thinking that results in incorrect information: young children will categorise the sun, wind, animals and humans as living while static plants and items as non-living. It is only when these young children grow and enter school that they gradually understand that some stationary objects (i.e., trees) are also alive, and that certain moving entities (i.e., cars) operate by mechanical principles, and that their movement does not indicate life (Kim, 2015).

Due to the broad, somewhat segmented nature of the literature related to prior knowledge, misconceptions about incorrect prior knowledge have been referred to as naïve concepts (Babai & Amsterdamer, 2008), earlier intuitions (Shtulman & Valcarcel, 2012), intuitive conceptions (Potvin et al., 2015), alternative conceptions (Garnett et al., 1995), preconceptions (Clement, 1993) or erroneous prior knowledge (Kendeou & van den Broek, 2007). These definitions often vary depending on the scope of the incorrect prior knowledge studied, or the incorrect prior knowledge's origin (Duit & Treagust, 2003). To ensure the consistent use of terminology throughout this PhD research, I refer to these misconceptions about incorrect prior knowledge, collectively, as *naïve thinking* in all chapters. For parallel purposes, I also refer to later learnt science knowledge as *science thinking*.

2.1.2 Naïve Thinking in Science Education

Children enter the classroom with experience- or intuition-based prior knowledge about how the world works (Carey, 1985; Guzzetti, 2000). While this prior knowledge is essential in helping them develop a proper grasp of the subjects they learn in classrooms, it is often difficult to correct their naïve thinking solely with educational instruction (Ackerman, 2007; Tricot & Sweller, 2014), which has led researchers to describe naïve thinking as being strong and robust. Such naïve thinking is found in various academic domains, including history (Limón, 2002), psychology (Taylor & Kowalski, 2004), politics (Nyhan & Reifler, 2010) and especially in science (Carey, 2000). Studies have found naïve thinking in numerous school science, not only in the aforementioned biology classes, but also in courses such as physics.

For example, while learning about the topic of electric circuits, students may think a single wire is sufficient for lighting up a bulb (Çepni & Keleş, 2006). Moreover, students often have naïve thinking about buoyancy, believing that larger objects sink more quickly and deeper than smaller ones (Hsin & Wu, 2011; Potvin et al., 2015).

Resistance of naïve thinking to alteration in science domains has been found in learners of all ages (Tippett, 2010), and frequently results from lack of clarity. For example, studies indicate that naïve thinking regarding buoyancy is rooted in students' confusion regarding the concepts of mass, volume, weight and density (Smith et al., 1985), compounded with the intuitive rule of 'more A–more B' in scientific reasoning (Stavy & Tirosh, 2000). Buoyancy naïve thinking may therefore start from a very early stage, such as when infants anticipate object behaviours using size cues (Wang & Baillargeon, 2008) and deduce weight from surface compression, associating heavier items with increased force (Hauf et al., 2012). Later in childhood, buoyancy reasoning based on object heaviness may emerge, leading to the conflation of the concept of weight with size (Treagust & Duit, 2008). Even among older children and adults, discerning the concept of buoyancy may become difficult when an object's volume masks its density (Potvin & Cyr, 2017).

2.1.3 Previous Understandings of Naïve Thinking and Science thinking

The shift from naïve to science thinking is known as the term *conceptual change*. Conceptual change encompasses various similar yet distinct theories (Carey, 2000; Chi, 2005; diSessa, 1993; Kendeou & O'Brien, 2014; Shtulman & Lombrozo, 2016; Vosniadou, 2013). This section focuses on conceptual change as it relates specifically to naïve thinking.

Early exploration of conceptual change was first approached from a philosophical perspective. Kuhn (1970) theorised that scientific advancement does not occur in small, incremental steps but rather through revolutions. During these revolutions, old and new understandings are incommensurable, meaning that new theories require phrasings distinct from older ones. Later, Carey (1991; 1999) argued

that learning science entails more than just rote learning, but requires fundamental conceptual shifts. Drawing on empirical research into children's naïve biological understanding, Carey (1985) suggested that children experience a simultaneous, pivotal change in their concepts at certain points, and these changes aid children when dealing with conflicts between naïve and science thinking as they acquire new scientific knowledge. In another view, DiSessa (1993) introduced the 'Knowledge in Pieces' model of conceptual change, in which a learner's knowledge was segmented into various elements that could be rearranged and reorganised in different contexts, then activated depending on the situation at hand.

Chi et al. (1994) have suggested that conceptual change might occur both within and across key ontological categories (i.e., matter, processes, mental states), and that changing concepts within the same category is easier than changing concepts across categories. Profound changes are therefore gradual and require a thorough overhaul and reorganisation of one's entire belief system and network of presuppositions. Building on comprehensive research into children's conceptual knowledge development, Vosniadou (1994) has proposed the Framework Theory model, wherein children build initial explanations of physical phenomena from their everyday experiences. When these initial explanations confront are confronted with scientific explanations, children assimilate new explanations through a constructive, enrichment-type processes.

All of the above theories posit that naïve thinking might be changed, replaced, re-organised or integrated to science thinking, and therefore presuppose that the two types of thinking cannot simultaneously exist in the mind. However, recent empirical studies have contradicted this view (Babai & Amsterdamer, 2008, 2008; Potvin et al., 2015a; Shtulman & Valcarcel, 2012), and argue that individuals may retain naïve thinking throughout their lives, regardless of age and educational level.

2.1.4 Explanatory Coexistence

The possibility of the coexistence of naïve and science thinking stems from an analysis of differences in trial responses when comparing congruent and incongruent trials. Congruent trials refer to experimental trials without conflicts between naïve and science thinking, while incongruent trials are experimental trials that involve conflicts. In Babai et al. (2010), tenth graders ($N = 58$; aged 15- to 16- years) were studied to determine whether they retained a belief in naïve animism, with the researchers using a computerised animism task comprising four categories of trials: congruent trials, which included moving and living things (i.e., a cat) and non-moving, non-living things (i.e., a cup); and incongruent trials, which included moving but non-living things (i.e., a car) and non-moving but living things (i.e., a tree). The researchers observed participants' accuracy and response times when judging whether each given item was living or non-living, and found that participants responded more accurately and quickly to congruent than incongruent trials. This suggested that naïve animism still existed even after the children had learned the correct concept of animism.

Another study, conducted by Shtulman and Valcarcel (2012), reviewed 200 congruent and incongruent statements across 10 science and mathematics domains. In congruent statements, the correctness of both naïve and science thinking aligned (i.e., 'Rocks are composed of matter' or 'Numbers are composed of matter'), meaning that participants did not need to resolve conflicts between scientific and naïve thinking. However, incongruent statements placed naïve thinking at odds with scientific reasoning (i.e., 'Fire is composed of matter' or 'Air is composed of matter'). Testing 150 undergraduate participants, the researchers found that across all testing domains, participant responses to congruent statements were faster and more accurate than to incongruent statements. Consistent with Babai et al. (2010), these results suggest that greater cognitive effort is required in cases of conflict between naïve and science thinking, with even adults needing more time to suppress deep-rooted naïve beliefs.

Beyond the discovery that naïve thinking can coexist with science thinking, Potvin et al. (2015) have found that naïve thinking may interfere with science thinking, and a learner's ability to suppress it. Using a computerised naïve thinking task about buoyancy, the researchers showed adolescent participants ($N = 128$, aged 14- to 15- years) simulated balls of three different sizes (large, medium and small) made of wood, lead and polyvinyl chloride. Participants were asked to determine which ball displayed on the screen would sink faster in water. A negative priming paradigm was designed with five types of trials: very congruent (small and low-density ball vs. large and high-density ball); congruent (medium and low-density ball vs. large and high-density ball); neutral (two balls of the same size); incongruent (medium but high-density ball vs. large but low-density ball) and very incongruent (small but high-density ball vs. large but low-density ball). Negative priming effect was observed when participants' performance (response times or errors) was greater on the test probes (targets were distractors inhibited on the previous item) than on the control probes (targets did not relate to the previous item). The experiment thus replicated the findings of Babai et al. (2010) and Shtulman and Valcarcel (2012), and found that when people were presented with highly incongruent interference they took a longer time to respond. Potvin et al. (2015) thus concluded that the level of suppression of naïve thinking required is associated with the degree of conflict between naïve and science thinking.

Taken together, these three studies, using both tasks involving pictorial (the Animism and Buoyancy tasks) and text (the Science Statement task) stimuli, found that when people confront naïve thinking, they spend more time on the issue at hand, and make more errors in their solutions. This phenomenon is consistently observed across different domains and among participants of various age ranges. Based on this evidence, two new theories have been proposed: the *Explanatory Coexistence* view (Shtulman & Valcarcel, 2012) and the *Knowledge Revision Component* framework (Kendeou & O'Brien, 2014).

The *Explanatory Coexistence* view highlights the possibility that naïve and science thinking might coexist (Shtulman & Lombrozo, 2016), with the simultaneity

of scientific and supernatural thinking within the same domain, or with naïve thinking existing alongside science thinking about the same phenomena. Shtulman and Lombrozo (2016) argue that this coexistence may be necessary in certain domains, where multiple theories may be maintained to match different complexity models to problems requiring different levels of precision (i.e., quantitative vs. qualitative predictions of molecular diffusion), or to enable participation in multiple discourses (i.e., formal vs. informal discussions of heat). The *Knowledge Revision Components* framework (Kendeou & O'Brien, 2014) focuses more on the cognitive process of how naïve and science thinking change. This framework comprises five principles: 1. The encoding principle, which posits that once naïve thinking is encoded into long-term memory, it remains there; 2. The passive activation principle, which suggests that naïve information presented in working memory has the potential to be activated; 3. The co-activation principle, which suggests that naïve and science thinking are co-activated when encountering conflicts; 4. The integration principle, which proposes that science thinking be integrated with naïve thinking during the revision process; and 5. The competing activation principle, which indicates that even after individuals learn science thinking, naïve thinking still exists and is competitively activated in cases where an individual is confronted with conflict.

Both the *Explanatory Coexistence* view and the *Knowledge Revision Components* framework posit that naïve and science thinking coexist and suggest the need to explain the process from naïve to science thinking in cases of counterintuitive conflict. As some empirical studies suggest that naïve thinking is suppressed during the process from naïve to science thinking (i.e., Shtulman & Valcarcel, 2012), later sections of this thesis call this process *explanatory coexistence*.

2.1.5 The Possible Underlying Mechanism of Explanatory Coexistence

Building on the explanatory coexistence of naïve and science thinking, numerous studies have explored possible underlying mechanisms for explanatory coexistence (e.g., Mason et al., 2019; Nadelson et al., 2018; Potvin et al., 2023;

Shtulman & Legare, 2020; Tardiff et al., 2017). Among these studies, executive functions, general cognitive ability and science education have emerged as main contributing factors and have therefore been considered extensively (findings from past research are detailed in later chapters). The next sections introduce these factors sequentially, beginning with these factors' definitions, followed by the interrelationships and positions in the current research. The study's research gap will thereafter be summarised, and research questions outlined, in order to explore the possible underlying mechanism of explanatory coexistence.

2.2. Executive Functions and Explanatory Coexistence

In studies examining the cognitive mechanism of explanatory coexistence, executive functions are a crucial contributing factor that deserve further consideration. This section introduces the concept of executive function, its subcomponents and their inter-relationships, in order to provide basic information about executive functions.

2.2.1 Overview of Executive Functions

The term *executive function* was initially introduced by Pribram (1973) in reference to prefrontal cortex functioning during the regulation of cognition, emotion and behaviour. Subsequent to this research, 83 constructs were incorporated under the umbrella term ‘executive function’ (Baggetta & Alexander, 2016). The most commonly used paradigm of executive functions is Miyake et al.'s (2000) three-factor model, which includes switching (shifting or cognitive flexibility), updating (working memory) and inhibition (inhibitory control). Some models have been proposed after Miyake et al. (2000) have included some higher-order elements of executive functions, such as problem solving, reasoning and planning, which are relevant to areas about learning and thinking (Collins & Koechlin, 2012; Diamond, 2013).

The three main subcomponents of executive function have laid a foundation for cognitive approaches to learning and performance within educational environments (Baggetta & Alexander, 2016). Numerous studies have reported on executive function’s crucial contribution to various academic fields, including mathematics, reading comprehension, literacy and physical development (e.g., Artuso et al., 2019; Bull et al., 2008; Bull & Lee, 2014; Butterfuss & Kendeou, 2018; Ellefson et al., 2020; Magalhães et al., 2020; Xu et al., 2022). Research has also found executive functions to be crucial to science learning, with Rhodes et al. (2014) finding executive function to predict adolescents’ ability to learn biology concepts (i.e., DNA) significantly and Tisgaridis et al. (2022) finding executive functions to be significantly correlated to adolescents’ performance on physics tests. Executive

functions thus seem to influence how people understand and learn science concepts (Rhodes et al., 2016; Vosniadou et al., 2018).

The following sections introduce these three main executive function components, discuss their interrelationship, then summarise current literature on these executive functions' association with explanatory coexistence before identifying this study's research gaps.

2.2.2 Inhibition

Inhibition, also called inhibitory control or inhibition ability, is an individual's ability to control their attention, thoughts, emotions or behaviours, refraining from possible distractions (Miyake et al., 2000). While various definitions of inhibition have been proposed (Stahl, 1994), this research adopts that of Diamond (2012), which identifies two main types of inhibition, behavioural and interference inhibition. Behavioural inhibition, also known as response inhibition, refers to the ability to suppress motor responses or inappropriate actions that distract goal-driven behaviour (Diamond, 2013; Miyake et al., 2000). This inhibition is usually assessed using non-selective stopping tasks that require participants to stop automatic responses according to the presentation of a conditional cue (Barkley, 1997; Cheung et al., 2004). Interference inhibition, also known as semantic or attentional inhibition, refers to the ability to suppress irrelevant information or ideas when faced with conflicts (Diamond, 2013; Miyake et al., 2000). This type of inhibition plays a role in avoiding distracting stimuli from one's external environment, allowing for selective information or the resistance of unwanted memories or thoughts (Diamond, 2013; Friedman & Miyake, 2004; Nigg, 2000; Stahl et al., 2014; Tiego et al., 2018). Interference inhibition is usually assessed using stimulus-response compatibility tasks that require individuals to ignore irrelevant distractors and selectively pay attention to target stimuli (Friedman & Miyake, 2004; Lazar et al., 2015; Postle et al., 2004; Stins et al., 2008).

2.2.3 Switching

Switching, also known as cognitive flexibility or shifting, refers to an individual's ability to switch between different *task-sets*, or rules (Dajani & Uddin, 2015; Ellefson et al., 2006; Miyake et al., 2000). Switching is essential for individuals to flexibly and quickly adapt to changing situations, moving from one task to another (Davidson et al., 2006). Solving multiple cognitive tasks demands the proper use of mental resources, in this context called *task-sets* (Rogers & Monsell, 1995). When individuals perform complex tasks, they are likely to be influenced by different *task-sets*, and intentional executive control must be used to resist the temptation to focus on irrelevant goals, focusing on the target *task-sets* (Duncan et al., 1990). Thus, switching requires not only the flexible and rapid execution of a new task, but also a delicate balance between control and flexibility, to ensure that the ongoing task is free from distraction while allowing for potential new changes (Rogers & Monsell, 1995). Thus, unlike inhibition, switching focuses on the disengagement from old tasks and the uptake of new ones (Dajani & Uddin, 2015; Huizinga et al., 2006; Monsell, 2003). During this process, switch cost is generated, which typically manifests in a need for increased response time or the presence of elevated error rates (Jersild, 1927; Spector & Biederman, 1976; Yeung & Monsell, 2003b).

2.2.4 Working Memory

Working memory is a cognitive function that involves containing, modifying and retrieving new information (Diamond, 2013; Miyake et al., 2000; Postle et al., 2004). In contrast to short-term memory, in which new information is retained for a short period of time, working memory emphasises manipulating information for defined purposes. Scholars have proposed various models for conceptualising working memory (see Chai et al., 2018), including the *Multi-Component* model (Baddeley, 2003) and the *Time Based Resource Sharing* model (Barrouillet et al., 2009). As explanatory coexistence emphasises the interactions between naïve thinking in long-term memory and newly learnt science thinking, the *Embedded Processes*

model, proposed by Cowan (2001; 2010), has been selected to explain working memory in this study. This model constructs working memory as a single functioning mode that retains novel and existing information in an accessible state, which is frequently applied in the performance and manipulation of tasks requiring mental operations. The model regards working memory as an aspect of long-term memory that is temporarily active due to attention processes. Prior to becoming the focus of attention, information associated with long-term memory requires activation, a process that also involves awareness, as awareness and attention have been deemed coterminous (Cowan, 1999, 2001). When information from long-term memory is activated and presented to an individual's awareness, the central executive, a supervisory system that manages information processing and activates both current memory traces and environmental stimuli through an attention mechanism, starts to integrate these cognitive processes (Dumontheil & Klingberg, 2012; Engle et al., 1999; Kane et al., 2004). Thus, when conflicts arise between naïve and science thinking, working memory not only facilitates the reception of current information, but also retrieves naïve or science thinking from long-term memory. The central executive then integrates and evaluates this information, guiding judgement and decision-making processes.

2.2.5 The Development and Interrelationship of Executive Functions

Executive functions emphasise development and articulation, and are therefore closely related to and mature with age (Munakata et al., 2011). Like other cognitive skills, executive functions are typically acquired through experience, through both reiterative practice and problem-solving, which reinforce the executive functions and render their related neural circuitry more efficient (Zelazo & Carlson, 2012). Evidence suggests that the cognitive processes implicated in completing performance-based executive function tasks may undergo developmental reorganisation during the transition from childhood to adulthood (Best & Miller, 2010). However, although executive functions are generally stable in adulthood, the maturation time for each executive function factor varies slightly: while inhibition (Bedard et al., 2002; van den Wildenberg & van der Molen, 2004) and working memory (Brockmole & Logie,

2013; Huizinga & Molen, 2011; Isbell et al., 2015) have been found to develop in early adulthood, there is debate as to when switching matures. Some studies find adult-stage switching levels by age 12 (Anderson, 2002; Crone et al., 2006), while others argue that it continues to grow until around age 29 (Reimers & Maylor, 2005; Zelazo et al., 2014).

Apart from the different maturation time of executive function sub-components, executive functions' structure may differ between children and adults (Miyake & Friedman, 2012). In their literature review of research on executive function development, Best and Miller (2010) argue that each executive function subcomponent develops traits differently. Van der Ven et al. (2013) suggests that inhibition and switching are hard to distinguish until the ages of six and seven, while van der Sluis et al. (2007) have identified working memory and switching in 9- to-12 years olds through confirmatory factor analysis. Notably, the assessment of executive functions may show a speed-accuracy trade-off effect with age, with adults having longer response times in order to achieve higher accuracy, thus some assessments used efficiency score in their research (Ellefson et al., 2017; Monsell, 2003).

An inter-correlated but separate feature of executive function is called unity and diversity, with unity referring to the correlations between different subcomponents and possible shared underlying cognitive capabilities under different executive function assessments, and diversity focusing on the differentiation of all three executive function components (Friedman & Miyake, 2017). Miyake et al. (2000) have adopted confirmatory factor analysis to explore the unity and diversity structure of executive functions among adults in their different assessments of each executive function component, formulating a robust model for distinguishing the three subcomponents among adults (Friedman & Miyake, 2017). Consequently, executive function should be considered not only as one cognitive ability, but also as a network of connections and separations between different executive function components.

There is disagreement among researchers regarding which executive function component is dominant (Davidson, 2006; Egner & Hirsch, 2005; Gernsbacher & Faust, 1991; Hanania & Smith, 2010; Nieuwenhuis & Nick Yeung, 2005; Rubens & Zanto, 2011). Diamond (2013) has argued that inhibition and working memory are

more fundamental executive functions than switching, with the latter building on the former two, yet the only consensus currently reached among researchers is that these factors generally depend upon one another and co-occur. For example, working memory contributes to the promotion of inhibition by facilitating the identification of relevant and appropriate information by retaining objectives within the mind, reducing the probability of inhibition errors. Inhibition also supports working memory, as operational working memory's restricted capacity is essential for inhibiting both internal and external interference in order to maintain goal orientation (Hasher et al., 2008; Hasher & Zacks, 1984).

Of particular interest, however, is the relationship between inhibition and switching, as several researchers believe there should be a mechanism that shuts down one cognitive system before the switch to a different task. Allport and Wylie (1999) hypothesise put forward *proactive interference* as the continued priming of a task and suppression of a currently intended task, with inhibition playing a simple but vital role in the switching process. However, Yeung and Monsell (2003a) argue that switching involves increased activation, rather than inhibition, a view that is supported by a range of empirical studies (Costa & Santesteban, 2004; Meuter & Allport, 1999; Philipp & Koch, 2009). Theoretically, if switching involves inhibiting one task and activating another, it would draw heavily on working memory to load the new task (Diamond, 2013). A recent neuroimaging study has found that both switching and working memory share common neural processes (Wang et al., 2022), suggesting a functional overlap between these two executive functions, yet this finding is inconsistent: others studies on working memory capacity and working memory load have found no strong links between working memory and switching, although Draheim et al. (2016) have argued that the failure to find such a link may be the result of measurement bias.

As measurements of executive functions mostly rely on behaviour tasks, researchers must inevitably mix different executive functions to complete specific tasks. Therefore, no consistent conclusion has been reached on specific distinctions

between the three executive functions, although it is acknowledged that all are closely linked.

2.2.6 Executive Functions and Explanatory Coexistence

Currently, empirical studies concerning executive functions and explanatory coexistence have largely focused on the study of inhibition (see Mason et al., 2019 and Masson et al., 2014), with some recent studies beginning to focus on other executive function components. The following is a review of these studies.

Inhibition Contribution. Studies using neuroimaging tools have found that brain regions associated with inhibition are activated when participants resolve questions involving conflicts between naïve and science thinking (Butterfuss & Kendeou, 2020; Jin et al., 2019; Potvin et al., 2015; Zhu et al., 2019). For example, Potvin et al.'s (2020) study using functional magnetic resonance imaging (fMRI) on pre-university and university chemistry professors' ($N = 17$, $M_{age} = 43$ years) chemistry naïve thinking, using a text-based chemistry science statement task (Shtulman & Valcarcel, 2012). Results indicated that during incongruent trials these participants' scans showed a heightened activation in the left Ventrolateral Prefrontal Cortex (linked to inhibition) and the left Dorsolateral Prefrontal Cortex (associated with inhibition and working memory). Similar results in the activation of brain areas related to inhibition aligned with other fMRI studies using different tasks (Allaire-Duquette et al., 2019; Brault Foisy et al., 2015; Masson et al., 2014; Zhu et al., 2019), with brain regions and waves showing consistent activation in cases of processing conflicts between naïve and science thinking.

Behavioural studies have also found inhibition to be a key contributor to explanatory coexistence, based on analyses of accuracy and response time (Brookman-Byrne et al., 2018; Butterfuss & Kendeou, 2020; Potvin et al., 2020). Assessing the involvement of behavioural inhibition (Go/No-Go task) and interference inhibition (numerical Stroop task) when solving text-based congruent and

incongruent (between naïve and science thinking) statements task among adolescents ($N = 90$, ages 13- to 15- years), Brookman-Byrne et al. (2018) have found the ability to suppress naïve thinking to be related to both behavioural and interference inhibition. In their findings, better response inhibition correlated with longer reaction times, whereas better interference inhibition correlated with higher accuracy in suppressing naïve thinking. These results demonstrated that two forms of inhibition contribute distinctively to explanatory coexistence.

Other studies have found that the connection between inhibition and explanatory coexistence persists across various age ranges. Measuring the picture-based Buoyancy task (as used in the study by Potvin et al., 2015) among school aged children ($n = 35$, $M_{age} = 11.1$ years), adolescents ($n = 45$, $M_{age} = 17.1$ years) and adults ($n = 47$, $M_{age} = 24.1$ years), Brault Foisy et al. (2021) analysed inverse efficiency scores, considering both accuracy and response time, and observed negative priming effect to be more prevalent in incongruent trials where there was a greater conflict between naïve and science thinking (i.e., comparing a small but high-density ball with a large but low-density one). This effect changed with age: assessed via response time, it was found in children across all incongruent trials, but was only evident in incongruent trials with larger conflicts among adolescents; in adults, no significant negative priming was detected. Such findings suggest that the interference of naïve thinking decreases from childhood to adulthood.

In contrast to the above study, Keleman et al.'s (2013) research measuring the link between interference inhibition (Stroop Colour task) and the text-based Science Statement task showed no significant contribution of interference inhibition and explanatory coexistence. This study assessed the accuracy (in limited or unlimited time conditions) of task completion among physical scientists ($n = 80$, $M_{age} = 36$ years), college undergraduates ($n = 179$, $M_{age} = 19$ years) and humanities scholars ($n = 73$, $M_{age} = 39$ years). Its inconclusive results signal that inhibition's contributions to explanatory coexistence, especially among adults, requires more exploration.

In sum, results of most neuroimaging and behavioural studies demonstrate a connection between inhibition and explanatory coexistence across various age ranges,

even among science experts. Moreover, some studies indicate the distinct contributions of different inhibition abilities among children and adolescents. However, how inhibition is involved in the explanatory coexistence process among adults still needs further investigation.

Switching and Working Memory Contributions. Switching and working memory have also been found to be involved in explanatory coexistence. Zaitchik et al.'s (2014) research on 79 young children's ($M_{age} = 6$ years 6 months) performance on executive functions is informative here. This research project combined three tasks – the Hearts & Flowers task (all three executive subcomponents), the Flanker Fish task (inhibition) and the Colour Words task (working memory) – with vitalist biology interviews comprising three parts: an Animism Interview assessing the child's understanding and classification of life when given a certain image; a Death Interview assessing the child's understanding of what it means to die and whether dead people no longer have any bodily or mental functions; and a Body Parts interview to assess the child's understanding of the location and function of a series of body parts. Scores from the verbal and picture-based Animism interview represented children's understanding of naïve animism thinking. Results found the aggregated executive function accuracy score (excluding the Colour Words task) positively correlated with scores from the Animism interview, even after removing variance due to age and general cognitive ability. Thus, Zaitchik et al. (2014) concluded that overall executive functions contributed to explanatory coexistence. Similar results have been observed in later studies measuring the link between animism naïve thinking and executive functions among young children (e.g., Bascandziev et al., 2018; Grenell & Carlson, 2021; Tardiff et al., 2017).

Similar studies have been conducted on groups of older children. Vosniadou et al. (2018) have conducted research on 64 fourth graders ($M_{age} = 117.9$ months) and 69 sixth graders ($M_{age} = 135.7$ months) using two Stroop tasks specifically designed to assess inhibition and switching. Using the text-based Re-Categorization task to assess participants' conceptual-understanding and the picture-and-text-based Sentence-

Picture Verification task (similar to the science statement task) to assess explanatory coexistence, these researchers found that the response times and accuracy of the switching Stroop task, rather than the inhibition Stroop task, better predicted performance on Sentence-Picture Verification task, both for congruent and incongruent trials. These findings indicate that switching might be a better predictor of explanatory coexistence than inhibition.

Ellefson et al. (2024) have further found that switching may play a significant role on explanatory coexistence among adults ($N = 167$, $M_{age} = 26.38$ years). Via an online experiment to assess interference inhibition (the Flanker task), switching (the Figure Matching task), animism naïve thinking (the picture-based Animism task) and general naïve thinking (the text-based Science Statement task), path analysis found that switching's response time and accuracy positively predicted response time and accuracy for both naïve thinking tasks. But only the response time of interference inhibition predicted the response time of the Animism task. These links were similar across congruent and incongruent trials. Moreover, while response inhibition did not directly predict explanatory coexistence, it was correlated with switching. These findings indicated that switching plays a significant role in achieving both greater accuracy and faster responses on answering naïve thinking tasks, while interference inhibition contributed more to participants' ability to answer naïve thinking tasks more rapidly. Behavioural inhibition, meanwhile, was found to indirectly contribute to explanatory coexistence.

An fMRI study conducted by Dumontheil et al. (2022) has assessed both interference and behavioural inhibition (Go/No-Go and numerical Stroop task), as well as general naïve thinking (picture-and-text-based Science Statement task) among 38 adolescents (aged 11-to 15-years). This study found that the brain regions activated during the completion of the two inhibition and science statement task only partially overlap, yet the neural underpinnings for these overlaps were confined to the parietal cortex, an area usually linked to working memory and visuospatial attention. Thus, these neuroscientific explanations suggested that, in addition to inhibition, working memory might be involved in the process of suppressing naïve thinking.

More recently, Dumontheil et al. (2023) have assessed how inhibition (Go/No-Go task) and working memory (verbal working memory: Backwards Digit task and spatial working memory: Frankie task) contribute to children's ($N = 372$, aged 7-to-10 years old) explanatory coexistence reasoning (pictur-and-text-based Science Statement task) via accuracy score. Results indicate that only among older children (aged 9-to-10 years old) do all executive functions positively correlate with explanatory coexistence in the correlation analysis. However, this study found that only verbal working memory positively contributes to explanatory coexistence in multiple regression.

By considering more executive functions, the findings of these studies suggest that all executive function components might have different contributions to explanatory coexistence. Apart from inhibition, switching and working memory also seem important.

2.2.7 Research Gaps on the Link Between Executive Functions and Explanatory Coexistence

Based on the above review of the previous studies, several research gaps regarding the link between executive functions and explanatory coexistence can be identified:

The Distinct Role of Each Executive Function Components. The above studies suggest that executive functions play a predictive role in explanatory coexistence, despite the fact that evidence has often solely focused on inhibition or considered working memory and switching as combined scores. Most early studies placed a predominant emphasis on inhibition, with significant results observed when measuring inhibition in separate tasks (Brookman-Byrne et al., 2018; Butterfuss & Kendeou, 2020) or incorporating inhibition into measurements within naïve thinking tasks (e.g., Allaire-Duquette et al., 2019; Potvin et al., 2015). These studies were used to argue that the principal process in explanatory coexistence is to inhibit naïve

thinking. Other studies suggested that, when taking other executive functions into account, inhibition might not always serve as the sole or optimal predictor (Bascandziev et al., 2018; Ellefson et al., 2024; Dumontheil et al., 2022; Grenell & Carlson, 2021; Vosniadou et al., 2018). Both Vosniadou et al. (2018) and Ellefson et al. (2024) found switching to be a better predictor than inhibition when predicting performance on naïve thinking tasks, especially for incongruent trials. By distinguishing between the cognitive process of switching and inhibition, these findings might imply that, in the process of suppressing naïve thinking, there is not only a need to inhibit naïve thinking but also to activate later-acquired science thinking. Moreover, as suggested by Dumontheil et al. (2022; 2023), the active engagement of science thinking would require working memory to retain and compare two types of thinking, and making a choice between them.

Nevertheless, despite these recent studies on the involvement of other executive function components in explanatory coexistence, studies are still lacking that measure distinct executive function components separately, together with explanatory coexistence. While some studies have measured distinct executive function components in young children (Bascandziev et al., 2018; Tardiff et al., 2020; Zaitchik et al., 2014), these studies often resort to a combined executive functions score in their subsequent analyses, possibly because it is challenging to distinctly identify the three components of executive functions for young children. Thus, further investigations exploring the specific contributions of different executive functions to explanatory coexistence should be conducted.

Due to the diversity of approaches when measuring executive functions (Best & Miller, 2010; Diamond, 2013; Miyake et al., 2000), some executive function tasks may not exclusively measure a singular ability. For example, the Stroop task, which involves colour-word challenges, has often been used to assess either inhibition or switching (Monsell, 2003). Certain complex working memory span tasks might intertwine with academic learning capabilities (Cowan, 2014; Engle et al., 1999). Moreover, some of the inhibition tasks used in previous studies have been combined with naïve thinking tasks (Allaire-Duquette et al., 2019; Jin et al., 2019; Potvin et al.,

2015; Zhu et al., 2019), resulting in executive function contributions becoming hard to distinguish. Hence, it is crucial to choose executive function tasks that minimise the interference between different executive function components, and to assess executive functions and naïve thinking separately (see more details in Appendix A.3.2).

Identifying the Influence of Task Modality. Previous studies have presented naïve thinking tasks in varying modalities, with each modality having its own performance advantages. Visual stimuli offer distinct advantages in comprehension, as they presented information in a manner that can be readily understood and applied (Ward, 1994). In contrast, text stimuli allow learners to control their reading pace and decide if the repetition of sections or sentences is necessary (Lehmann & Seufert, 2020). Texts are powerful in conveying abstract knowledge and facilitating inference due to their descriptive nature, and allow new information to be directly interpreted from their representation (Bourisseau et al., 1965). Additionally, some studies on modality suggest that the processing of both pictorial and text stimuli require an involvement of prior knowledge, as prior knowledge mitigates the effects of limited external information, lower working memory capacity and deficits in propositional representation (Adams & Campbell, 1999; Mayer et al., 2001; Miller & Stine-Morrow, 1998; Moreno & Mayer, 1999). Pictorial stimuli in particular have been analysed more intensively in cases of complex content and limited prior knowledge among learners, when compared with text stimuli (Carney & Levin, 2002).

In previous studies, the Animism (Babai et al., 2010), Buoyancy (Allaire-Duquette, Bélanger, et al., 2019; Potvin et al., 2015a) and certain Circuit Diagram (Zhu et al., 2019) tasks have assessed individual naïve concepts through repetitive pictorial stimuli. These measurements did not require reading comprehension, and thus provided a more direct measure of participants' intuitive understanding of a single naïve concept. In contrast, Science Statement tasks (Brookman-Byrne et al., 2018; Potvin et al., 2015; Vosniadou et al., 2018) incorporated a broader range of naïve thinking concepts from various disciplines using text stimuli. These statement measurements required certain reading comprehension skills, aligning more with

everyday naïve thinking scenarios, and thus offered a comprehensive reflection of how people might respond when confronted with conflicts between naïve and science thinking.

However, it remains unclear whether executive function involvement is consistent across different task modalities. In the recent study by Ellefson et al. (2024), switching appeared to be predictive in both the picture-stimuli and sentence stimuli task, while inhibition seemed involved only in the picture-stimuli Animism task. Thus, executive functions' level of engagement might vary across different naïve thinking task modalities, meaning that further exploration is needed to understand how executive functions contribute to explanatory coexistence in various task modalities.

Identifying the Influence of Trial Types. Previous studies have not thoroughly distinguished the involvement of executive functions in congruent and incongruent trials. In most studies, although naïve thinking tasks have included stimuli for both congruent and incongruent trials, comparison between congruent and incongruent trials is not clear (Allaire-Duquette et al., 2019; Brault Foisy et al., 2015; Brookman-Byrne et al., 2018; Dumontheil et al., 2022; Jin et al., 2019; Mason et al., 2019; Potvin et al., 2015; Zhu et al., 2019). In congruent trials, where naïve and science thinking lead to the same conclusion, participants need only activate the relevant concept without suppression, while in incongruent trials, where naïve and science thinking are in conflict, it becomes necessary to inhibit naïve in favour of science thinking.

Typically, participants are less accurate and take longer to respond in incongruent situations (Shtulman & Valcarcel, 2012). If congruent trials require only the activation of a corresponding concept, then inhibition and switching may not be necessary to suppress naïve thinking, with working memory instead serving to recall relevant knowledge. Therefore, it is crucial to observe whether executive functions are consistently involved across the two trial types. Although studies by Vosniadou et al. (2018) and Ellefson et al. (2024) separated congruent and incongruent trials, both lacked an assessment of working memory, meaning that future research should aim to

provide a more comprehensive analysis distinguishing between congruent and incongruent trial types. This analysis should also include an assessment of all three executive functions.

Overviewing Different Task Measurement Metrics. Finally, the above studies have used different measurement metrics, which may contribute to inconsistent results. For example, some studies on adolescents or adults have used both accuracy and response time (Brookman-Byrne et al., 2018; Ellefson et al., 2014), or measured accuracy while controlling for total response time (Keleman et al., 2013), while others, on younger children, possibly out of consideration for the impact of interference on responses, have primarily focused on analysing task accuracy (Zaitchik et al. 2014). Such inconsistency in measurement approaches could be one reason for variability in earlier findings. This issue is particularly notable in the context of inhibition tasks, as most inhibition tests (i.e., the Stroop or Go/No-Go task) involve simple response button presses, so that adolescents and adults may be influenced by the speed-accuracy trade-off problem (Bruyer & Brysbaert, 2011). Therefore, it may be necessary to adopt a more standardised scoring method for these tests that provides a comprehensive measure of the relationship between cognitive abilities and explanatory coexistence.

2.3. General Cognitive Ability and Explanatory Coexistence

General cognitive ability, also referred to as intelligence or non-verbal reasoning in different studies (Brydges et al., 2012; Burkart et al., 2017; Conway et al., 2002; Ellefson et al., 2020; Kane et al., 2004), is the ability to reason, solve problems and think abstractly without task-specific knowledge; it is a core mechanism of human learning (Diamond, 2002; Kovacs & Conway, 2019; Krawczyk, 2012). The following sections use the term *general cognitive ability* throughout, in order to explore the concept's potential connections with executive functions, science learning and explanatory coexistence.

2.3.1 General Cognitive Ability, Executive Functions and Education Outcomes

General cognitive ability, also known as reasoning or intelligence, develops throughout childhood and adolescence and is associated with a range of other cognitive abilities, including executive functions, across developmental stages (Chierchia et al., 2019.; Friedman et al., 2006; Ger & Roebbers, 2023; McGrew, 2009). Interpretations of the relationship between general cognitive ability and executive functions varies across the research literature. While some studies regard general cognitive ability as a more fundamental cognitive function when compared to executive functions (Ackerman et al., 2005; Blair, 2006; Ellefson et al., 2020; Engelhardt et al., 2017; Xu et al., 2020), others find that executive functions predict general cognitive ability (Brydges et al., 2012; Burkart et al., 2017; Conway et al., 2002; Engel et al., 2010; Engelhardt et al., 2016). In a study involving eight executive functions and four general cognitive ability assessments, Brydges et al. (2012) found that executive functions positively predicted different forms of general cognitive ability among children. Meanwhile, Ellefson et al. (2020) have used computerised executive functions and general cognitive ability assessments to assess children in both England and Hong Kong, and found general cognitive ability has positively predicted executive functions in both sites. Furthermore, Dumontheil and Klingberg

(2012) have found enhancements in general cognitive ability during adolescence to correlate with concurrent improvements in executive functions. While there is no consensus as to the causal relationship between general cognitive ability and executive functions, current research suggests a positive relationship.

General cognitive ability has also been found to be a strong predictor of academic performance (Costa & Faria, 2018; Gottfredson, 2002; Kuncel et al., 2004; Lozano-Blasco et al., 2022). Two separate meta-analyses have summarised past research on general cognitive ability and its academic related impact. The first, by Lozano-Blasco et al. (2022), encompasses 27 studies and focuses on academic performance, measured via both test scores and school performance. The second, by Costa & Faria (2018), includes 46 studies and concentrates on academic achievement (similar to academic performance, with a focus on academic outcomes). Both meta-analyses find a significant predictive role for general cognitive ability on academic achievement or performance.

General cognitive ability has been found to be crucial for success in various science domains (Hegarty & Waller, 2004; Kell et al., 2013; Lynn & Mikk, 2009; Uttal et al., 2013). Lynn and Mikk (2009) have found that the national scores of 56 countries in reading comprehension, mathematical ability and science understanding show a correlation of .84 in terms of general cognitive ability ('national intelligence'). Another longitudinal study spanning nearly three decades and involving 320 adolescents found that individuals under the age of 13 with better general cognitive ability, assessed via mathematical or verbal reasoning abilities, have greater creative achievements at 38 years old, especially in the fields of business, healthcare, law, academia and STEM (Science, Technology, Engineering and Mathematics) subjects (Kell et al., 2013). These findings indicate that general cognitive ability, much like executive functions, positively impacts educational outcomes, particularly in science education. However, the mechanisms of this influence remained unclear. It is possible that general cognitive ability contributes to explanatory coexistence during scientific learning, thereby facilitating a more rapid understanding of scientific theories.

2.3.2 General Cognitive Ability and Explanatory Coexistence

While no studies have hitherto delved into the specific influence of general cognitive ability on explanatory coexistence, some studies focused on young children and teenagers have set general cognitive ability as a control variable before measuring the contribution of executive functions to explanatory coexistence, to mixed results (Brookman-Byrne et al., 2018; Dumontheil et al., 2022; Grenell & Carlson, 2021; Shtulman & Legare, 2020; Vosniadou, Pnevmatikos, Makris, et al., 2018; Zaitchik et al., 2014). Using the Peabody Picture Vocabulary Test to assess general cognitive ability among young children, Zaitchik et al. (2014) found no statistically significant link between general cognitive ability and naïve animistic thinking in interviews but did find it to be positively associated with the learning of other biological concepts. Grenell and Carlson (2021) used the Stanford–Binet Intelligence Scales to gauge general cognitive ability and found that verbal assessment of general cognitive ability positively predicted overall comprehension of scientific questions, including both congruent and incongruent conditions. These pieces of evidence suggest that the involvement of general cognitive ability during early childhood reflects an overall understanding of scientific knowledge and does not necessarily contribute to the suppression process of naïve thinking.

In the adolescent study, Vosniadou et al. (2018) used the Raven scale to measure general cognitive ability, and found it to be positively linked with response times for answering congruent science statements but not linked with accuracy and response times for incongruent statements. Brookman-Byrne et al. (2018) and Dumontheil et al. (2022), meanwhile, have both used the Wechsler Abbreviated Scale of Intelligence in two separate adolescent studies, with the former finding no link between general cognitive ability and explanatory coexistence and the latter finding slightly contrasting results. Both found that while general cognitive ability positively contributed to the accuracy of answers to incongruent trials, it negatively contributed to the response times of answers to incongruent trials. Both of these studies on adolescents also found that general cognitive ability had a positive effect on the

learning of science concepts, but its role in explanatory coexistence was not clear. Additionally, Dumontheil et al.'s (2022) findings suggest that adolescents with higher general cognitive ability might take longer to respond to questions requiring explanatory coexistence. Still, the specific link between general cognitive ability and explanatory coexistence still requires further investigation.

2.3.3 Research Gaps on the Link Between General Cognitive Ability and Explanatory Coexistence

Many studies indicate a strong association between general cognitive ability and executive functions, as well as general cognitive ability and science learning (Allaire-Duquette et al., 2019; Dumontheil et al., 2022; Jin et al., 2019; Masson et al., 2014; Vosniadou et al., 2018). However, the specific contributions of general cognitive ability to explanatory coexistence are unclear, as most previous studies have merely treated general cognitive ability as a control variable (Brookman-Byrne et al., 2018; Vosniadou et al., 2018). To date, several studies have suggested that general cognitive ability may be more associated with understanding science thinking (as measured by congruent trials) rather than suppressing naïve thinking (as measured by incongruent trials) (Grenell & Carlson, 2021; Vosniadou et al., 2018; Zaitchik et al., 2014b). In contrast, executive functions seem to play a role in both understanding science thinking (i.e., congruent trials or statements) and suppressing naïve science thinking (i.e., incongruent trials or statements). Thus, executive functions and general cognitive ability likely play different roles during the suppression process.

As with the above discussion on executive functions and explanatory coexistence, it is also worth investigating whether task modalities and measurements influence general cognitive ability's link with explanatory coexistence. However, unlike most executive function tasks that require a simple response, the completion of general cognitive tasks usually requires a longer time for thinking reasoning. Dumontheil et al. (2022) have demonstrated how results from adults' cognitive assessments (i.e., executive functions and general cognitive ability) are more likely to

be influenced by speed-accuracy trade-off (Bruyer & Brysbaert, 2011). Individuals with higher general cognitive abilities may be more willing to spend greater time answering explanatory coexistence tasks. Therefore, consistent with considerations for executive functions, it is essential to consider general cognitive ability's varying impacts on both accuracy and reaction time when considering its contribution to explanatory coexistence.

2.4. Science Education and Explanatory Coexistence

2.4.1 *How Science Education Contributes to the Explanatory Coexistence*

Neuroimaging and behavioural studies spanning from adolescents to the elderly adults have revealed that scientific competences and education levels can influence individuals' ability to suppress naïve thinking (Allaire-Duquette et al., 2019; Brault Foisy et al., 2015; Mason & Zaccoletti, 2021; Zhu et al., 2019). Foisy et al. (2015) have researched the area of 'naïve mechanics thinking' (i.e., "heavier objects falls faster") among adults, using an fMRI study to gauge responses to three types of stimuli: scientific (incongruent trials: two objects fall at the same speed), naïve (incongruent trials: heavy objects fall faster) and control (congruent trials: light objects fall faster). Adult participants ($N = 29$, 19 novice and 10 experts; $M_{age} = 23.5$ years) were asked to respond to stimuli by pressing on a keyboard. The researchers found no significant difference in brain activation between novices and experts when answering scientific and control trials, yet experts' dorsolateral prefrontal cortex and ventrolateral prefrontal cortex area were more active than those of novices when responding to naïve stimuli; these differences suggest experts' better suppression of naïve thinking when compared with novices. Furthermore, although there was no difference in anterior cingulate cortex activation between experts and novices, both groups detected the conflict between naïve beliefs about free falling objects and the science knowledge they had learned in school, even if novices had not yet reached the stage where they could entirely overcome their naïve thinking to provide a scientific answer.

Another study on 24 adolescents aged 16-to-18, conducted by Allaire-Duquette et al. (2019), tested participants' science competency levels by presenting them with a science statement task comprising 128 statements. fMRI results demonstrated that students with higher science competency exhibited stronger activation in the dorsolateral prefrontal cortex and ventrolateral prefrontal cortex area when answering incongruent statements, when compared with congruent statements. These findings

suggested that individuals with greater science competence deployed greater inhibition when faced with conflicting information. These results provided evidence from a neuroimaging perspective that experts, or individuals with high science competence, were better at explanatory coexistence than novices, or individuals with low science competence. Similar results have been found in behavioural studies among adults (Goldberg & Thompson-Schill, 2009; Shtulman & Harrington, 2016). For example, Shtulman and Harrington (2016) have used the science statements task to test explanatory coexistence in young adults ($n = 104$; aged 18- to 22- years) and older participants of various professions ($n = 45$; aged 51- to 87- years) and found responses to the Science Statement task to be predicted by statement subjects and participants' professions. Science professors were found to give more accurate, faster answers to the science statements when compared with humanities professors and non-professors, leading the researchers to conclude that extensive scientific education or exposure to accurate scientific knowledge might enhance explanatory coexistence performance. Thus, both neuroimaging and behavioural studies have highlighted the role of science education in improving individuals' performance on explanatory coexistence.

2.4.2 Science Education, Executive Functions and General Cognitive Ability

Although some studies have found the development of executive functions to be influenced by educational activities such as music (Bayanova et al., 2024; Jaschke et al., 2018) and physical education (Rudd et al., 2019), limited research has been conducted on the relationship between formal education, especially science education, and the development of executive functions (Shanmugan & Satterthwaite, 2016). Both cross-sectional and longitudinal data have found a positive link between education and individuals' general cognitive abilities (Cahan & Cohen, 1989). For example, Cahan and Cohen (1989) have used regression-discontinuity on a sample of approximately 11,000 students to show how education is a primary factor in boosting general cognitive ability scores with age, especially on verbal assessments. Moreover,

Falch and Sandgren Massih's (2011) analysis of a longitudinal dataset comprising 1,547 participants have found that, when controlling for general cognitive ability at 10 years old, education increases general cognitive ability by 2.9 to 3.5 points annually by age of 20. Thus, enhancement of general cognitive ability due to education surpasses the general effects of maturation. Nevertheless, research is still lacking on the link between science education in particular and the development of general cognitive ability.

While studies have found that the duration of education might influence executive functions and general cognitive ability, the causal relationship between education, executive functions and general cognitive ability has been difficult to disentangle (Deary & Johnson, 2010; Peng et al., 2018; Van Der Maas et al., 2006). Ritchie et al. (2015) have posited two ways that education may be linked with cognitive ability (including both executive functions and general cognitive ability). One possibility is that individuals with higher cognitive ability in childhood are more likely to remain in educational settings for longer periods of time, and gain higher qualifications. In such cases, the duration of receiving education does not have causal prediction on individual's original cognitive ability levels. On the other hand, education might improve cognitive ability by conferring on people the skills and knowledge needed to gain higher scores on different cognitive ability assessments.

Additionally, while the above studies indicate that duration of education is linked with improved general cognitive ability and executive functions, it remains uncertain whether these cognitive gains translate to specific learning outcomes, or how they manifest in such outcomes. Ritchie et al. (2013) have found that although education significantly enhances higher accuracy on general cognitive ability tests later in life, it does not affect participants' response time. These findings align with Shtulman and Harrington's (2016) research on explanatory coexistence, which suggests that increased scientific education improves accuracy scores and response times on explanatory coexistence tasks. Moreover, these researchers did not find response time differences between congruent and incongruent trials to significantly differ across science education. Thus, Shtulman and Harrington (2016) propose that

improvements in explanatory coexistence tasks may arise more from direct knowledge accumulation, especially in terms of accuracy. Consequently, extended exposure to scientific information, or simply mastering more scientific content, might enable individuals to answer questions more accurately and quickly without the engagement of cognitive abilities.

2.4.3 Research Gaps on the Link Between Science Education and Explanatory Coexistence

Current research on explanatory coexistence shows task performance to be greater among science experts when science education levels are divided into the basic categories of ‘novel’ and ‘expert’. However, it remains unclear whether science education levels influence explanatory coexistence performance directly through knowledge accumulation (i.e., by increasing the accuracy of explanatory coexistence-related questions) or affecting cognitive abilities such as executive functions and general cognitive ability (i.e., by shortening cognitive process response times), or by increasing overall response efficiency (i.e., making lower accuracy but in quick speed). Therefore, in addition to understanding the relationships between different measurements, one should also consider how to better measure science education in order to observe its connections with other variables. Existing research’s treatment of science education as a mere category does not allow for investigation into this type of education’s causal mechanisms, and may result in a loss of statistical power due to reduced variability, confining analysis to between-group comparisons (Altman & Royston, 2006). Science education must therefore be measured with continuous variables, which will both retain more nuanced information and encompass a range of scientific education exposures, while also allowing for a greater diversity of analytical methods (Ragland, 1992).

2.5. Research Gaps and Research Questions in Study 1

Several research gaps have been identified in the above literature review that require further exploration. To date, no studies have analysed the distinct contributions of various executive functions to explanatory coexistence comprehensively. Moreover, research exploring other potential influence factors, such as general cognitive ability and science education, and how these factors intersect with executive functions to influence explanatory coexistence, is lacking. It is also unknown as to whether these contributory factors consistently predict outcomes across different forms of naïve thinking tasks. Finally, deeper analysis and exploration into the differences between different trial types is needed. Study 1 of this PhD research therefore assesses adult participants using behavioural measurements to explore the relative contributions of executive functions, general cognitive ability and science education to explanatory coexistence.

Due to the developmental factors associated with children's responses, adult participants are used to mitigate potential inaccuracies (Ellefson et al., 2024) and offer a broader range of science education experiences. Behavioural measurements allow for the collection of statistical data on both accuracy and response time (Ellefson et al., 2017; Liu & Watanabe, 2012), allowing this research to explore the influence of the speed-accuracy trade-off effect during participant responses. In studies assessing cognitive abilities and explanatory coexistence, the most commonly used metrics are those that quantify individual variability and gauge individuals' abilities to perform tasks directly (Maris & Maas, 2012), as well as ones that measure participant response times (for correct trials), demonstrate the complexity of reasoning and reflect individuals' attention (De Boeck & Jeon, 2019; Brookman-Byrne et al., 2018; Grenell & Carlson, 2021; Vosniadou et al., 2018). Response time and accuracy both shed light on fundamental mechanisms of the cognitive process and highlight the intricate reasoning involved in task solving (Bruyer & Brysbaert, 2011; De Boeck & Jeon, 2019; Denes & Pizzamiglio, 2020), while response accuracy varies according to times allocated to participants for response (Heitz, 2014). Although precise description is

challenging, the ongoing discourse on speed-accuracy trade-off indicates that participants, especially adult participants, typically allocate greater time to their responses in order to enhance accuracy (Bruyer & Brysbaert, 2011; Liu & Watanabe, 2012; Zimmerman, 2011). Depending on their response patterns, participants may choose either liberal (rapid but error-prone) or conservative (emphasising accuracy over speed) responses and, with their initial level of baseline information, may reach different decision thresholds, making more or less informed decisions (Lubczyk et al., 2022). Based on the above considerations of accuracy and response times and following the methodology outlined by Ellefson et al. (2017), efficiency is adopted as the third metric for this study to account for the speed-accuracy trade-off effect of participant responses. A more detailed equation for efficiency calculation is provided in Appendix A.6.

The first study of this PhD research thus uses behavioural tasks and efficiency to holistically address the following research questions:

Research Question 1: What are the relative contributions of executive functions, science education and general cognitive ability to explanatory coexistence?

Research Question 2: Do task modalities (i.e., picture stimuli tasks versus sentence stimuli tasks) affect the involvement of executive functions, science education and general cognitive ability in explanatory coexistence?

Research Question 3: Do executive functions, science education and general cognitive ability contribute differently to congruent and incongruent trials in explanatory coexistence?

Chapter 3: A Cross-Cultural Exploration of Explanatory Coexistence

Chapter 2 discussed the possible involvement of executive functions, general cognitive ability and science education in explanatory coexistence. However, many of the studies reviewed in Chapter 2 have been overly reliant on participants from Western, Educated, Industrialised, Rich and Democratic (WEIRD) nations (Henrich et al., 2010). Recent research has argued that psychological and educational research that relies too heavily on WEIRD participants may be subject to biases, as findings thought to represent human universals may not generalise to individuals in other cultural contexts (Nzinga et al., 2018; Wong & Cowden, 2022). If a study's selected sample does not adequately represent the broader population, the study outcomes' research validity may be questioned (Oliveira & Baggs, 2023). This PhD therefore gathers data from participants who have grown up in different cultural environments, and in so doing will reveal whether previous findings from studies on WEIRD participants are generalisable to different cultures, and can thus give insight into how culture shapes human development from a cognitive perspective.

The 'Confucian-Socratic' framework is often utilised when delineating cultural influences between Eastern and Western contexts (Tweed & Lehman, 2002). East Asian regions (i.e., China, Japan, South Korea) influenced by Confucian thought are said to have developed a style of moral self-perfection that regards learning as a moral virtue. In contrast, Western regions (i.e., Europe, North America) have been strongly influenced by a Socratic tradition in which the values of rationalism and individualism promote a critical interrogation of the world (Huntington, 1993; J. Li, 2012; Peters, 2015). Despite the schematic utility of this model, *culture* should be viewed as a dynamic system rather than a static entity, and using data from multiple countries to represent a single culture may introduce bias (Xu et al., 2020). This study's data is therefore limited to sample sets from China and the United Kingdom, in order to prevent oversimplifying or stereotyping differences between the East and the West. Thus, in addition to examining participants from the United Kingdom in Study 1, this study further examines whether its United Kingdom results can be replicated among

Chinese participants. Study 2 is a follow-up study, replicating this study's research project within a Chinese context.

Chapter 3 summarises the current state of science learning and explanatory coexistence research in cross-cultural contexts. Findings from a cognitive perspective (executive functions and general cognitive ability), and explanatory coexistence in cross-cultural studies, are then discussed. Given that educational systems are largely determined by different countries' varying policies, differences in approach in science education in both China and the United Kingdom are summarised to provide a relevant background before proposing Study 2's two main research questions.

3.1 Science Learning and Explanatory Coexistence in Cross-Cultural Contexts

Variances in scientific assessment outcomes between Eastern and Western students have been observed in numerous global assessments. In the 2022 Programme for International Student Assessment (PISA), students aged 15 years old from Singapore (*Mscore* = 561), Japan (*Mscore* = 547) and Macau, China (*Mscore* = 543), ranked the top three positions in scientific achievement, significantly surpassing the global average (*Mscore* = 485) (OECD, 2022). Moreover, six of the top 10 regions in science scores were in East Asian countries. Another global evaluation, the 2019 Trends in International Mathematics and Science Study (TIMSS), found that the top three countries for international science benchmarks among pupils aged 9-to-11 years old were Singapore (intermediate benchmark = 93), South Korea (intermediate benchmark = 95) and Russia (intermediate benchmark = 92) (TIMSS, 2019). International assessment results therefore suggest that Asian students achieve higher scores on science assessments than their global peers, and it is worth exploring what educational processes contribute to such outcomes.

Previous studies have identified factors such as students' cognitive abilities, personalities, socio-economic status and learning environments (Boyle et al., 2020; Chmielewski, 2019; Gebauer et al., 2021; Kim, 2019; Kuwabara & Smith, 2012; Zaroff et al., 2014) as accounting for Asian students' high performance on international assessments. As

the mastery of scientific concepts is essential in the process of science learning, as mentioned in Chapter 2, identifying the underlying mechanism of explanatory coexistence, and whether this mechanism is shared or differs between the Chinese and United Kingdom contexts, could provide a more generalised understanding of influential factors in the science learning process. However, as previously stated, research exploring explanatory coexistence has mainly been conducted in Western countries such as Canada (Potvin et al., 2020), the United States (Zaitchik et al., 2014) and Greece (Vosniadou et al., 2018), making the current understanding of explanatory coexistence within cross-cultural contexts limited.

Two empirical studies have been conducted on influence factors within non-WEIRD countries, both from China. Zhu et al. (2019) conducted Event-Related Potential (ERP) studies with 27 university science major participants ($M_{age} = 22.8$ years), using the electrical naïve thinking task (Potvin et al., 2014). The researchers tested for P2, N2 and LPP components: the P2 component tests the indicative of expectation processes, comparing perceptual inputs with implicit memory, and exhibited larger amplitudes for congruent trials; the N2 component tests error detection and conflict monitoring, and increased for incongruent trials; the LPP component, which encompasses conscious error recognition, conflict resolution and response selection, was greater in incongruent trials, especially when stimuli were correct in naïve but incorrect in science thinking. Overall results suggested that naïve thinking is invoked automatically when evaluating the plausibility of incoming information, but later inhibited by subsequently retrieved scientific knowledge.

A similar study was conducted by Jin et al. (2019) comparing experts' ($n = 11$; $M_{age} = 24.1$ years) and novices' ($n = 9$; $M_{age} = 23.8$ years) respective performances of a speed-distance naïve thinking task. Participants were presented with an animation of two cars on parallel lanes in a grassy setting, both starting from the same point and arriving simultaneously at the same destination. The animations portrayed three scenarios: 1) Incongruent but correct, showing a scientifically accurate depiction that illustrated the car moving on a curved path moving faster than its counterpart on a straight path; 2) Congruent but false, a misconception-based depiction wherein speed is determined by distance, with the

car moving on a straight path travelling faster than its counterpart on a curved path; and 3) Incongruent and false, a misconception-based depiction wherein speed is determined by position, with both cars travelling at identical speeds. Functional Near-Infrared Spectroscopy (fNIRS) found different brain activation areas between novices and experts via the presentation of PE, with heightened PE denoting increased brain activation. Results found similar differences between experts and novices as those found in Western participants (Brault Foisy et al., 2015; Masson et al., 2014), while experts showed higher activation than novices during false trials. Notably, experts recorded significantly lower PE values in the right Inferior Frontal Gyrus (IFG) than novices. As the IFG encompasses regions associated with inhibition, this study's research results suggested that experts use greater inhibition to suppress their naïve thinking, while novices are less confused by naïve thinking interferences.

Findings from the above study not only reinforce the contrast between novices and experts in science education, previously mentioned in Chapter 2 (Brault Foisy et al., 2015; Masson et al., 2014), but also indicate that Chinese participants may adopt cognitive resources similar to their Western counterparts under the explanatory coexistence process. Moreover, both studies point to the possible involvement of inhibition during explanatory coexistence. Thus, these results collectively suggest that the underlying mechanism of explanatory coexistence might be similar, both in terms of inhibition and science education level.

3.2 Cognitive Ability and Academic Performance in Cross-Cultural Context

Cultures influence how individuals perceive, interpret and respond to various phenomena (Kelkar et al., 2013; Oliveira & Baggs, 2023; Pérez-Arce, 1999; Zaroff et al., 2014), shaping individuals' cognitive development from childhood to adulthood (Miller-Cotto et al., 2022). One fundamental assumption in cognitive research is that despite their different cultural backgrounds, humans rely on the same evolved neural structural foundations (Chua et al., 2005). Nevertheless, different individuals' life experiences can lead to variations in the cognitive developmental process, with Miller-Cotto et al. (2022) noting that the development of human cognitive structures often depends on an individual's surrounding environment. Significant influence factors such as poverty, adversity and stress can potentially impact an individual's cognitive development (Blair & Raver, 2016; Lipina & Evers, 2017; Miller-Cotto et al., 2022; Rutter, 1981).

The Individualist/Collectivist model has frequently been used to explain cognitive differences between cultures. East Asian cultures are thought to embrace collectivist perspectives under the influence of the Confucian tradition, which emphasises social harmony and the alignment of individual and communal goals. Consequently, individuals are expected to conform to their surrounding group and suppress their personal self-expression to some extent. In contrast, Western cultures accentuate individual autonomy, driven by the influence of Socratic tradition (Aikenhead & Jegede, 1999; Greenfield et al., 2006; Tobin & Hayashi, 2011). These cultural variances introduce societal norms and expectations into distinctive educational processes and societal activities. For instance, parents from collectivist cultures tend to be authoritarian and 'effort-oriented', while parents from individualistic cultures are more authoritative, encouraging independence and personal expression (Hsin & Xie, 2014). In school settings, teachers from collectivist cultures offer more proactive task-related guidance (Lan et al., 2011), while those from individualistic cultures focus more on encouraging personal

exploration. Meanwhile, neurological explanations suggest that cultural values not only modulate behaviour but also regulate cortical-level neural activity. For example, when compared to individuals from collectivist cultures, those from individualistic cultures have more extensive cortical areas for self-expression (Chiao et al. 2009; Zhu et al. 2007).

3.2.1 Executive Functions in a Cross-Cultural Context

Some evidence from executive function research indicates that children from East Asian countries develop certain executive function components earlier than their Western counterparts (Chen et al., 2010; Ellefson et al., 2017; Oh & Lewis, 2008; Xu et al., 2020). Oh and Lewis (2008) found 36-to-59-month-old Korean children's ($n = 76$) accuracy of inhibition and switching to be higher than that of their 36-to-62-month-old Western counterparts ($n = 43$), even reaching ceiling effects on these measures. Using computerised tasks to compare the executive functions of children aged 11-to-12 from mainland China ($n = 453$; $M_{age} = 11.89$ years), Hong Kong ($n = 371$; $M_{age} = 12.21$ years), and the United Kingdom ($n = 487$; $M_{age} = 11.91$ years) through the use of efficiency score (combined accuracy and response times), Xu et al. (2020) revealed that executive function development among Hong Kong children was two years ahead of their peers from the United Kingdom. While the performance difference between China and Hong Kong was significant, this difference was smaller than the difference observed when comparing either Asian region with the United Kingdom. These findings thus indicate that children from East Asian countries have earlier development in one or more executive function components.

However, some recent views suggest that using a uniform method to assess the cognitive abilities of children from different countries may not accurately reflect their true cognitive abilities, as demonstrated within their own countries (Gaskins & Alcalá, 2023; Hughes, 2023; Sabbagh et al., 2006; Scheidecker et al., 2022). For instance, in a study on delay of gratification, assessing self-regulation or inhibition, conducted among children aged 4-to-5 in Japan ($n = 80$) and the United States ($n =$

58), Niebaum and Munakata, 2023 found that Japanese children had a longer gratification delay (reaction time) for food as compared to gifts, while American children had a longer gratification delay for gifts as compared to food. This difference may arise from cultural emphases, with Japan having students wait before eating in school, while in the United States children might place a greater emphasis on waiting to open gifts. Thus, even when using the same tasks, cross-cultural research can be challenging due to participants from different countries being unfamiliar with the content related to task measurements (Kuwabara & Smith, 2012; Lamm et al., 2018; Munakata et al., 2020).

Adult studies have notably shown more inconclusive results when comparing executive function outcomes within cross-cultural context. Ellefson et al.'s (2017) cross-cultural, cross-generational study measuring children's and their parents' executive functions in Hong Kong and the United Kingdom using computerised assessments found that, while Hong Kong children performed better on all executive function measures in efficiency performance, parents' results ($n = 266$, $M_{age} = 43.84$ years in Hong Kong; $n = 275$, $M_{age} = 44.06$ years in the United Kingdom) showed no significant differences. Meanwhile, Kelkar et al. (2013) compared adults from India and the United States ($N = 42$, $M_{age} = 25.4$ years) and found that while American participants had greater accuracy on category switching tasks (assessing switching) and verbal fluency tests (assessing working memory), Indian participants responded faster on Trial Making tests (assessing working memory). Given these findings, cognitive skill differences observed during childhood may not fully extend into adulthood (Ellefson et al., 2017), with cultural influences on adult executive functions being more diversely reflected in the assessment metrics (i.e., accuracy, response time, efficiency) of different executive function tasks.

3.2.2 General Cognitive Ability in a Cross-Cultural Contexts

In the area of general cognitive ability research, numerous early studies from different countries have revealed mixed outcomes (see the meta-analysis by Costa &

Faria, 2018) and raised methodological issues. Some scholars argue that cultural influences in different regions cause variations in general cognitive ability's manifestations (Sternberg, 2008; Yiend et al., 2019), and note therefore that results from a single measurement may not exhibit universality in cross-cultural contexts. Other studies have noted that attempts to interpret measurement results across cultures using standards or research from different cultural backgrounds may result in the failure of a measurement's factorial invariance, potentially leading to misconceptions regarding individual or group scores (Chen et al., 2010; Wilson et al., 2023). Consequently, while there is currently no consistent and reliable conclusion that suggests substantial disparities in general cognitive ability among individuals from different regions, further research is still needed to gain a more nuanced understanding of cross-cultural cognitive ability.

3.2.3 Robust Links Between Cognitive Ability and Academic Performance in Cross-Cultural Context

Although the limitations of purely cognitive task measurement make comparisons of cognitive ability across different countries difficult, robust links between cognitive abilities and academic performance have been found across various countries (Ellefson et al., 2017; Georgiou et al., 2020; Lan et al., 2011; Thorell et al., 2013). Using accuracy analysis to measure executive functions (inhibition, working memory and attention control) and mathematical/reading abilities on young children in both China and the United States ($N = 258$, aged 3.4-to-5.5- years), Lan et al. (2011) found that Chinese children significantly outperformed their American counterparts in inhibition and attention control tasks. Nevertheless, no significant statistical difference was found among participants in working memory performance, or regarding the associations between executive function components and achievement. Similarly, Georgiou et al. (2020) tested executive functions and general cognitive ability in predicting mathematical and reading abilities in Chinese and Canadian children ($N = 301$; $M_{age} = 82.12$ months in Canadian, $M_{age} = 85.85$ months

in China), finding that while only working memory predicted these abilities among Canadian children, both working memory and inhibition played a role among Chinese children. These researchers therefore demonstrated that culture influences executive functions' contribution to academic achievement, with inhibition in particular playing a stronger role in Chinese children.

Critically, both of the aforementioned studies mainly relied on computer-based experiments to directly measure children's executive functions and academic performance. In contrast, Thorell et al. (2013) explored the relationship between executive functions and academic performance among children from Sweden, Spain, Iran and China (total $N = 482$, aged 6- to 11- years) using parent- and teacher-targeted questionnaires. Notably, results showed that Chinese parents generally rated their children as having more executive deficits when compared with parents in other countries, while inhibition and working memory were found to be related to academic achievement in all four countries. The inconsistency between parental rating and direct assessment of children's abilities in this case may further reflect the impact of culture on cognitive ability assessments, with the researchers arguing that Chinese culture's strong emphasis on self-regulation may result in a stricter behavioural control mind-set on the part of Chinese teachers and parents (Mann et al., 1992; Welsh, 2002).

More recently, Hughes (2023) has reviewed executive function measurements in cross-cultural research and argued that, due to the unity and diversity among different components of executive functions and the variety of executive function measures, it is critical to consider measurement invariance when evaluating the conceptual equivalence of executive function tasks for different cultural groups across countries. To assess measurement invariance in executive functions, Ellefson et al. (2020) have also explored whether these executive functions act as a mediator between socio-economic status and numeracy skills using efficiency score. Testing four executive functions (inhibition, working memory, switching and planning efficiency) in children from Hong Kong, China, and the United Kingdom ($n = 835$, aged 9-to-16 years) using structure equation modelling, these researchers found that

executive functions exhibited measurement invariance between the two countries, meaning that all four measures could stably constitute the latent variable of executive functions, significantly influencing numeracy development. Follow-up analysis found a stable predictive role of general cognitive ability on numeracy across different countries and genders. Thus, these findings suggest a positive role for executive functions in predicting academic achievement that may remain stable even when considering cross-cultural factors. In sum, despite variations in executive function test scores across different countries, executive functions and general cognitive ability are both essential for academic performance.

3.3 The Education System in the United Kingdom and China

A country's education system influences the form in which students receive science knowledge. Due to the differences in these systems between countries, this section briefly reviews the education systems relevant to this study, those of the United Kingdom and China, as well as the exposure time to science education these countries provide. Critically, the United Kingdom comprises four countries – England, Scotland, Wales, and Northern Ireland – and each show some variation in their educational systems. As the researcher and most of the participants in this study are based in England, this study primarily refers to the English education system. Additionally, given the wide geographical range in China and the differences in educational systems among Chinese provinces, this study pertains to the country's most commonly used education system.

3.3.1 The United Kingdom Education System

The United Kingdom's education system comprises four main sections: primary, secondary, further and higher education (GOV.UK, 2024). Primary school commences when students are aged four-to-five years old and lasts until age 11, when students transition to secondary school. This education stage lasts until age 16, when students take their General Certificate of Secondary Education (GCSE) examinations,

following which they have the option of extending their education through two pathways, either academic (typically a two-year course, such as Advanced Level, or A-Level, courses) or vocational (vocational qualifications such as the National Vocational Qualifications or the Business and Technology Education Council Qualifications, or BTECs). In higher education, undergraduate university courses generally span three-to-four years (four years in Scotland), masters' courses one-to-two years (with MPhil degrees taking about two years in most institutions) and Doctoral courses extending for three-to-four years.

Science education is introduced in primary school, where students must take science, mathematics and computing as part of the United Kingdom's compulsory national curriculum but are not assessed in these subjects through examinations. In secondary education, all students study English, Maths, Sciences, a Humanities discipline and a Modern Language, and schools provide a variety of optional subjects, either in the sciences or the humanities, which students may choose from according to their own interests. The GCSE examinations at the end of secondary education assess students' skills in 8-to-12 subjects, both compulsory (i.e., English, Maths, 2/3 Sciences, History/Geography, a Modern Language) and optional. No compulsory subjects are required after secondary education, so students are free to choose topics of their own interest or ones that they feel will most benefit their future studies. For example, students can choose to study three subjects for their A-Levels from an array of options, including English Language & Literature, Maths, Biology, Physics, Chemistry, History, Geography, Psychology, Economics, Art, Information Technology and Modern Foreign Languages (i.e., French, German, Spanish). Grades from these courses will be used to apply to universities. Some university subjects may require students to take A-Level courses, but this situation varies across universities. In sum, science education is only compulsory in primary and secondary education, and whether students decide to continue their science education after age 16 depends on their preference.

3.3.2 The Chinese Education System

The Chinese education system has four main sections, primary education, junior secondary education, senior secondary education and higher education (Ministry of Education of the People's Republic of China, 2022). Primary education begins for students at around age six and lasts until age 12, when students transition to junior secondary education, which continues until age 15. Together, primary and junior secondary education in China are often referred to as ‘nine-year compulsory education’. At the end of junior secondary education, students must take their high school entrance examination (*Zhongkao*). For their senior secondary education, students either receive high school (three year) or Secondary Vocational School (two year) education, after which they can take China’s college entrance examination (*Gaokao*) to determine the universities or colleges they may enter. Higher education generally spans three (college, career-oriented institutions) or four (university) years, while postgraduate courses last three years and doctoral courses four-to-five years.

Unlike the United Kingdom’s course selections, Chinese, Maths and English are the three compulsory subjects taught throughout China’s primary, junior secondary and senior secondary education. Science may be introduced in primary school, but varies across schools and is not compulsory. In junior high school, science education becomes compulsory. While science subjects may differ slightly across cities, biology, physics and chemistry are considered three compulsory subjects in all areas. In senior high school, students can choose one of two educational tracks, focused on the humanities (history, politics, and geography) or science (biology, physics, and chemistry) tracks. In the first year of senior high school, students must study all subjects in both tracks, but in their second year they are streamed into one of the two tracks, and for their university entrance examinations they can select either a humanities- or science-track-based exam. Students have traditionally entered university and chosen their majors according to their entrance examination score, yet reforms scheduled for 2025 are set to abolish the strict division between the humanities and sciences tracks. Instead, students will be allowed to freely choose

three subjects in which they excel from among the disciplines of biology, physics, chemistry, history, politics and geography for their examination.

To conclude, science education is only compulsory for every student in China's junior secondary education and first year of senior secondary education. Whether students continue to study science during their second year of senior secondary education or decide to choose a secondary vocational school education (career-oriented education), ultimately depends on their preference.

3.3.3 A Comparison of the United Kingdom and Chinese Education Systems

Science education is important in both the United Kingdom and China, although the United Kingdom system delivers this education in a more flexible manner (for example, there are more A-level subjects from which students can choose). Still, China's science track includes a course of compulsory subjects such as biology, chemistry and physics, which may offer more comprehensive science knowledge than that of the United Kingdom. For adults from the United Kingdom, science education levels may be more related to specific subject-related interests, while the science education levels of Chinese adults may be more related to their science or humanities education track. Consequently, it is difficult to directly separate science and non-science students when conducting cross-cultural comparisons of different education systems.

3.4 Research Gaps and Research Questions in Study 2

Individuals from different countries currently exhibit varying performances on standardised assessments such as PISA and TIMSS, and the reasons underlying these differences require further explanation. While learning outcomes may be determined by various factors, as mentioned in Chapter 2, explanatory coexistence has been found to be crucial to science learning outcomes and cognitive abilities, and science education may greatly contribute to explanatory coexistence.

The previous section reviewed studies that addressed science learning, cognitive abilities and science education systems from a cross-cultural perspective. Although some cross-cultural studies of executive functions and general cognitive ability have found that children in Eastern countries may develop executive functions faster, the use of the same assessments across countries could introduce bias. Research on adults has found different results to the simpler comparisons of accuracy used in previous child studies. Additionally, the differences observed among adults by Ellefson et al. (2024) and Kelkar et al. (2013) when using efficiency as a measurement metric, as compared with accuracy or response time, suggest that there might be varying response strategies across different countries, and considering a single metrics might not be enough. Overall, when viewed in light of cross-cultural perspectives, more complex relationships between the task measurement results of cognitive abilities may be at work. Therefore, it is essential to observe an efficiency score during computerised tasks that considers both accuracy and response times collectively.

While the positive link between cognitive abilities and academic performance seems robust, the variety of educational systems and cultural differences across different countries make identifying the most significant influence factors of explanatory coexistence for individuals of different cultural backgrounds uncertain. To better understand whether there is a common underlying mechanism during the process of explanatory coexistence reasoning, a key research question is whether the link between executive functions, general cognitive ability, science education and explanatory coexistence can be generalised across countries.

Thus, Study 2 addresses the following two research questions by applying the Chinese version of the experimental materials used in Study 1:

Research Question 4: Are the response patterns of executive functions, general cognitive ability and explanatory coexistence similar among adult participants from the United Kingdom and China?

Research Question 5: Is the relationship between science education, executive functions, general cognitive ability and explanatory coexistence robust across the United Kingdom and China?

Chapter 4 Method

4.1 Participants

The current research (both Studies 1 and 2) involve adult participants from both China and the United Kingdom. In Study 1, a total of 262 participants ($M_{age} = 23.45$ years, $SD_{age} = 6.53$) from the United Kingdom were used in the final analysis, with the majority providing demographic information. More detailed demographic information is presented in Table 4.1. Study 2 added a total of 234 participants ($M_{age} = 22.88$ years, $SD_{age} = 3.55$) from China, with the majority providing demographic information. More detailed demographic information is presented in Table 4.2.

Despite changes to the research plan outlined in the first preregistration (see Appendix A.8), no additional participants were recruited. For both the first and second pre-registrations, a Sample Size Calculator (Soper, 2024) with the following parameters was used: anticipated effect size = 0.2 (small effect), desired statistical power = 0.8, latent variables = 2, observed variables = 7 and a probability level = 0.05 to calculate minimum sample size. Results indicated that a minimum of 223 participants would be necessary to detect the anticipated effect size. While general cognitive ability and science education were included in the second pre-registration (with latent variables = 2, observed variables = 9), the minimum required sample size remained 223 when using the same sample size calculator. Therefore, I did not collect additional data. The final path analysis included participants from China ($n = 234$) and the United Kingdom ($n = 267$), satisfying both the minimum requirement of 200 participants (Bandalos, 2014) and the rule of thumb of five participants per parameter (Wang & Rhemtulla, 2021). In addition, the current sample size also aligns with the requirements outlined in Appendix A, Table A14 for conducting later analyses (correlation and path analysis).

Table 4.1*Demographic information for participants from the United Kingdom*

Category		Count	Percentage
Gender ($n = 262$)			
	Male	128	48%
	Female	124	48%
	non-binary	3	1%
	missing	7	3%
First Language ($n = 262$)			
	English	242	92%
	non-English	3	1%
	missing	18	7%
Range		<i>Mean</i>	<i>SD</i>
Age	18-77	23.45	6.53
$(n = 257, \text{missing} = 5)$			
Education Level	10-24	17.21	2.81
$(n = 257, \text{missing} = 5)$			
Highest Biology Education	0-22	10.86	3.30
$(n = 257, \text{missing} = 5)$			
Highest Biology Education	0-14	10.00	2.90
$(n = 257, \text{missing} = 5)$			

Table 4.2*Demographic information for participants from the United Kingdom*

Category		Count	Percentage
Gender ($n = 234$)			
	Male	109	47%
	Female	98	42%
	non-binary	0	0%
	missing	27	11%
First Language ($n = 234$)			
	Chinese	223	95%
	non-Chinese	0	0%
	missing	11	5%
	Range	<i>Mean</i>	<i>SD</i>
Age	18-35	22.88	3.55
($n = 226$, missing = 8)			
Education Level	14-24	18.73	3.34
($n = 200$, missing = 34)			
Highest Biology Education	0-22	11.15	2.42
($n = 227$, missing = 7)			
Highest Biology Education	0-22	11.57	4.11
($n = 227$, missing = 7)			

4.2 Methods

4.2.1 Demographic Information Questionnaire

The questionnaire included items about age, gender, and educational background. Specifically, to assess participants' experience with science education, a question was included regarding their highest level of education in science subjects, such as biology, physics, and two additional subjects, which participants were required to fill out themselves. For all questions related to education level, options ranged from primary school to completed PhD degrees, in alignment with the educational systems of the United Kingdom and China (see Appendix A, Table A2).

4.2.2 Computerised Behaviour Tasks

Behavioural Inhibition task – Stop Signal task. The Stop Signal task was adapted from the study conducted by Ellefson et al. (2017). Using a football field as the background, the task featured an image of a football placed in either the left half or the right half of the field. Participants were instructed to press the left arrow key if the football appeared in the left half of the field and the right arrow key if it appeared in the right half. They were also told to stop pressing any keys if they heard a whistle during any of the trials. There was a short time interval (250 ms) between the visual presentation of the football image and the sound of the whistle. This interval increased by 50 ms in the next trial if the participant made a correct response, and it decreased by 50 ms if the response was incorrect. There were 108 trials in total, presented in three blocks in random order, with a fixation trial in between experimental trials. The whistle trials accounted for 20% of the total trials, with equal distribution across each block. A central fixation cross lasting for 500 ms was shown between trials, and no time restriction was imposed on this task.

Interference Inhibition task – Flanker task. A revised version of the classic Flanker task (Ellefson et al., 2024; Eriksen & Eriksen, 1974; McDermott et al., 2007) was used. There were 108 trials in total, presented in three blocks (36 trials each) in random order with a fixation trial between experimental trials. In a typical trial, a row of five fish were presented. The target fish was in the middle, being surrounded by four non-targets, or “flanked” fish that were swimming in the same or opposite direction as the target one. The target fish swam towards the right direction in half of the trials, in which case participants should press the right arrow key. The other half of trials featured the target fish swimming left, and therefore the left arrow key should be pressed in this case. There were three types of trials: 1) congruent trials in which flanked fish were in the same direction as the target fish; 2) simple incongruent trials

in which the flanked fish were in opposite direction from the target fish; and 3) complex congruent trials in which the four flanked fish were in different directions, irrespective of the direction in which the target fish swam. There were 12 trials that belonged to each of the three trial types in each of the three blocks. A central fixation cross lasting for 500ms was shown in between trials, and no time restriction was set in this task.

Switching task – Figure Matching task. A simplified version of the traditional task-switching paradigm (Rogers & Monsell, 1995) was used to assess switching. Modifications to the task were made to accommodate the diversity in age and nationality among participants (Ellefson et al., 2006, 2024). Each trial involved four simultaneous events. In each trial, a central target stimulus was presented, with its shape being either a triangle or a circle and its colour either blue or red. Meanwhile, two small figures were displayed at the bottom left and right sides of the screen, respectively. Participants were instructed to match the two small figures with the target stimulus based on whether they shared the same shape or colour. A response selecting the figure on the lower left could be made by pressing the left-arrow key, while the choice of the figure on the lower right corresponded to the right-arrow key. There were four blocks in total: two pure blocks, where either shape or colour was consistently matched, and two mixed blocks, in which the instruction for figure-matching alternated every two trials (e.g., colour-colour-shape-shape-colour-colour-shape-shape). One mixed block began with two colour trials, while the other began with two shape trials. In total, there were 128 trials, randomised across the four blocks, with a fixation trial between experimental trials. The sequence of these four blocks was randomly assigned to participants. A central fixation cross, lasting for 500 ms, was shown between trials.

Working Memory task – Corsi Block task. A modified version of the original Corsi block task was used (Corsi, 1972; Ellefson et al., 2017). This task was divided into two blocks: the forward block and the backward block. In both blocks, a total of nine boxes were displayed on the screen, with some or all of them lighting up in a predetermined order. Participants were required to memorise the order in which the boxes lit up. In the forward block, they were instructed to click on the boxes in the original order, whereas in the backward block, they had to click in the reverse order. The number of boxes that lit up increased incrementally from two to nine across trials. Each trial was presented twice, resulting in a total of 18 forward items and 18 backward items. Trials in the forward block assessed short-term memory, while those in the backward block indexed working memory. In accordance with the research aims, only the results from the backward trials were included in the analyses.

General Cognitive Ability task – Matrix Reasoning task. The MaRs-IB (Matrix Reasoning Item Bank) task was used to assess participants' general cognitive ability (Chierchia et al., 2019). This task is analogous to Raven's Progressive Matrices, a behavioural measure commonly employed in studies on executive functions. The task consisted of 80 trials presented in random order. Each trial displayed a 3x3 matrix comprising nine cells, eight of which contained abstract shapes, while one cell, located in the bottom right-hand corner, was empty. Participants were required to select the shape that best fit the empty cell from four given options. To determine the correct shape, participants needed to identify universal patterns in the shapes already present in the matrix, which varied in colour, shape, position, and size. A maximum of 30 seconds was allotted for each response.

Naïve Thinking task - Animism task. The computerised Animism task, adapted from the Living and Non-Living Classification Test (Babai et al., 2010), was designed to evaluate explanatory coexistence regarding animism concepts in biology. In each trial, a grayscale drawing of an object appeared at the centre of the screen, and participants were required to classify the object as either living or non-living by pressing the left arrow key (for living) or the right arrow key (for non-living). The task included 56 objects equally divided into eight blocks, with seven objects in each block. The objects were categorised into four groups: animals (living and moving, such as mammals and insects), plants (living and non-moving, such as trees and flowers), inanimate static objects (non-living and not moving, such as tools like hammers and screwdrivers, furniture, and immobile natural landscape components like mountains), and inanimate dynamic objects (non-living but capable of motion, such as vehicles like motorcycles and tractors, and celestial bodies like stars and the moon). Each drawing was presented once in a randomised order within its block, and the task consisted of three blocks, each containing 56 trials, for a total of 168 trials. A central fixation cross, lasting 500 ms, was shown between trials, and no time restriction was imposed on participants for this task.

Naïve Thinking task - Buoyancy task. The Buoyancy task, developed by Potvin et al. (2015), was designed to evaluate the explanatory coexistence of buoyancy concepts in physics. In each trial, two balls of different sizes and materials were displayed on the left and right sides of the screen, and participants were asked to select the ball most likely to sink if placed in a water tank. Participants pressed the left arrow key if they believed the ball on the left side was more likely to sink, and the right arrow key if they believed the ball on the right side was more likely to sink. The balls varied in density (high: steel, medium: wood, low: plastic foam) and size (large, medium, small). Trials were categorised into five conditions: (1) very intuitive, where a large, high-density ball was paired with a small, low-density ball; (2) intuitive, where a large, high-density ball was paired with a medium, low-density ball or a medium, high-density ball was paired with a small, low-density ball; (3) neutral, where size and density levels were similar for both balls (e.g., a medium, high-density ball compared with a medium, low-density ball); (4) counter-intuitive, where a large, low-density ball was paired with a medium, high-density ball or a medium, low-density ball was paired with a small, high-density ball; and (5) very counter-intuitive, where a large, low-density ball was paired with a small, high-density ball. The task consisted of 162 trials evenly divided into three blocks, with 54 trials per block, presented in random order. Each block included 18 intuitive trials (six of which were very intuitive), 18 counter-intuitive trials (six of which were very counter-intuitive), and 18 neutral trials. A central fixation cross lasting 500 ms was displayed between trials, and participants were allowed a maximum of 5,000 ms to respond to each trial.

Naïve Thinking task - Science Statement task. The task was designed to measure general science knowledge and consisted of 64 statements selected from a statement bank on natural phenomena (Shtulman & Valcarcel, 2012). In each trial, a statement was presented at the center of the screen, and participants were instructed to judge whether the statement was true or false. They pressed the left arrow key if the statement was true and the right arrow key if it was false. The trials included four types of statements: (1) statements true under both naïve and scientific theories (e.g., "Rocks are composed of matter"), (2) statements false under both naïve and scientific theories (e.g., "Numbers are composed of matter"), (3) statements true under naïve theories but false under scientific theories (e.g., "Fire is composed of matter"), and (4) statements true under scientific theories but false under naïve theories (e.g., "Air is composed of matter"). The statements were evenly divided into eight sets of physics-related topics (mass, weight, density, force, velocity, gravity, acceleration, and stars) and eight sets of biology-related topics (death, reproduction, metabolism, microbes, kinship, contamination, mutation, and life). All statement trials were presented in a single block in random order, with a fixation trial between experimental trials. A central fixation cross lasting 500 ms was shown between trials, and no time restriction was imposed on participants for this task. The complete set of statements is provided in Appendix G.

4.3 Procedure

Research consent form, information sheet, and a demographic information questionnaire were provided to participants via Qualtrics (see Appendices C & D & E). After obtaining the demographic information from participants, a link to a battery of computerised performance tasks on PsychoPy was provided to them (Peirce et al., 2019). Referencing the logistics for the wider projects, Thinking Games (Ellefson et al., 2017), I implemented a mixed sequence of task difficulty in subsequent task designs. This approach ensures that all participants can more effectively persist in completing all tasks while remaining engaged (Ellefson et al., 2024). The tasks were completed in the following order : 1) the Stop Signal task, 2) the Figure Matching task, 3) the Flanker task, 4) the Corsi Block task, 5) Physics task, 6) Scientific Statement task, 7) the Matrix Reasoning task, and 8) the Animism task, with each task taking about 5–15 mins (overall around 90mins). During the tasks (apart from the Matrix Reasoning task), participants were instructed to respond as fast as possible while keeping accurate. Accuracy and response time data were recorded.

Participant recruitment involved reaching out to potential participants from universities from the United Kingdom and China via personal or institutional contacts via bilingual posters detailing the research. Funded by Homerton Research Grant, I provided 5GBP amazon vouchers in the United Kingdom and equivalent WeChat envelopes in China to boost participation rates and maintain motivation (Bentley & Thacker, 2002; Largent et al., 2014).

4.4 Statistical Analyses Procedures

Building on the detailed explanations of data analytic strategies outlined in Chapter 4, I organised the results of my PhD research into two studies. Study 1 focuses on analysing the data collected from participants in the United Kingdom. Study 2 extends this analysis by incorporating data collected from participants in

China and comparing the results between the two datasets. For both studies, I followed a consistent approach, conducting preliminary and main analyses using R (R Core Team, 2020). This section provides a detailed account of the specific procedures and R packages used in the analysis. The R scripts for all analyses are available on OSF: <https://osf.io/75kr8/>.

4.4.1 Preliminary Analysis

Before conducting formal data analysis to address my research questions, I performed a series of preliminary analyses on the raw data collected from the United Kingdom and China. After data wrangling, I first examined demographic information, including the mean, standard deviation, and normality of all computerised tasks. Next, I conducted age adjustment checks and converted raw scores into z-scores. In the third step, I assessed differences between congruent and incongruent trials across the three naïve thinking tasks. Following this, I checked for multicollinearity and homoscedasticity to ensure the data met the assumptions required for regression analysis. Finally, I conducted confirmatory factor analysis to evaluate the possibility of creating latent variables for executive functions and explanatory coexistence. However, the results indicated that structural equation modelling could not be used in subsequent analyses. Table 4.3 summarises the analyses performed in each step of the preliminary analysis, along with the criteria for the selected methods and the R packages used.

Table 4.3*Details of Steps in Preliminary Analysis*

Steps	Analysis	Criteria or Formula	R packages
Data Wrangling and Descriptive Statistics	Calculate accuracy, response Time and efficiency for each computerised task for every participant	Using formula presented in Appendix A.6, Table A9.	<i>tidyverse</i> package (Wickham et al., 2019) <i>Base R</i> (R Core Team, 2021)
	Calculate science education score	Highest Biology Education + Highest Physics Education + Optional Study in Biology or Physics/2(or 3 if there is optional courses).	<i>tidyverse</i> package (Wickham et al., 2019) <i>Base R</i> (R Core Team, 2021)
	Detect outliers	Data outside +3/-3 standard deviation are regarded as outliers, checking using box plot.	<i>tidyverse</i> package (Wickham et al., 2019)
	Detect missing values	Missing value in demographic information (As mentioned in Appendix A.5, all PsychoPy tasks were completed).	<i>Base R</i> (R Core Team, 2021) <i>ggplot2</i> package (Wickham, 2016)
	Descriptive statistics of mean, standardised deviation and normality of raw data	Significant result of Shapiro–Wilk test show non-normal distribution.	<i>Base R</i> (R Core Team, 2021)

Table 4.3 continue (1)

Age adjustment and z score conversion	Calculated residuals from a linear, quadratic or cubic regression with age, comparing results of significance.	Significant result of linear, quadratic or cubic regression show the impact of age	<i>Base R</i> (R Core Team, 2021)
	Z score conversion	Convert age adjusted scores into z scores	<i>Scales</i> package (Wickham et al., 2023)
	Normality checking	Significant result of Shapiro–Wilk test indicates non-normal result	<i>Base R</i> (R Core Team, 2021)
Naïve thinking task check	Check individual and overall discrepancies in accuracy and response time between congruent and incongruent trials	Significant result in mixed-effect model shows differences in performance of congruent and incongruent trials	<i>lmer4</i> package (Bates et al, 2015)
Multicollinearity and homoscedasticity check	Checking Multicollinearity via Variance Inflation Factor (VIF) and Tolerance	Variance Inflation Factor (VIF) < 10 and Tolerance > 0.2 indicate no Multicollinearity issues	<i>olsrr</i> package (Hebbali, 2024)
	Checking Homoscedasticity via Goldfeld-Quandt Test	Non-significant Goldfeld-Quandt Test results indicate data homoscedasticity	<i>lmtest</i> package (Zeileis & Hothorn, 2002)

Table 4.3 continue (2)			
Confirmatory Factor Analysis	Evaluate whether data could form latent variables of executive functions, and explanatory coexistence. General cognitive ability and science education were measured as single indicator latent variables	Goodness-of-fit index (see Appendix A Table A16)	<i>lavaan</i> package (Rosseel, 2021)
	Checking the link between each latent variable	Check significance and beta of covariances between each latent variable.	<i>lavaan</i> package (Rosseel, 2021)

4.4.2 Main Analysis

In Study 1, the main analysis consisted of two parts: correlation analysis and path analysis. In Study 2, prior to conducting correlation and path analyses, I presented scatterplots and performed a multigroup analysis to determine whether the data collected from the United Kingdom and China should be analysed separately. For the path analysis models in both studies, I considered the potential for just-fitting issues in the full model. As described in Appendix A.9, I addressed this by removing one non-significant path across all models, creating a full model with one degree of freedom to facilitate both path analysis and multigroup analysis (Ellefson et al., 2024). All main analyses were performed using z-scores of age-adjusted data. Table 4.4 provides an overview of the analyses conducted at each step of the main analysis, along with the criteria for the selected methods and the R packages used.

Table 4.4*Details of Steps in Main Analysis*

Steps	Analysis	Criteria or Formula	R packages
Correlation	Pearson's correlation to explore the correlation between each variable.	$r = .50$ large effect size $r = .30$ medium effect size $r = .10$ small effect size	<i>Hmisc</i> package (Harrell, 2023)
Path Analysis	Explore the complex relationship between dependent variables (executive functions, general cognitive ability & science education) and independent variables (explanatory coexistence).	Goodness-of-fit index (see Appendix A.9.1 Table A.16) $\beta = .50$ large effect size $\beta = .30$ medium effect size $\beta = .10$ small effect size	<i>lavaan</i> package (Rosseel, 2021)
Scatterplots	Visualise response pattern of accuracy and response time of each task across two countries.	Observing the response pattern	<i>ggplot2</i> package (Wickham, 2016)
Multigroup analysis	Explore whether the complex links between all variables in data collected from the United Kingdom is similar with data collected from China.	Non-significant chi-square probability indicates invariance between countries.	<i>lavaan</i> (Rosseel, 2021) <i>car</i> package (Fox & Weisberg, 2019)

Chapter 5 Results and Discussion for Study 1: Analyses Using the United Kingdom Data

In this chapter, I analyse the data collected from the United Kingdom for Study 1 and discuss the results. The analysis primarily summarises findings using efficiency metrics. The chapter aims to address the three research questions proposed in Chapter 2: (1) What are the relative contributions of executive functions, science education, and general cognitive ability to explanatory coexistence? (2) Do task modalities (i.e., pictorial stimuli versus text stimuli) affect the involvement of executive functions, science education, and general cognitive ability in explanatory coexistence? and (3) Do executive functions, science education, and general cognitive ability contribute differently to congruent and incongruent trials in explanatory coexistence? As discussed in Chapter 3, correlation and path analyses were conducted to examine the overall trials, congruent trials, and incongruent trials in the naïve thinking tasks to collectively address these questions. This chapter begins with a summary of the preliminary analysis, followed by the presentation of the main analyses, namely correlation and path analyses. A discussion of the results in relation to the three research questions is then provided. Analyses based on accuracy and response time are included in Appendix H and Appendix I. Research questions related to Study 2, which involve comparisons between the United Kingdom and China, are addressed in Chapter 6.

5.1 Preliminary Analysis (the United Kingdom)

All results from preliminary analysis on efficiency for the United Kingdom data were presented in Appendix J. Descriptive statistics analysis was conducted to summarise the mean, standard deviation, minimum value and maximum value of the efficiency data. Then, I conducted age adjustment and results indicated that age significantly influenced efficiency of executive functions (curvilinear). After conducting age adjustment and *z* score conversion, the Shapiro-Wilk test results indicated that most of the efficiency adhered to a non-normal distribution. The congruent and incongruent trials for the three naïve thinking tasks differed in the expected direction for efficiency based on raw score analysis: efficiency for the Science Statement, the Animism, and the Buoyancy tasks in congruent trials was higher than in incongruent trials. Significant large effect size correlations were found between congruent and incongruent trials in the Animism task ($r(260) = .82, p < .001$), the Buoyancy task ($r(260) = .75, p < .001$), and the Science Statement task ($r(260) = .68, p < .001$).

Multicollinearity and homoscedasticity results showed no multicollinearity and no heteroskedasticity problems in all efficiency models. Finally, I conducted a confirmatory factor analysis to test whether the efficiency data could form latent variables in structural equation modelling. Model fits of overall trials ($\chi^2(23) = 43.14, p = .004, CFI = .95, TLI = .91, RMSEA = .07, SRMR = .05$), congruent trials ($\chi^2(23) = 36.09, p = .040, CFI = .97, TLI = .95, RMSEA = .05, SRMR = .04$), and incongruent trials ($\chi^2(23) = 31.82, p = .104, CFI = .98, TLI = .96, RMSEA = .04, SRMR = .04$) all suggested good model fit to proposed model using the efficiency. However, collinearity problem was found between latent variable executive functions and explanatory coexistence, with overall trials ($\beta = 1.00, p < .001$), congruent trials ($\beta = 1.00, p < .001$), and incongruent trials ($\beta = .97, p = .006$), indicating the need to use path analysis to analyse all variables separately.

5.2 Correlation Analysis (the United Kingdom)

Correlation analyses results using efficiency were presented in Table 5.1. As the specific r and significance were presented in the table, the current section only summarises correlation results.

Science Education score showed significant correlation with small effect size with the Flanker task. There were significant intercorrelations with small to large effect size within four executive function tasks, and intercorrelations with small to large effect size within three naïve thinking tasks. However, all other correlations are not statistically significant.

In overall trials, the Matrix Reasoning task showed significant correlation with small effect size with the Science Statement task. The Stop Signal task showed significant correlation with small to medium effect size with all three naïve thinking tasks. The Figure Matching task and the Flanker task indicated significant correlation with small to large effect size with all three naïve thinking tasks. The Corsi Block task showed correlation with medium effect size with the Animism task and the Buoyancy task.

In congruent trials, the Matrix Reasoning task showed significant correlation with small effect size with the Science Statement task. The Stop Signal task, the Figure Matching task, and the Flanker task indicated significant correlation with small to large effect size with all three naïve thinking tasks. The Corsi Block task showed significant correlation with medium effect size with the Animism task and the Buoyancy task.

In incongruent trials, Science Education showed significant correlation with small effect size with the Animism task. The Matrix Reasoning task showed significant correlation with small effect size with the Science Statement task. The Stop Signal task showed significant correlation with small to medium effect size with the Animism task and the Buoyancy task. The Figure Matching task indicated

significant correlation with small to large effect size with all three naïve thinking tasks. The Flanker task and the Corsi Block task indicated significant correlation with small to medium effect size with the Animism task and the Buoyancy task.

Table 5.1*Correlation Analysis on Efficiency (the United Kingdom)*

Variables	1	2	3	4	5	6	7	8
<i>Overall Trials</i>								
1. Science Education	-							
2. Matrix Reasoning	-.05							
3. Stop Signal	.10	-.01	-					
4. Figure Matching	.07	-.03	.29***	-				
5. Flanker	.12*	-.09	.39***	.55***	-			
6. Corsi Block	.10	.02	.25***	.35***	.43***	-		
7. Science Statement	.05	.19**	.16**	.34***	.15*	-.01		
8. Animism	.09	-.08	.37***	.61***	.65***	.36***	.22***	-
9. Buoyancy	.14*	.02	.36***	.54***	.49***	.33***	.29***	.56***
<i>Congruent Trials</i>								
1. Science Education	-							
2. Matrix Reasoning	-.05	-						
3. Stop Signal	.10	-.01	-					
4. Figure Matching	.07	-.03	.29***	-				
5. Flanker	.12*	-.09	.39***	.55***	-			
6. Corsi Block	.10	.02	.25***	.35***	.43***	-		
7. Science Statement	.02	.15*	.19**	.32***	.16*	.04		
8. Animism	.04	-.07	.36***	.59***	.64***	.34***	.25***	-
9. Buoyancy	.11	.05	.37***	.55***	.46***	.30***	.32***	.51***
<i>Incongruent Trials</i>								
1. Science Education	-							
2. Matrix Reasoning	-.05	-						
3. Stop Signal	.10	-.01	-					
4. Figure Matching	.07	-.03	.29***	-				
5. Flanker	.12*	-.09	.39***	.55***	-			
6. Corsi Block	.10	.02	.25***	.35***	.43***	-		
7. Science Statement	.08	.24***	.07	.28**	.07	-.09		
8. Animism	.13*	-.07	.33***	.59***	.60***	.34***	.18**	-
9. Buoyancy	.11	.03	.28***	.48***	.41***	.27***	.23***	.48***

Notes. * $p < .05$; ** $p < .01$; *** $p < .001$; Correlations $< .30$ are small effect size, between $.30$ and $.50$ are medium effect size, and $> .50$ are large effect size.

5.3 Path Analysis (the United Kingdom)

The path models used efficiency and the results showed different patterns across three trial types.

In overall trials (see Figure 5.1a), the Matrix Reasoning task positively predicted the Science Statement task ($\beta = .21, p = .034$) with a medium small effect size. The Stop Signal task positively predicted the Buoyancy task ($\beta = .16, p = .004$) and the Animism task ($\beta = .10, p = .025$) both with small effect size. The Flanker task positively predicted the Buoyancy task ($\beta = .20, p = .002$) with small effect size and the Animism task ($\beta = .42, p < .001$) with medium effect size. The Figure Matching task positively predicted the Science Statement task ($\beta = .40, p < .001$), the Buoyancy task ($\beta = .36, p < .001$), and the Animism task ($\beta = .35, p < .001$) with medium effect size.

In congruent trials (see Figure 5.1b), the Matrix Reasoning task positively predicted the Buoyancy task ($\beta = .08, p = .031$) with very small effect size. The Stop Signal task positively predicted Buoyancy Task ($\beta = .18, p = .001$) and the Animism task ($\beta = .11, p = .023$) with small effect size. The Flanker task positively predicted the Buoyancy task ($\beta = .15, p = .012$) with small effect size and the Animism task ($\beta = .42, p < .001$) with medium effect size. The Figure Matching task positively predicted the Science Statement task ($\beta = .33, p < .001$), the Buoyancy task ($\beta = .40, p < .001$), and the Animism task ($\beta = .32, p < .001$) with medium effect size.

In incongruent trials (see Figure 5.1c), the Flanker task positively predicted Buoyancy ($\beta = .16, p = .024$) with small effect size and the Animism task ($\beta = .36, p < .001$) with medium effect size. The Figure Matching task positively predicted the Science Statement task ($\beta = .38, p = .003$), the Buoyancy task ($\beta = .35, p < .001$), and the Animism task ($\beta = .36, p < .001$) with medium effect size. All three path models showed good model fits (see Table 5.2 for model fit).

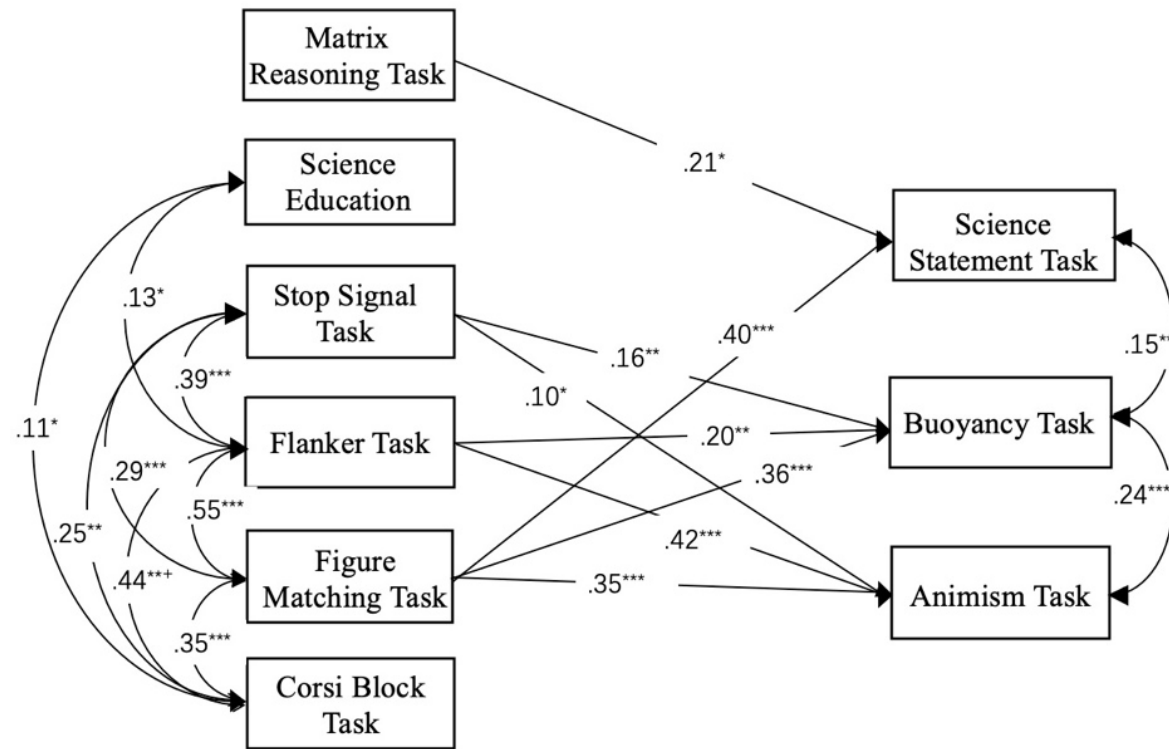
Table 5.2*Path Analysis Model Fit Indices Summary on Efficiency (the United Kingdom)*

<i>Trials</i>	$\chi^2(df)$	<i>CFI</i>	<i>TLI</i>	<i>AIC</i>	<i>BIC</i>	<i>RMSEA</i>	<i>SRMR</i>
Overall	.60(1)	1.00	1.03	6167.53	6356.66	.00	.01
Congruent	.28(1)	1.00	1.07	6199.80	6388.92	.00	.00
Incongruent	.64(1)	1.00	1.03	6236.68	6425.80	.00	.01

Notes. Significant χ^2 was presented with *; * $p < .05$; ** $p < .01$; *** $p < .001$; χ^2 = Chi-square; df = degree of freedom; *CFI* = Comparative Fit Indices; *TLI* = Tucker and Lewis' Index; *AIC* = Akaike Information Criterion; *BIC* = Bayesian Information Criterion; *RMSEA* = Root Mean Square of Approximation; *SRMR* = Standardized Root Mean Square; Robust *CFI*; *TLI*; *RMSEA* and *SRMR* were reported in the Table.

Figure 5.1a

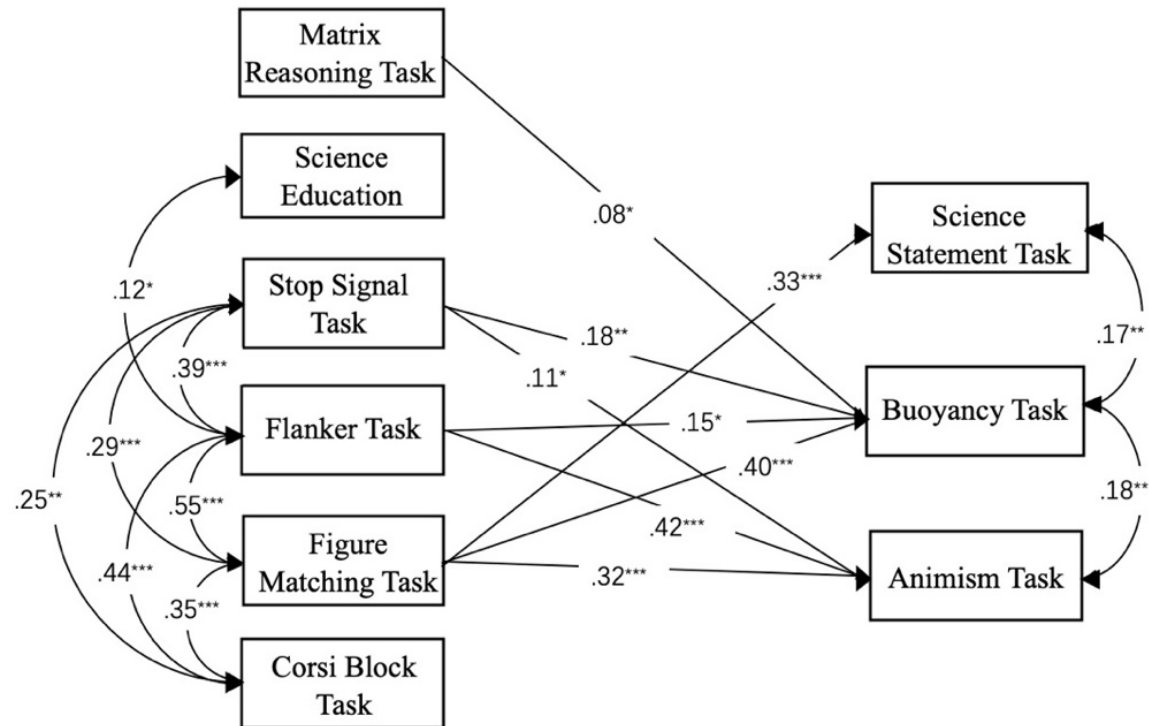
Path Analysis for Overall trials on Efficiency (the United Kingdom)



Notes. Standardised Path Coefficients were reported. * $p < .05$; ** $p < .01$; *** $p < .001$.

Figure 5.1b

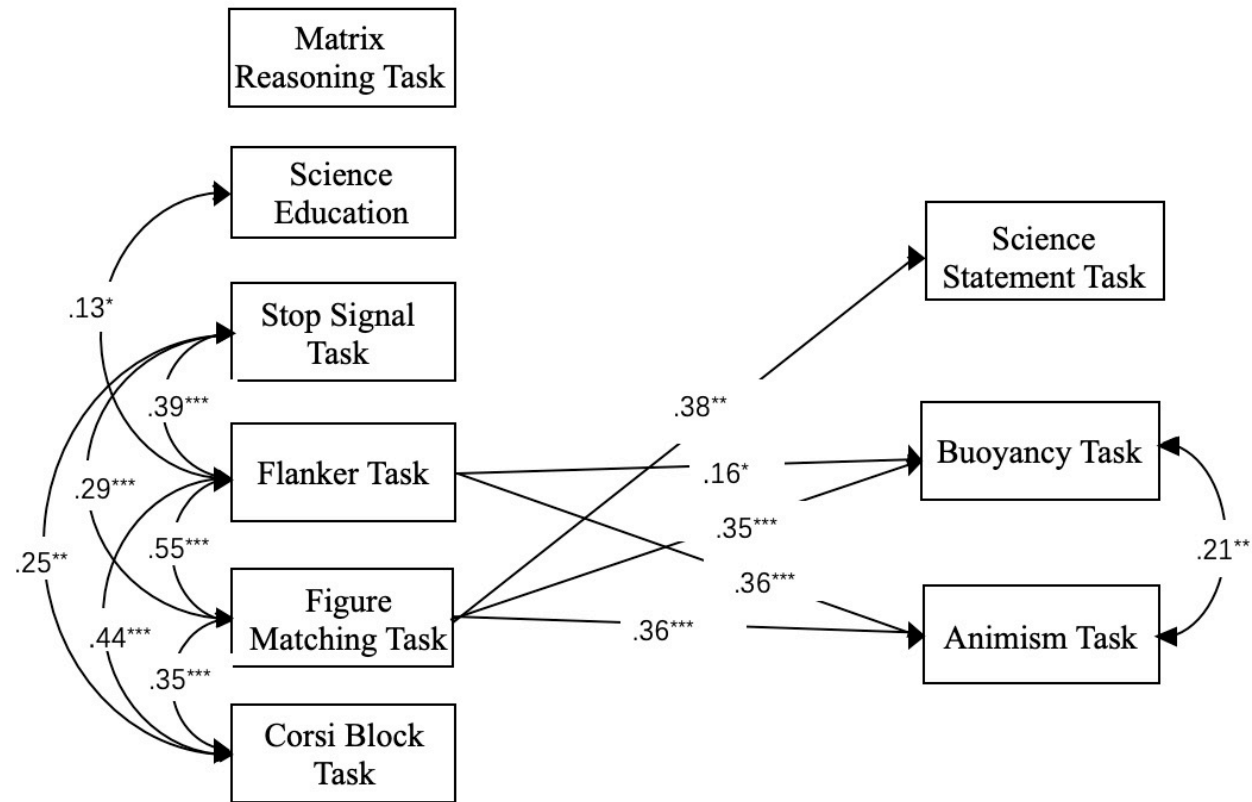
Path Analysis for Congruent trials on Efficiency (the United Kingdom)



Notes. Standardised Path Coefficients were reported. * $p < .05$; ** $p < .01$; *** $p < .001$.

Figure 5.1c

Path Analysis for Incongruent trials on Efficiency (the United Kingdom)



Notes. Standardised Path Coefficients were reported. * $p < .05$; ** $p < .01$; *** $p < .001$.

5.4 Discussion on Study 1

As outlined in the Introduction, Study 1 was inspired by my master's research on exploring the links between switching and inhibition in explanatory coexistence (Ellefson et al., 2024). Due to the various obstacles imposed by COVID-19, this research was restricted to a concentrated focus on the roles of inhibition and switching within explanatory coexistence. Study 1 expanded upon this investigation by further incorporating working memory, general cognitive ability and science education into the research scope, and whether the results obtained from my master's research could be replicated in the current study. However, considering that this PhD research encompassed two studies, results from Study 1 are only briefly discussed, in this section. A more comprehensive discussion of the full PhD's research questions are presented in Chapter 7. Table 5.3 lists the research questions discussed in this section, as they relate to Study 1.

Table 5.3*Summary of Study 1 Analysis and Corresponding Discussion*

Research Questions	Analysis Used	Results	Discussion Sections
Research Question 1: What are the relative contributions of executive functions, science education and general cognitive ability to explanatory coexistence?	Correlation (Overall trials) and Path Analysis (Overall trials; Figure 5.1a)	Executive functions were the most important factors associated with explanatory coexistence	5.4.1 Relative Contributions of Executive Functions, General Cognitive Ability and Science Education on Naïve Thinking Tasks
Research Question 2: Do task modalities (i.e., pictorial stimuli versus text stimuli) affect the involvement of executive functions, science education and general cognitive ability in explanatory coexistence?	Correlation (Overall trials) and Path Analysis (Overall trials; Figure 5.1a)	Picture-based task recruiting both inhibition and switching; text-based task recruiting switching and general cognitive ability.	5.4.2 Different Naïve Thinking Task Modalities Require Different Skills
Research Question 3: Do executive functions, science education and general cognitive ability contribute differently to congruent and incongruent trials in explanatory coexistence?	Correlation (Congruent and Incongruent trials) and Path Analysis (Congruent and Incongruent trials; Figure 5.1b & Figure 5.1c)	Overall contributing pattern of executive functions, science education and general cognitive ability on explanatory coexistence were similar across trial types.	5.4.3 Congruent and Incongruent Trial Types Require Different Skills

5.4.1 Relative Contributions of Executive Functions, General Cognitive Ability and Science Education on Explanatory Coexistence

This section discusses the contributions of executive functions, general cognitive ability and science education to explanatory coexistence, as derived from the results of Study 1. All discussions in the following sections are based on efficiency analysis.

Contribution of Executive Functions. Overall, executive functions were found to be better predictors of explanatory coexistence than general cognitive ability and science education. Results of the correlation analysis showed that behavioural inhibition, interference inhibition and switching positively correlated with all three naïve thinking tasks with medium-to-large effect sizes, while working memory only linked positively with the Animism and Buoyancy tasks with medium effect sizes. These results suggested that, when each component of executive functions was observed individually, all components linked with explanatory coexistence, and that this association was relatively strong from an effect size perspective.

Results aligned with the view that all executive functions are essential in supporting scientific reasoning. The explanatory coexistence process requires the use of working memory to maintain beliefs, coupled with the flexible switching of these beliefs in ever-changing contexts, and the need to inhibit unnecessary information (Vosniadou et al., 2018). Nevertheless, path analysis results demonstrated that among executive functions, inhibition and switching, but not working memory, directly predicted explanatory coexistence. More specifically, behavioural and interference inhibition as well as switching played different roles in predicting the performances of different naïve thinking tasks. Both behavioural inhibition (small effect) and interference inhibition (small-to-medium effect) positively predicted the Animism and Buoyancy tasks, while switching predicted all three naïve thinking tasks with medium

effect sizes. Comparing prediction and predictive power, these results imply that, when considering executive functions' internal relationships, switching may be a stronger predictor for explanatory coexistence than inhibition and working memory.

Contribution of General Cognitive Ability. General cognitive ability is often regarded as a control variable alongside executive functions, and is usually significantly linked with explanatory coexistence (Bascandziev et al., 2018; Brookman-Byrne et al., 2018). In this study, correlation results indicated that general cognitive ability was only associated with the Science Statement task with a small effect size in overall trials, and the Animism task with a small effect size in incongruent trials. Such results align with previous findings assessing general cognitive ability and science statement naïve thinking tasks (Brookman-Byrne et al., 2018; Dumontheil et al., 2022, 2023). However, this study's results departed from previous studies showing a significant contribution from general cognitive ability to animism naïve thinking tasks (Bascandziev et al., 2018; Zaitchik et al., 2014). A possible explanation for this divergence is that general cognitive ability (assessed via the Matrix Reasoning task in the current study) is more associated with complex naïve thinking tasks that require a greater amount of reasoning and understanding. The Science Statement task covers a wide range of science knowledge, necessitating the use of general cognitive ability to understand and reason these concepts. However, the Animism task contains only single science concepts that do not necessarily require general cognitive ability to understand or reason through. That said, in congruent trials general cognitive ability was found to contribute to the Buoyancy task, albeit with a very small effect size (.08). This might be attributed to the Buoyancy task's design, where two balls appear simultaneously on the screen and require that participants engage in a greater amount of reasoning during their responses. Nevertheless, given the minimal effect size, further verification is necessary to determine whether this association is needed. In all, general cognitive ability may be

more involved in the complex the Science Statement task, or related to task modality, as discussed in the following sections.

Science Education. Science education has been assessed in terms of dichotomous variables (i.e., experts vs. novice) in previous studies (Brault Foisy et al., 2021; Jin et al., 2019; Mason et al., 2019), which have found that experts exhibit greater brain activations when answering naïve thinking tasks when compared to novices. However, in the current research, correlation results for United Kingdom participants showed that science education, when considered as a continuous variable, only correlated significantly with performance in the Buoyancy task, with a small effect size. These results may be due to the fact that animism and buoyancy knowledge is usually taught in early childhood or by middle school, and thus used more in science classes (Brault Foisy et al., 2021; Carey, 2000). Those who continue to study science gradually use the animism and buoyancy concept more frequently, leading to better performance on the Animism and Buoyancy tasks among participants with higher levels of science education. Still, no significant correlations were found between science education and the Animism and Science Statement tasks, which may be due to the fact that the general science concepts (in the Science Statement task) require greater knowledge and are harder to be accepted by people with various science education levels (Shtulman & Valcarcel, 2012). Path analysis results showed that science education did not directly predict performance on any naïve thinking tasks. However, science education correlated positively with interference inhibition, meaning that science education may influence explanatory coexistence through executive functions; this link, however, requires further investigation.

In sum, among United Kingdom participants, both correlation and path analysis results indicated that executive functions have stronger predictive power for explanatory coexistence when compared with general cognitive ability and science

education, regardless of whether a correlation or predictive relationship exists, or the strength of the predictive power. Within executive functions, switching emerged as the strongest predictor.

5.4.2 Different Modalities Require Different Skills

In the current research, executive function, general cognitive ability and naïve thinking tasks all involved different task modalities (see Table 5.4). According to Ellefson et al., (2024), task modality may be another factor influencing the relative contributions of different factors in explanatory coexistence.

Table 5.4

Summary of Each Task's Modality

Measurement	Task	Modality of Stimuli
Interference inhibition	Flanker task	Picture
Behavioural inhibition	Stop Signal task	Picture & Auditory
Switching	Figure Matching task	Picture
Working memory	Corsi Block task	Picture
General cognitive ability	Matrix reasoning task	Picture
Naïve thinking task	Animism task	Picture
Naïve thinking task	Buoyancy task	Picture
Naïve thinking task	Science Statement task	Text

Study results indicated that different contributing factors played different roles when task modality varied. Behavioural and interference inhibition only predicted task performance with pictorial stimuli (the Animism and Buoyancy tasks) while switching consistently predicted performances in all three naïve thinking tasks containing both pictorial and text stimuli. Additionally, general cognitive ability only predicted performance in the naïve thinking task containing text stimuli (the Science Statement task), while working memory and science education did not predict any naïve thinking tasks. These results highlight how different cognitive skills are presented within explanatory coexistence through different modalities and are consistent with findings and discussions from previous research involving adult participants. Stricker et al. (2021) have shown inhibition to not be involved in tasks with statements containing science and mathematics concepts, arguing that, as explanations in certain domains require lengthier statement sentences when compared with others, these sentences potentially require more response time that can impact their relationship with inhibition. Similarly, Ellefson et al. (2024) have found that, while switching constantly predicts all modalities of naïve thinking tasks, interference inhibition only predicts response times in picture-based task (the Animism task), meaning that inhibition may be absent from more complex text-based tasks.

Working memory did not predict performances in all naïve thinking tasks, as shown in this study's path analysis. As a visual working memory measurement, as discussed above, was used, this finding echoed Dumontheil et al.'s (2023) research among children, and suggests that visual working memory does not affect explanatory coexistence directly, regardless of modality. General cognitive ability only predicted explanatory coexistence in text-based, but not picture-based, naïve thinking tasks, a finding that aligns with previous research measuring general cognitive ability and naïve thinking tasks including science statements (Vosniadou et al. 2015; 2018). The differential impact of pictorial versus text stimuli may explain the association between performances in executive function tasks and those in general cognitive ability tasks,

as the former contain simpler information while the latter are more complex and require greater reflection. Therefore, as discussed in the previous section, general cognitive ability is more involved in more complex explanatory coexistence processes.

Unlike previous studies using pictorial stimuli, science education showed no direct influence on naïve thinking tasks (Jin et al., 2019; Mason et al., 2019; Zhu et al., 2019). One possible explanation for this finding might be that naïve or science thinking as indexed in all tasks were relatively simple. Adults already know these concepts well, making the influence of science education difficult to detect. As previously mentioned, since there is limited research on the impact of science education on executive functions, more investigations are needed. In sum, results from this study replicated those obtained by Ellefson et al. (2024), supporting their argument that cognitive abilities involved in processing explanatory coexistence reasoning differ across task modalities.

5.4.3 Congruent and Incongruent Trials Require Different Skills

Results from correlation and path analysis found that the involvement of executive functions, general cognitive ability and science education in explanatory coexistence was similar, but not exactly identical, across congruent and incongruent trials. Both correlation and path analysis results consistently indicated interference inhibition and switching to be robustly associated with naïve thinking tasks in all trials, and that science education and working memory did not predict naïve thinking tasks in any trial. General cognitive ability and behavioural inhibition, however, only predicted naïve thinking tasks in congruent trials.

Congruent trials are trials with no conflict between naïve and science thinking. The focus of these trials was more on the understanding of naïve and science knowledge, rather than the need for a suppression process. Correlation results showed that science education did not correlate with congruent trials in the Buoyancy task,

while path analysis results showed that both types of inhibition and switching positively predicted congruent trials of naïve thinking tasks, with similar effect sizes (small to medium) to congruent trials. However, although general cognitive ability also predicted the Buoyancy task, the effect size is very small with .08 when compared with this link in overall trials.

In incongruent trials, where conflicts emerged between naïve and science thinking, participants' need to convert between the two types of thinking needed to partake in a more complex cognitive process. Correlation results for the incongruent trials therefore showed some differences when compared with the congruent trials for the Science Statement task, with only switching and general cognitive ability correlating with performance. Still, science education correlated with incongruent trials in the Animism task, and this link did not show in overall and incongruent trials. In path analysis, results found that only interference inhibition and switching contributed to explanatory coexistence with similar effect size (small- to- medium), compared to overall trials.

Thus, this study's results showed that when separating overall trials of the naïve thinking task into congruent and incongruent trials, switching and interference inhibition consistently contributed to both trial types. These two executive function components were not only essential for learning of scientific knowledge (in congruent trials), but also necessary for the suppression of naïve thinking and science thinking activation (in incongruent trials), thereby contributing to explanatory coexistence. This finding aligns with previous research separating congruent and incongruent trials (Ellefson et al. 2024, Vosniadou et al. 2015; 2018), which suggests that both inhibition and switching are vital in the science reasoning process.

Behavioural inhibition, however, did not predict performance in incongruent trials. This result also aligns with Ellefson et al.'s (2024) findings that behavioural inhibition does not directly, but correlates with other executive functions that do directly, predict explanatory coexistence. One possible reason for this finding, as

inferred from previous studies, is that distinct performances of behavioural and interference inhibitions in explanatory coexistence can be observed (Brookman-Byrne et al., 2018). Individuals with better behavioural inhibition may spend more time coming up with correct responses in incongruent trials, leading to decreased efficiency, while in congruent trials where conflicts are lacking, individuals with better behavioural inhibition will spend less time and with higher accuracy on responses, making behavioural inhibition significantly contribute to efficiency in congruent trials.

General cognitive ability only predicted the Buoyancy task in congruent trials, with very small effect size, but not in incongruent trials. This finding echoes previous studies assessing general cognitive ability (Dumontheil et al., 2022; Vosniadou et al., 2018), suggesting that general cognitive ability may contribute more to learning scientific concepts than suppressing naïve thinking; such a link, however, requires further exploration as the effect size of this link is too small and may lack predictive power. Overall, results of the current study show that when conflicts arise between naïve and science thinking, executive functions, rather than general cognitive ability, are generally more involved in explanatory coexistence.

5.4.4 Links Between Study 1 and Study 2

Results from Study 1 were largely consistent with those of my master's work (Ellefson et al., 2024), suggesting that executive functions, particularly switching and inhibition, were key factors in the reasoning process of explanatory coexistence. However, as discussed in Chapter 3, current literature on explanatory coexistence seldom includes research targeting non-WEIRD participants, which can lead to potential bias due sample selection. Thus, the question of whether Study 1's results were generalisable across different cultural contexts remained unresolved. Therefore, the next chapter provides results from Study 2, a study conducted with Chinese participants, that used the same research tools to ascertain Study 1's robustness.

Together, these two studies could provide an understanding of the mechanisms underlying explanatory coexistence, as viewed in different, culturally diverse contexts.

Chapter 6 Results and Discussion for Study 2: Analyses Using the United Kingdom and Chinese data

In this chapter, I analyse the data collected from both the United Kingdom and China to address the two research questions proposed in Study 2 using efficiency metrics: (1) Are the performances of executive functions, general cognitive ability, and explanatory coexistence similar among adult participants from the United Kingdom and China? (2) Is the relationship between science education, executive functions, general cognitive ability, and explanatory coexistence robust across the United Kingdom and China? Before presenting the main analysis, I provide the preliminary analysis of the combined data from the United Kingdom and China, as well as the data specific to China. In the main analysis, as outlined in Chapter 4, I first present scatterplots illustrating the relationships between accuracy and response time for various tasks across the two regions to address the first research question. Following this, I report the results of the multigroup analysis. Significant findings from the multigroup analysis led to the separation of data from the United Kingdom and China for further analysis. Correlation and path analyses using the Chinese data are presented in the same structure as in Chapter 6, including analyses of overall, congruent, and incongruent trials. Finally, I provide a discussion of the results corresponding to the two research questions in Study 2. Additional results related to accuracy and response time are included in Appendix K and Appendix L.

6.1 Preliminary Analysis (China)

All results from preliminary analysis on efficiency for the United Kingdom data were presented in Appendix M. Descriptive statistics analysis was conducted to summarise the mean, standard deviation, minimum value and maximum value of the efficiency data. Then, age adjustment test shown age significantly influenced the efficiency (linear). After conducting age adjustment and z score conversion, the Shapiro-Wilk test results indicated that most of the efficiency adhered to a non-normal distribution.

The congruent and incongruent trials for the three naïve thinking tasks differed in the expected direction for efficiency based on raw score analysis: efficiency for Science Statement, Animism, and the Buoyancy tasks in congruent trials was higher than in incongruent trials. Significant large effect size correlations were found between congruent and incongruent trials in the Animism task ($r(260) = .85, p < .001$), the Buoyancy task ($r(260) = .80, p < .001$), and the Science Statement task ($r(260) = .78, p < .001$).

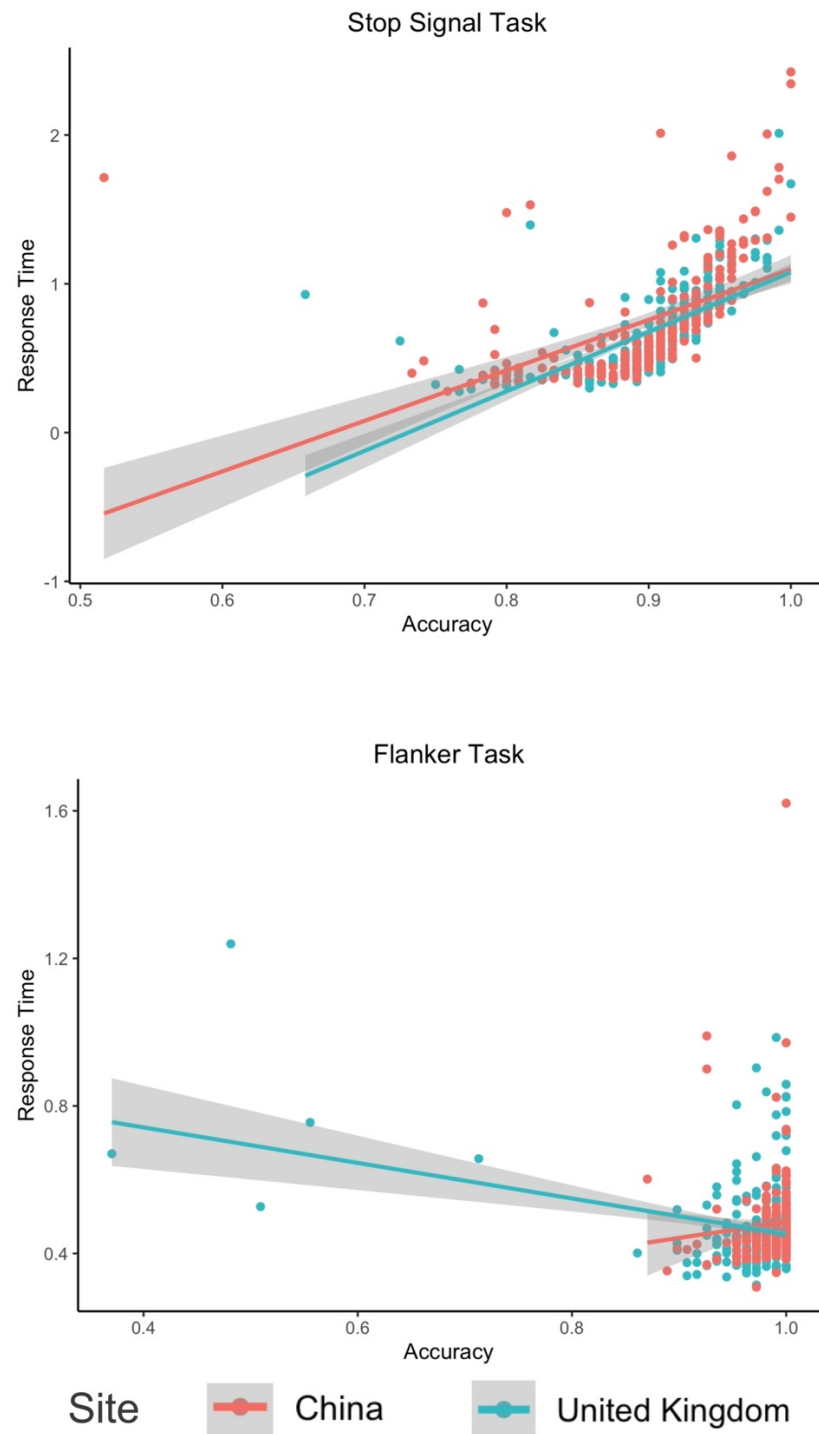
Multicollinearity and homoscedasticity results showed no multicollinearity and no heteroskedasticity in all efficiency models. Finally, I conducted a confirmatory factor analysis to test whether the efficiency data could form latent variables in structural equation modelling. Model fits of overall trials ($\chi^2(23) = 35.07, p = .051, CFI = .96, TLI = .93, RMSEA = .05, SRMR = .05$), congruent trials ($\chi^2(23) = 33.93, p = .066, CFI = .96, TLI = .94, RMSEA = .05, SRMR = .04$), and incongruent trials ($\chi^2(23) = 33.95, p = .066, CFI = .96, TLI = .93, RMSEA = .05, SRMR = .05$) all suggested good model fit to proposed model using the efficiency. However, collinearity was found between latent variable executive functions and explanatory coexistence, with overall trials ($\beta = .86, p < .001$), congruent trials ($\beta = .87, p < .001$), and incongruent trials ($\beta = .83, p = .006$), indicating the need to use path analysis to analyse all variables separately.

6.2 Response Patterns Between the United Kingdom and China

The scatter plots (see Figure 6.1a-g) indicated that the relation between accuracy and response time between the two countries. Scatter plots shows Chinese participants typically took more time to provide more accurate answers across all tasks and trials. More detailed interpretation of these results is presented in later discussion.

Figure 6.1a

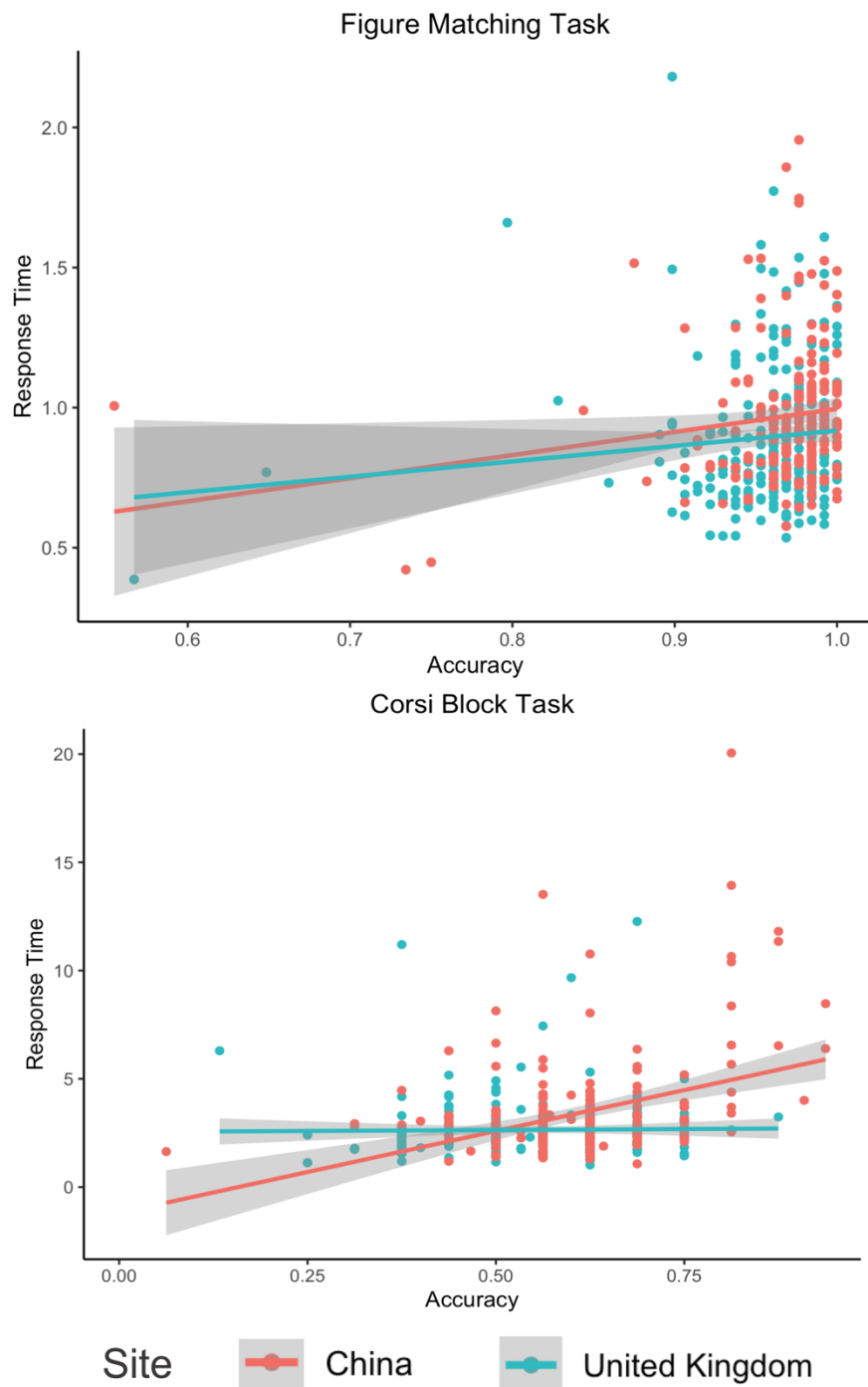
Scatterplot of the Link Between Accuracy and Response Time Between China and the United Kingdom: Stop Signal Task and Flanker Task



Notes. Solid lines represent best-fitting regression line. Grey areas represent 95% confidence interval; Response Time here refers to response time to correct trials.

Figure 6.1b

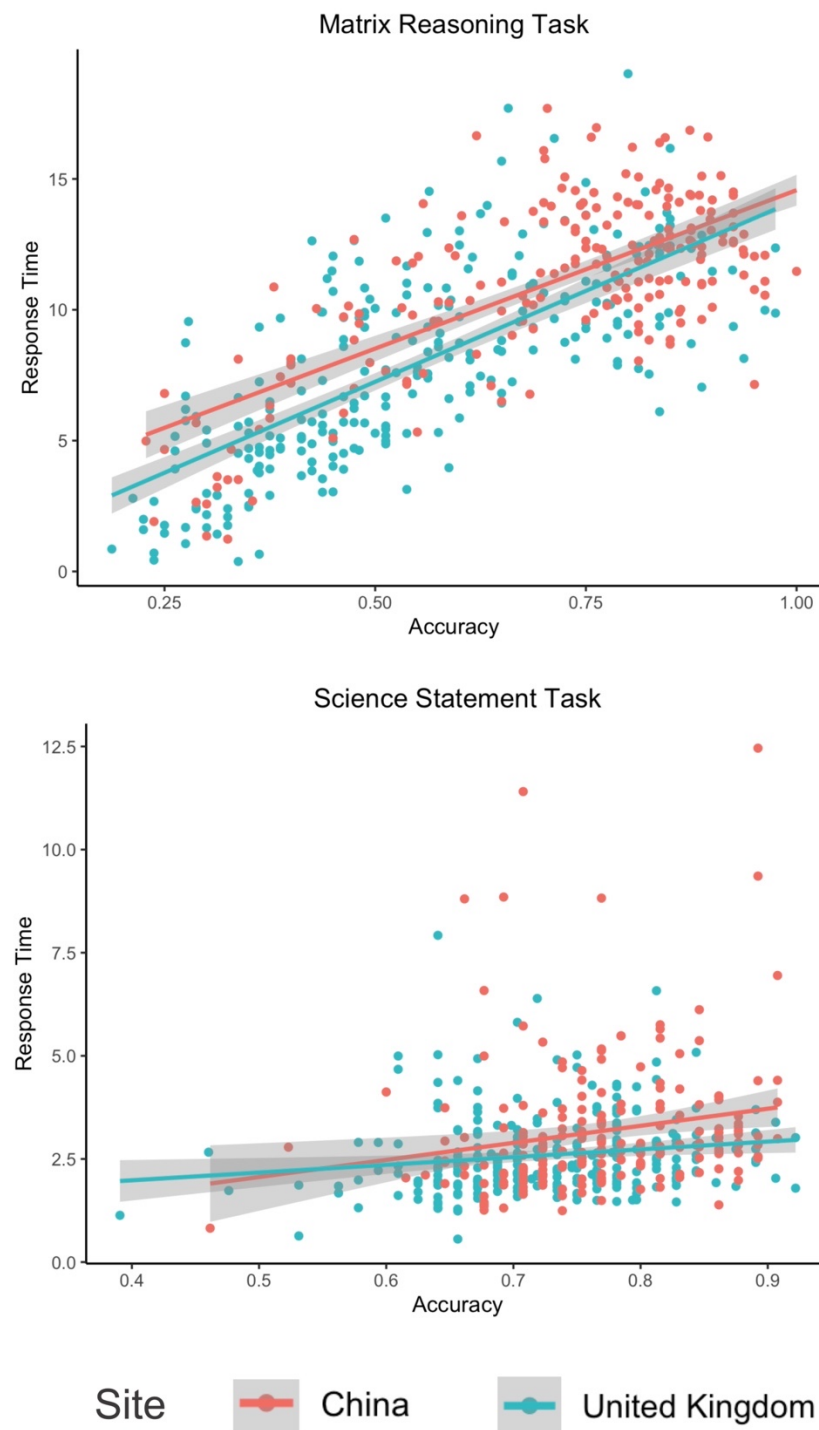
Scatterplot of the Link Between Accuracy and Response Time Between China and the United Kingdom: Figure Matching Tasks and Corsi Block Task



Notes. Solid lines represent best-fitting regression line. Grey areas represent 95% confidence interval; Response Time here refers to response time to correct trials.

Figure 6.1c

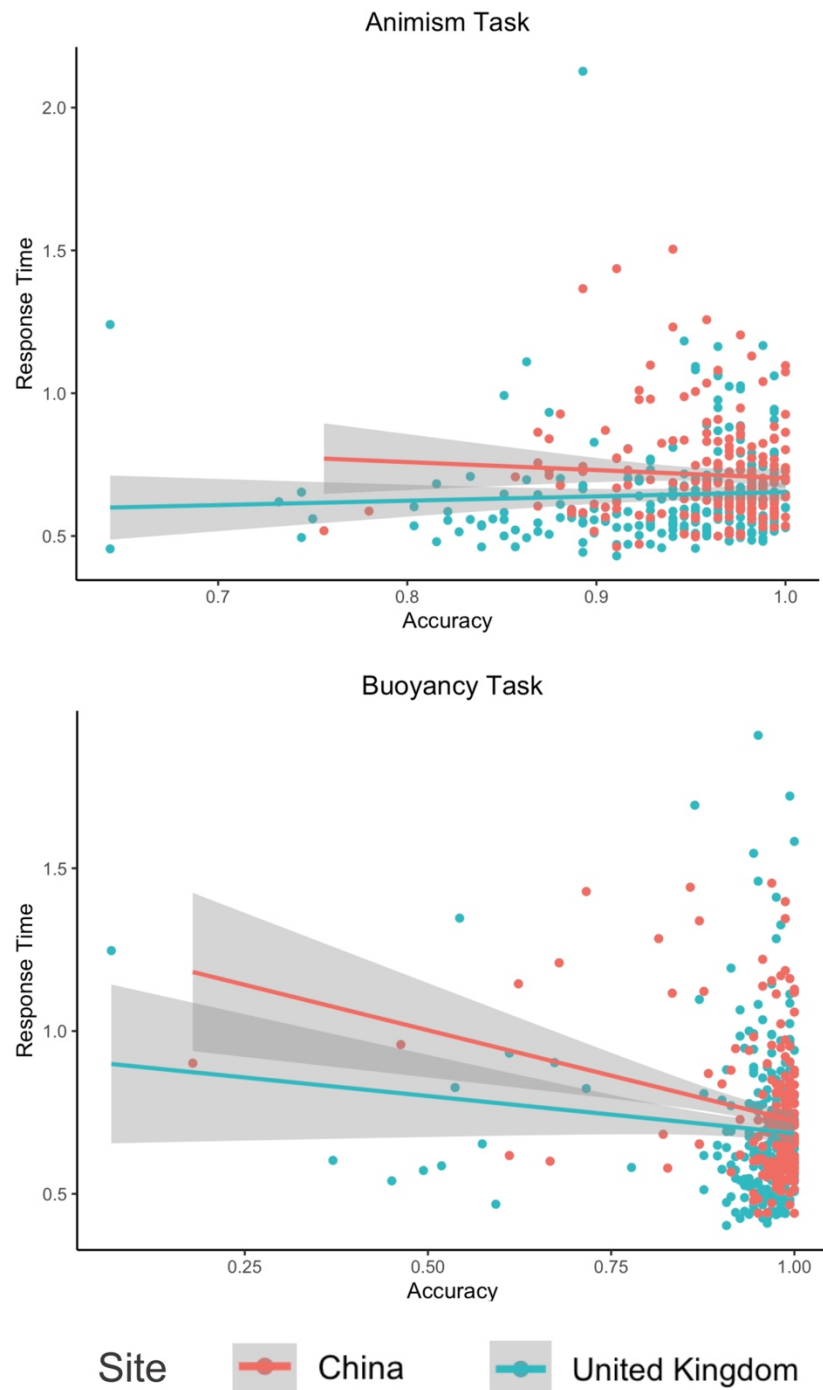
Scatterplot of the Link Between Accuracy and Response Time Between China and the United Kingdom: Matrix Reasoning Task and Science Statement Tasks



Notes. Solid lines represent best-fitting regression line. Grey areas represent 95% confidence interval; Response Time here refers to response time to correct trials.

Figure 6.1d

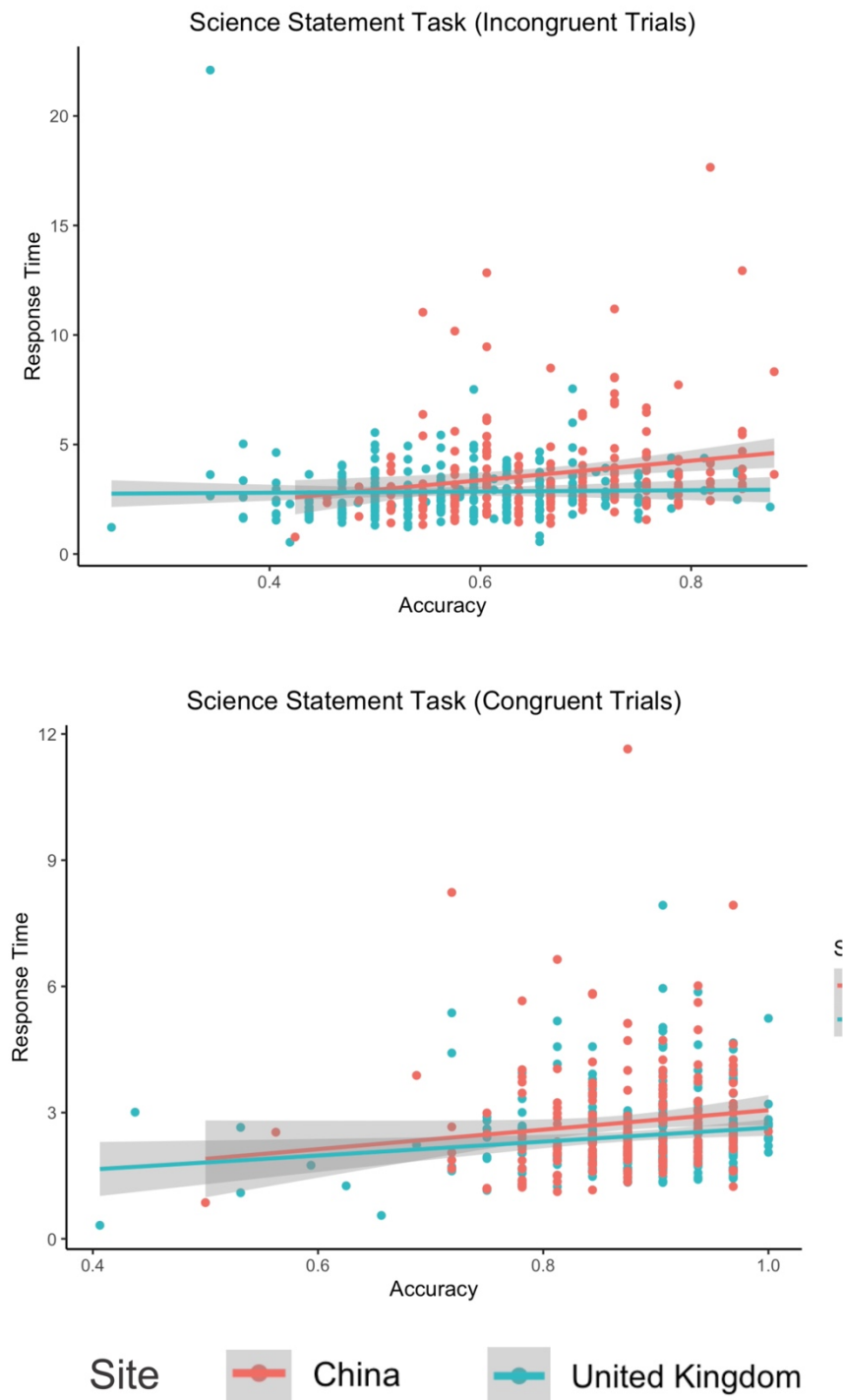
Scatterplot of the Link Between Accuracy and Response Time Between China and the United Kingdom: Animism Task and Buoyancy Tasks



Notes. Solid lines represent best-fitting regression line. Grey areas represent 95% confidence interval; Response Time here refers to response time to correct trials.

Figure 6.1e

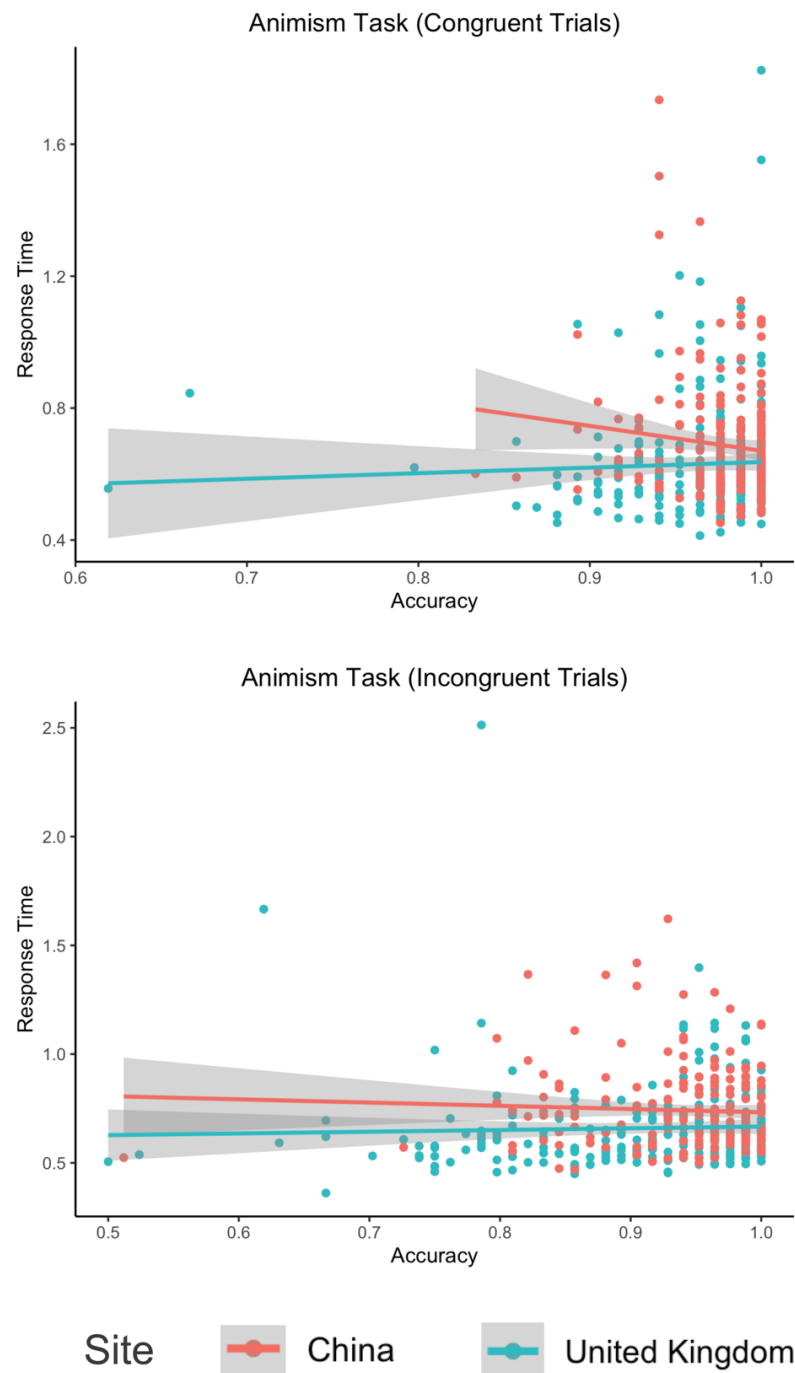
Scatterplot of the Link Between Accuracy and Response Time Between China and the United Kingdom: Congruent and Incongruent Trials of Science Statement Tasks



Notes. Solid lines represent best-fitting regression line. Grey areas represent 95% confidence interval; Response Time here refers to response time to correct trials.

Figure 6.1f

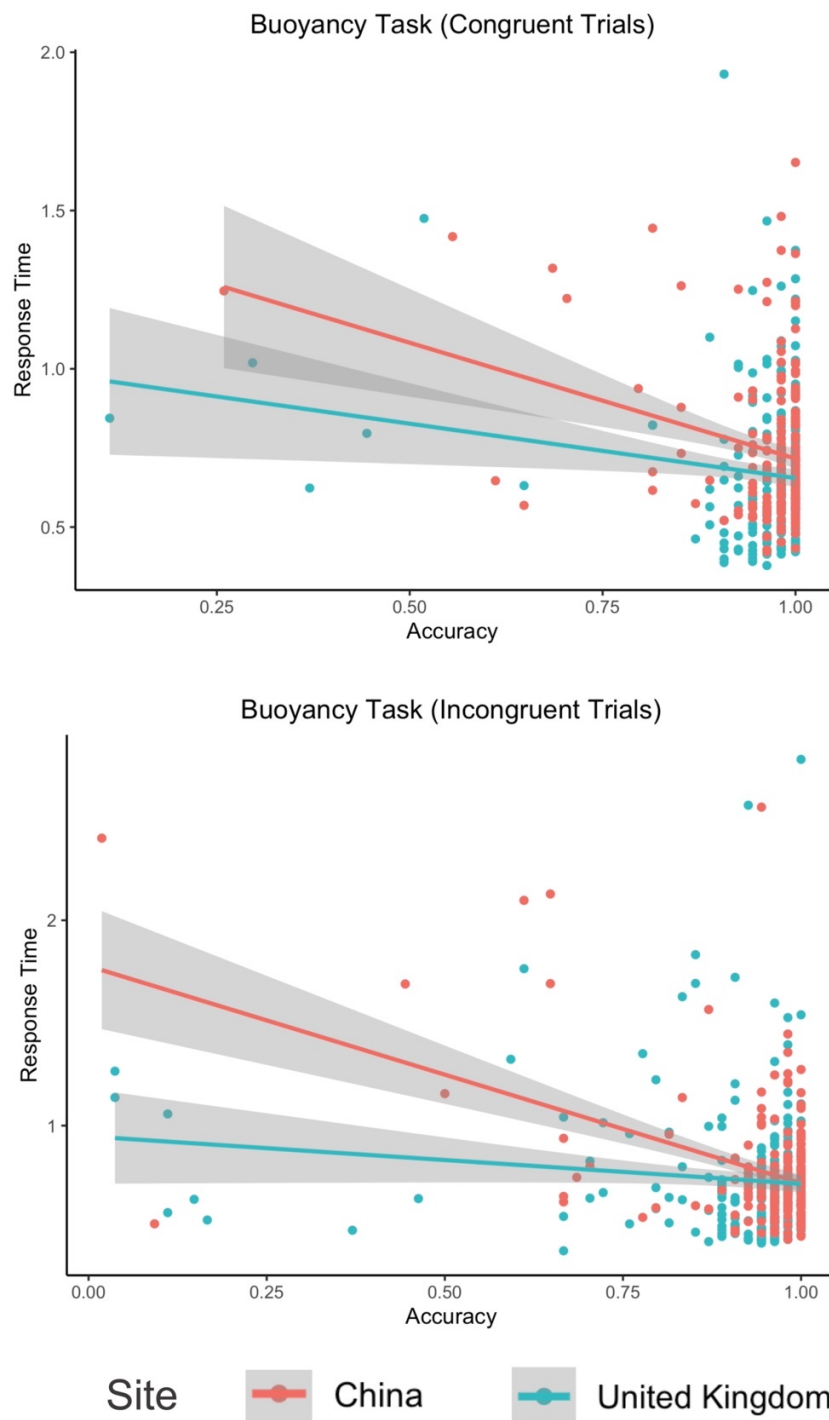
Scatterplot of the Link Between Accuracy and Response Time Between China and the United Kingdom: Congruent and Incongruent Trials of Animism Tasks



Notes. Solid lines represent best-fitting regression line. Grey areas represent 95% confidence interval; Response Time here refers to response time to correct trials.

Figure 6.1g

Scatterplot of the Link Between Accuracy and Response Time Between China and the United Kingdom: Congruent and Incongruent Trials of Animism Tasks



Notes. Solid lines represent best-fitting regression line. Grey areas represent 95% confidence interval; Response Time here refers to response time to correct trials.

6.3 Preliminary Analysis (the United Kingdom and China)

In this section, I first conducted a preliminary analysis of the combined data, as well as a multigroup analysis. All results from preliminary analysis on efficiency for the United Kingdom data were presented in Appendix N. Descriptive statistics analysis was conducted to summarise the mean, standard deviation, minimum value and maximum value of the efficiency data. Then, age adjustment test shown age significantly influenced the efficiency (curvilinear). After conducting age adjustment and z score conversion, the Shapiro-Wilk test results indicated that most of the efficiency adhered to a non-normal distribution.

Multicollinearity results indicated no multicollinearity in all efficiency models. However, homoscedasticity tests revealed heteroskedasticity in models where the dependent variables are overall trials and incongruent trials of the Animism task. Finally, I conducted a confirmatory factor analysis to test whether the efficiency data could form latent variables in structural equation modelling (see Appendix N, Table N6). Model fits of overall trials ($\chi^2(23) = 51.66, p = .001, CFI = .95, TLI = .92, RMSEA = .07, SRMR = .04$), congruent trials ($\chi^2(23) = 51.30, p = .001, CFI = .95, TLI = .92, RMSEA = .06, SRMR = .04$), and incongruent trials ($\chi^2(23) = 33.95, p = .066, CFI = .93, TLI = .89, RMSEA = .07, SRMR = .05$) all suggested good model fit to proposed model using the efficiency. However, collinearity was found between latent variable executive functions and explanatory coexistence, with overall trials ($\beta = .95, p < .001$), congruent trials ($\beta = .95, p < .001$), and incongruent trials ($\beta = .93, p = .006$), indicating the need to use Path Analysis to analyse all variables separately.

6.4 Multigroup Analysis Between the United Kingdom and China

Multigroup Analysis revealed significant differences between the tested model and the constrained model for both overall and congruent trials when comparing the

data from the United Kingdom and China using efficiency (see Table 6.1). This result indicates that there was no statistically significant measurement invariance between these two sites when assessing performance on overall trials and congruent trials, but there was measurement invariance in incongruent trials. Consequently, this finding suggested that I needed to test participants from the United Kingdom and China separately for overall and congruent trials, but together for incongruent trials. For consistency reasons, I reported results from separate analyses (analysis for Chinese data only) in the main text. Thus, as I had already reported results from the United Kingdom data in chapter 5, the following sections 6.5 and 6.6 report results from Chinese data.

Table 6.1*Fit statistics and Significance Tests for Multigroup Analysis on Efficiency*

Model fits	Overall trials		Congruent trials		Incongruent trials	
	Baseline	Constrained	Baseline	Constrained	Baseline	Constrained
χ^2	2.24	47.32	1.65	52.17	1.93	38.43
df	2	28	2	28	2	28
CFI	1.00	.98	1.00	.98	1.00	.99
TLI	.99	.95	1.02	.94	1.00	.97
$RMSEA$	.02	.05	.00	.06	.00	.04
$SRMR$	.01	.04	.01	.04	.01	.04
$\Delta\chi^2$	-	45.14	-	50.977	-	36.41
Δdf	-	26	-	26	-	26
p -value	-	.011*	-	.002**	-	.084

Notes. Significant χ^2 was presented with *; * $p < .05$; ** $p < .01$; *** $p < .001$; χ^2 = Chi-square; df = degree of freedom; CFI = Comparative Fit Indices; TLI = Tucker and Lewis' Index; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; $RMSEA$ = Root Mean Square of Approximation; $SRMR$ = Standardized Root Mean Square; Robust CFI ; TLI ; $RMSEA$ and $SRMR$ were reported in the Table.

6.5 Correlation Analysis (China)

Correlation analysis results using efficiency were presented below (see Table 6.2). As the specific r and significance are presented in the table, the current section only summarises correlation results.

There were significant intercorrelations with small to large effect size within four executive function tasks, and significant intercorrelations with small to medium effect size within three naïve thinking tasks across all three trial types. However, all other correlations are not statistically significant.

In overall trials, the Matrix Reasoning task showed a significant correlation with a small effect size with the Science Statement task. The Stop Signal task showed a significant correlation with a small effect size with both the Animism task and the Buoyancy task. The Figure Matching task, the Flanker task, and the Corsi Block task showed significant correlations with small to medium effect sizes with all three naïve thinking tasks.

In congruent trials, the Matrix Reasoning task showed significant correlation with small effect size with the Science Statement task. The Stop Signal task showed significant correlation with small effect size with the Animism task and the Buoyancy task. The Figure Matching task, the Flanker task, and the Corsi Block task indicated significant correlation with small to medium effect size with all three naïve thinking tasks.

In incongruent trials, the Matrix Reasoning task showed significant correlation with small effect size with the Science Statement task. The Stop Signal task and the Flanker task showed significant correlation with small to medium effect size with the Animism task and the Buoyancy task. The Figure Matching task and the Corsi Block task indicated significant correlation with small to medium effect size with all three naïve thinking tasks.

Table 6.2*Correlation Analysis on Efficiency (China)*

Variables	1	2	3	4	5	6	7	8
<i>Overall Trials</i>								
1. Science Education	-							
2. Matrix Reasoning	.06							
3. Stop Signal	.05	.00	-					
4. Figure Matching	.04	.07	.32***	-				
5. Flanker	-.01	-.01	.32***	.57***	-			
6. Corsi Block	-.05	.09	.17**	.22***	.36***	-		
7. Science Statement	.01	.27***	.05	.27***	.16*	.17**		
8. Animism	.09	.04	.22***	.49***	.49***	.16*	.33***	-
9. Buoyancy	.07	.11	.20**	.46***	.45***	.19**	.23***	.46***
<i>Congruent Trials</i>								
1. Science Education	-							
2. Matrix Reasoning	.06	-						
3. Stop Signal	.05	.00	-					
4. Figure Matching	.04	.07	.32***	-				
5. Flanker	-.01	-.01	.32***	.57***	-			
6. Corsi Block	-.05	.09	.17**	.22***	.36***	-		
7. Science Statement	.03	.25***	.07	.33***	.21**	.16*		
8. Animism	.10	.00	.22***	.46***	.49***	.16*	.33***	-
9. Buoyancy	.05	.09	.19**	.42***	.40***	.16*	.24***	.41***
<i>Incongruent Trials</i>								
1. Science Education	-							
2. Matrix Reasoning	.06	-						
3. Stop Signal	.05	.00	-					
4. Figure Matching	.04	.07	.32***	-				
5. Flanker	-.01	-.01	.32***	.57***	-			
6. Corsi Block	-.05	.09	.17**	.22***	.36***	-		
7. Science Statement	.01	.28***	.05	.21**	.11	.15*		
8. Animism	.07	.07	.20**	.49***	.46***	.15*	.30***	-
9. Buoyancy	.05	.13	.18**	.44***	.41***	.21**	.22***	.44***

Notes. * $p < .05$; ** $p < .01$; *** $p < .001$; Correlations $< .30$ are small effect size, between $.30$ and $.50$ are medium effect size, and $> .50$ are large effect size.

6.6 Path Analysis on Efficiency (China)

The path models used efficiency and the results showed different patterns across three trial types.

In overall trials (see Figure 6.2a), the Matrix Reasoning task positively predicted the Science Statement task ($\beta = .24, p < .001$). The Flanker task positively predicted the Buoyancy task ($\beta = .27, p = .005$) and the Animism task ($\beta = .31, p < .001$). The Figure Matching task positively predicted the Science Statement task ($\beta = .25, p < .001$), the Buoyancy task ($\beta = .28, p < .001$), and the Animism task ($\beta = .31, p < .001$).

In congruent trials (see Figure 6.2b), the Matrix Reasoning task positively predicted the Science Statement task ($\beta = .22, p < .001$). The Flanker task positively predicted the Buoyancy task ($\beta = .23, p = .011$) and the Animism task ($\beta = .33, p < .001$). The Figure Matching task positively predicted the Science Statement task ($\beta = .30, p < .001$), the Buoyancy task ($\beta = .27, p < .001$), and the Animism task ($\beta = .27, p < .001$).

In incongruent trials (see Figure 6.2c), the Matrix Reasoning task positively predicted the Science Statement task ($\beta = .26, p = .001$). The Flanker task positively predicted Buoyancy ($\beta = .23, p = .015$) and the Animism task ($\beta = .28, p < .001$). The Figure Matching task positively predicted the Science Statement task ($\beta = .19, p = .011$), the Buoyancy task ($\beta = .29, p < .001$), and the Animism task ($\beta = .32, p < .001$). All three path models showed good model fits (see Table 6.3 for model fit).

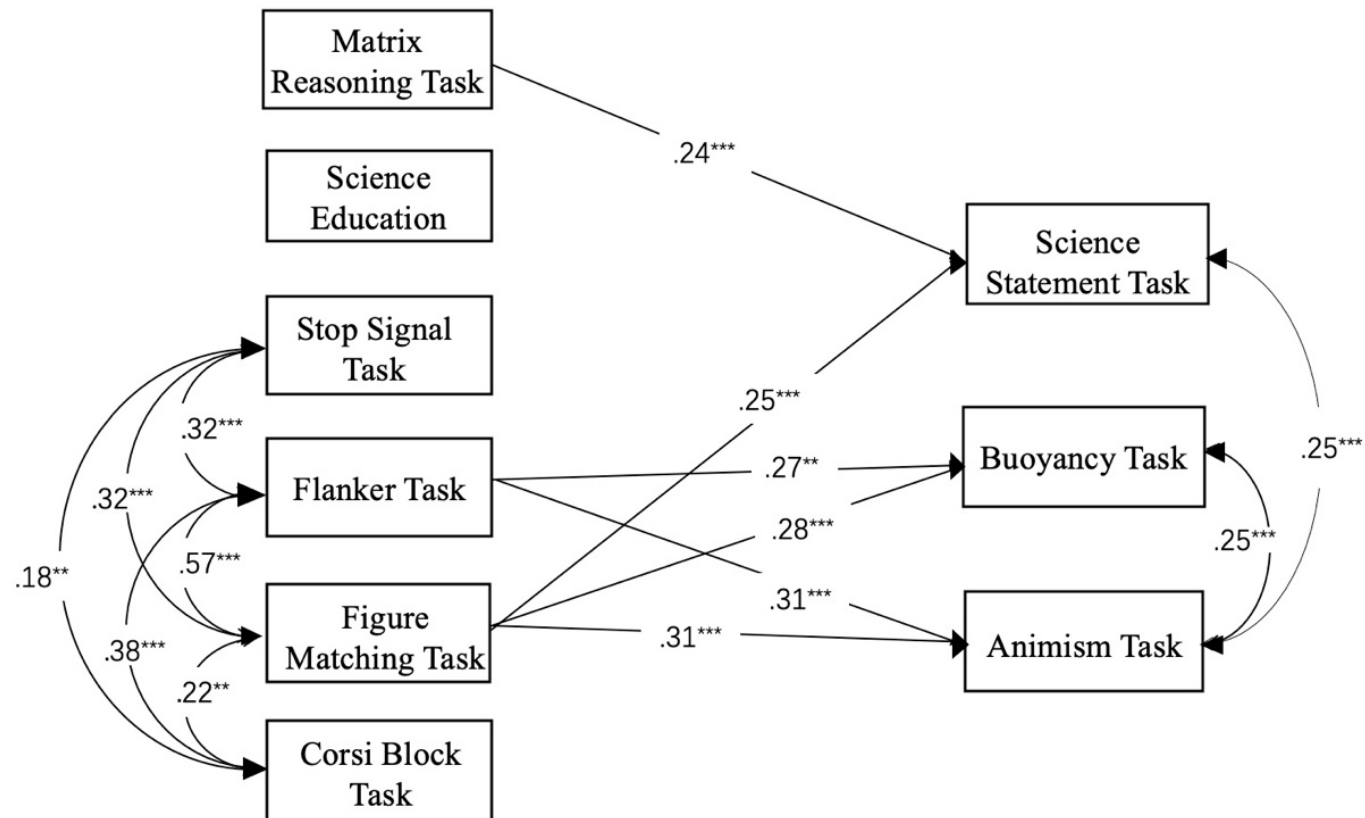
Table 6.3*Path Analysis Model Fit Indices Summary on Efficiency (China)*

<i>Trials</i>	$\chi^2(df)$	<i>CFI</i>	<i>TLI</i>	<i>AIC</i>	<i>BIC</i>	<i>RMSEA</i>	<i>SRMR</i>
Overall	.04(1)	1.00	1.11	5661.19	5844.32	.00	.00
Congruent	.01(1)	1.00	1.13	5682.90	5866.04	.00	.00
Incongruent	.07(1)	1.00	1.11	5684.02	5867.15	.00	.00

Notes. Significant χ^2 was presented with *; * $p < .05$; ** $p < .01$; *** $p < .001$; χ^2 = Chi-square; df = degree of freedom; *CFI* = Comparative Fit Indices; *TLI* = Tucker and Lewis' Index; *AIC* = Akaike Information Criterion; *BIC* = Bayesian Information Criterion; *RMSEA* = Root Mean Square of Approximation; *SRMR* = Standardized Root Mean Square; Robust *CFI*; *TLI*; *RMSEA* and *SRMR* were reported in the Table.

Figure 6.2a

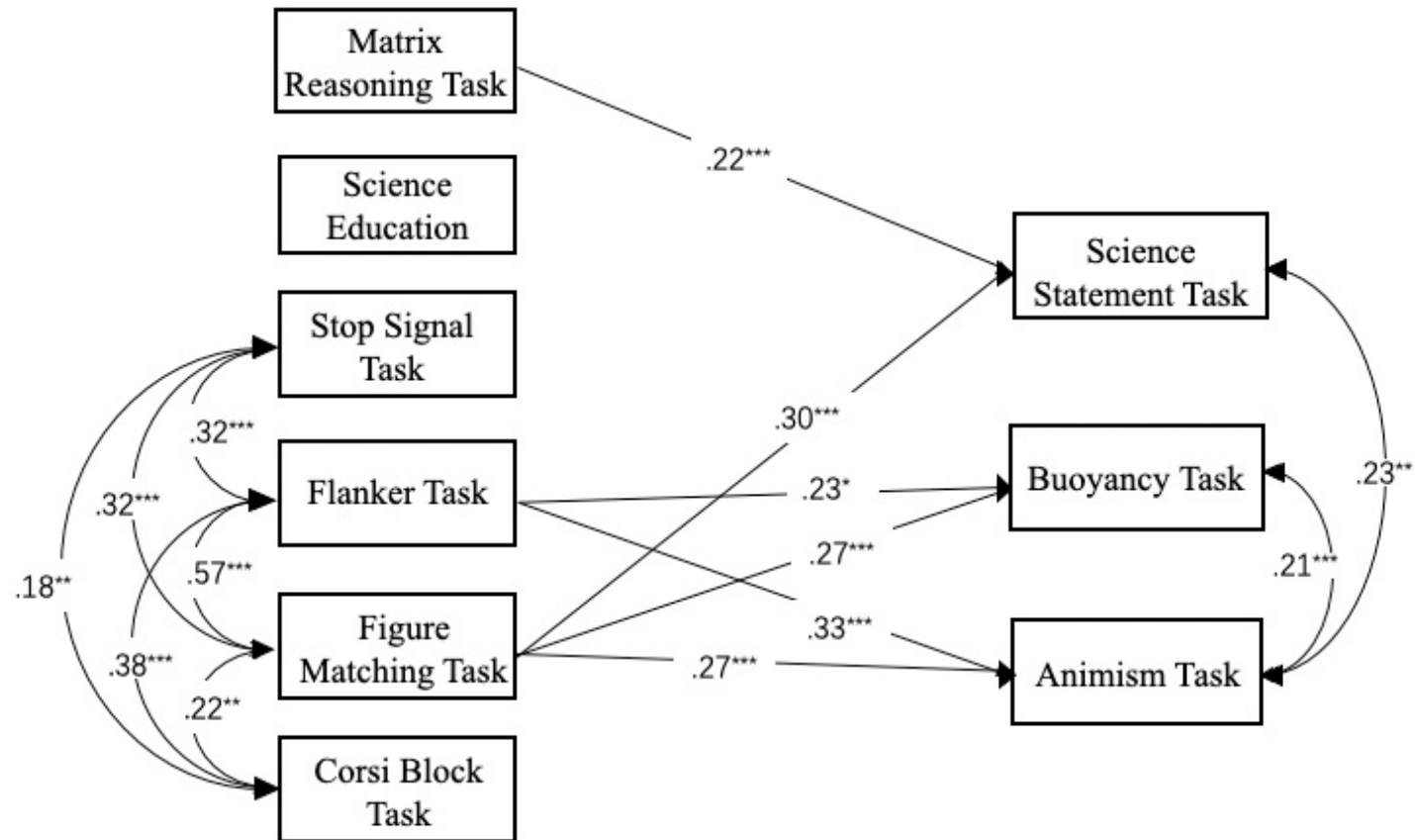
Path Analysis for Overall trials on Efficiency (China)



Notes. Standardised Path Coefficients were reported. * $p < .05$; ** $p < .01$; *** $p < .001$.

Figure 6.2b

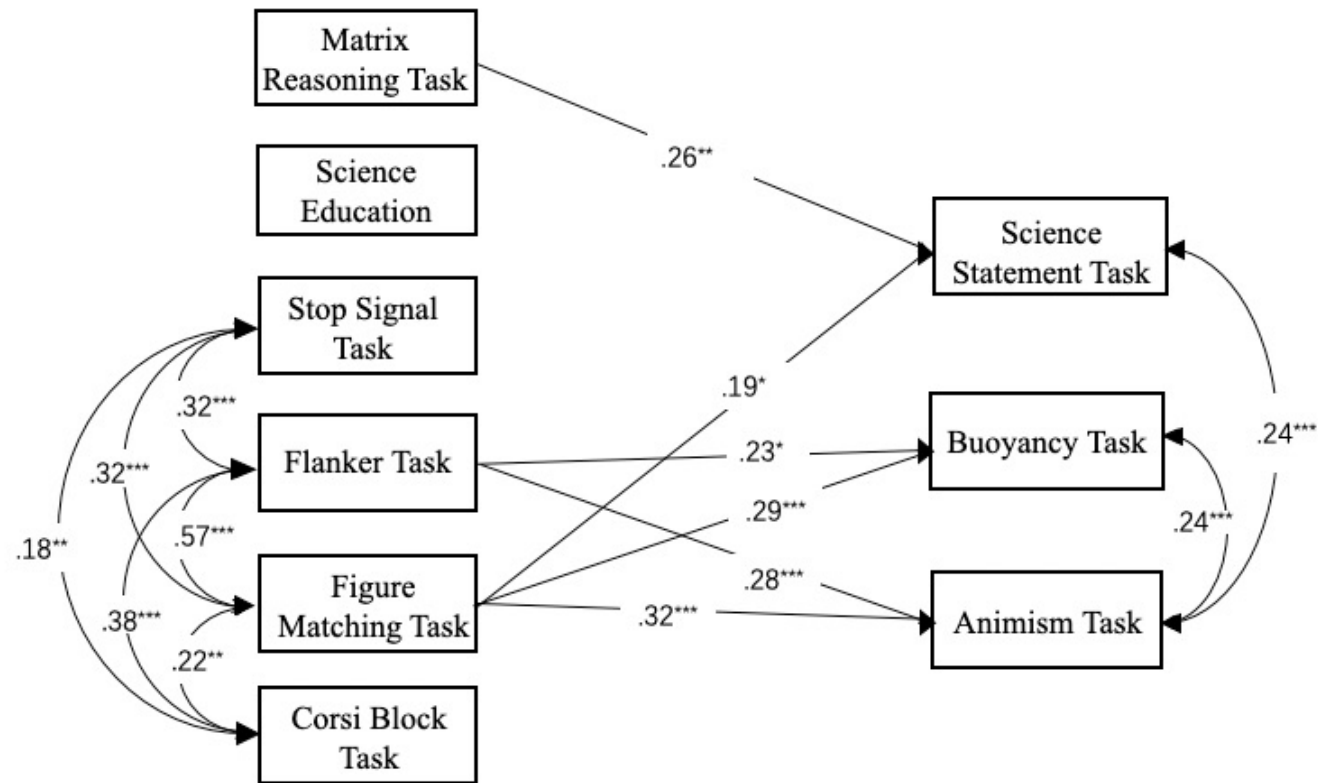
Path Analysis for Congruent trials on Efficiency (China)



Notes. Standardised Path Coefficients were reported. * $p < .05$; ** $p < .01$; *** $p < .001$.

Figure 6.2c

Path Analysis for Incongruent trials on Efficiency (China)



Notes. Standardised Path Coefficients were reported. * $p < .05$; ** $p < .01$; *** $p < .001$.

6.7 Discussion on Study 2

Extending upon Study 1, Study 2 explored the underlying mechanism of explanatory coexistence within a cross-cultural context. Study 2 addressed two key research questions. The following section also provides a detailed discussions of the findings as they correspond to this PhD thesis' overall study questions, as outlined in Chapter 3. Table 6.4 lists these research questions, as well as their links to Study 1.

Table 6.4*Summary of Analysis and Corresponding Discussion for Study 2*

Research Questions	Link with Study 1	Analysis Used	Results	Discussion Sections
Research Question 4: Are the performances of executive functions, general cognitive ability and explanatory coexistence similar among adult participants from the United Kingdom and China?	n.a.	Scatter Plots (Overall; Congruent; Incongruent trials; Figure 6.1a-6.1g)	Chinese participants generally spent more time on achieving higher accuracy when compared with participants from the UK	6.7.1 Response Pattern Between the United Kingdom and Chinese data

Table 6.4 continue

Research Question 5: Is the relationship between science education, executive functions, general cognitive ability and explanatory coexistence robust across the United Kingdom and China?	Research Question 1	Measurement Invariance (Overall; Congruent; Incongruent Trials; Table 6.1) Correlation (Overall trials; Table 6.2) and Path Analysis (Overall trials; Figure 6.2a)	There were inconsistencies in how behavioural inhibition contributed to explanatory coexistence, but results of the task modalities and trial types were similar across countries.	6.7.2.1 Differences of the Underlying Mechanism of Explanatory Coexistence Between the United Kingdom and China
	Research Question 2	Correlation (Overall trials; Table 6.1) and Path Analysis (Overall trials; Figure 6.2a)		6.7.2.2 Similar Findings When Considering Task Modality Between the United Kingdom and China
	Research Question 3	Correlation (Congruent trials & Incongruent trials; ; Table 6.2) and Path Analysis (Congruent trials & Incongruent trials; Figure 6.2b & Figure 6.2c)		6.7.2.3 Different Measurement Invariance Results Among the Two Trial Types Between the United Kingdom and China

6.7.1 Response Patterns of Participants from the United Kingdom and China

Regarding different response patterns to the various tasks from participants in the United Kingdom and China, Chapter 3 posited that comparing the single scores of cognitive abilities across countries could result in bias. By using efficiency scores, which comprehensively reflect both accuracy and response times, I was able to integrate the scatterplots depicting the relationship between response time and accuracy to showcase the differences and similarities in participants' response patterns.

In terms of executive functions and general cognitive ability, the scatterplot showed that among participants who achieved higher accuracy, Chinese participants generally took longer times on all four executive function tasks, as well as on the general cognitive ability task, when compared with participants from the United Kingdom. This response pattern might be due to different strategies adopted by participants across the two countries, with Chinese participants tending to ensure more accurate responses and United Kingdom participants tending to answer more quickly. This difference in response patterns could be impacted by cultural factors: for example, Chinese children are aware from a young age that their parents expect them to excel academically, and that their level of accuracy will be evaluated strictly on exams (Ng & Wei, 2020). Consequently, the Chinese participants in this study may have taken greater care when considering their answers for each trial. Asian parents have also been found to place greater emphasis on self-control and self-regulation than their Western counterparts (i.e., Roos et al., 2017), which may be a reason why Chinese participants were more willing to spend additional time contemplating their answers. Conversely, participants from the United Kingdom did not spend additional time to ensure correct answers. This difference in cognitive test strategies has been observed in other cross-cultural studies considering both accuracy and response times (Kelkar et al., 2013).

Notably, the measurement of executive function tasks in this research was adopted from Ellefson et al., (2017). These researchers found no statistically significant differences on the efficiency of executive function between the United Kingdom and Hong Kong (Chinese)

adults when using efficiency scores. Although this study did not conduct a direct comparison in its research on efficiency, the scatterplot of study results might explain why efficiency results appear to be similar among adults in both countries. As efficiency is influenced by both accuracy and response times, differences in response patterns might lead to similar efficiency outcomes. Still, it is worth noting that Ellefson et al.'s (2017) research is both a cross-cultural and cross-generation study, and that their adult participants were parents of child participants (M_{age} of United Kingdom participants = 44.06; M_{age} of Hong Kong participants = 43.84). Participants of this study were mostly undergraduate or postgraduate students (M_{age} of United Kingdom participants = 23.45; M_{age} of Chinese participants = 22.89). As executive functions generally peak in early adulthood, then stabilise before gradually declining from the ages of 30-to-40 and into older age (de Oliveira et al., 2012; Ferguson et al., 2021; Huizinga et al., 2006), results from the current study may better represent cognitive abilities among young adults in the United Kingdom and China.

Moreover, since previous cross-cultural research on executive functions has concentrated on children (see the review by Schirmbeck et al., 2020), the scatterplot results in this research provide more information on the interaction between accuracy and response times when answering cognitive tasks. Compared with previous research using accuracy to compare performance in executive function tasks and concluding that superior performance among children in collectivist cultures might overlook different response strategies (Chen et al., 1998; Oh & Lewis, 2008; Sabbagh et al., 2006), this study's results offer a more diverse interpretation of the similarities and differences in executive functions in a cross-cultural context.

Scatterplots showed that the overall response patterns of all three naïve thinking tasks across all trials indicated different response patterns in both countries. While Chinese participants generally spent longer response times and had a higher accuracy when compared with United Kingdom participants, differences were smaller when considering only participants who achieved higher accuracy scores, especially in incongruent trials. These response patterns may have been caused by two reasons. Firstly, Chinese participants

recruited for the current research had higher average levels of science education (see Section 4.1), and might have placed greater focus on the accuracy of their answers, spending longer times making responses even during simple trials. Second, according to OECD (2018), Chinese students typically spend more time studying than their United Kingdom counterparts, and previous studies have shown that extended periods of science education tend to improve accuracy in responses but not necessarily response time (Goldberg & Thompson-Schill, 2009; Shtulman & Harrington, 2016). Chinese participants may have a higher motivation to achieve correct answers due to the high expectations set by their parents from a young age, and the significant emphasis placed on scores within the Chinese society in general (Hau & Ho, 2012; Ng & Wei, 2020).

That said, in this study's incongruent trials, which require the suppression of naïve thinking and more cognitive processing while recalling learned knowledge, participants from both countries tended to spend more time contemplating to ensure accuracy. This pattern suggests that incongruent trials indeed require a greater amount of cognitive processing. While participants from the United Kingdom responded quickly in congruent trials, they also needed more time to contemplate their answers in incongruent trials. Thus, when complex scientific reasoning is required, both United Kingdom and Chinese participants might have exhibited similar strategies in responding to naïve thinking tasks.

Overall, the scatterplots reveal information about the response patterns beyond simple data comparison, showcasing the relationship between response times and accuracy for the participants from the two countries. These results extend upon the exploration of the relationships of executive functions, general cognitive ability, science education and explanatory coexistence, and have implications for future analysis from a speed-accuracy trade-off perspective. However, given the current limited understanding of cognitive abilities and explanatory coexistence within cross-cultural contexts, further verification and discussion are essential to fully grasp the reasons behind these varied results.

6.7.2 Relative Contributions of Executive Functions, General Cognitive Ability and Science Education on Naïve Thinking Tasks are Different Between the United Kingdom and China

In exploring the links between executive functions, general cognitive ability, science education and explanatory coexistence among Chinese and United Kingdom participants, a multigroup analysis was conducted (See Table 6.1). Its results showed significant differences in the relationship between all variables across two countries. Data from each country was then separated for analysis, and Table 6.4 answers the three research questions put forward in Study 1 with information from the Chinese dataset.

Differences and Similarities in the Underlying Mechanism of Explanatory Coexistence Between the United Kingdom and China. Results of the multigroup analysis indicated significant differences, suggesting that the contributions of science education, executive functions and general cognitive ability to explanatory coexistence differed between participants from the two countries. Correlation analysis revealed that, despite similar patterns between Chinese and United Kingdom participants, some differences still existed. First, correlation in Chinese participants indicated that science education did not correlate performance in the Buoyancy task, and interference inhibition did not predict performance in the Science Statement task. Moreover, working memory was found to be significantly correlated with performance in the Science Statement task among Chinese participants, but not among United Kingdom participants.

Regarding the path analysis, the only significant differences between the two countries were that behavioural inhibition did not predict performance in the Buoyancy and Animism tasks among Chinese participants across overall, congruent and incongruent trials. General cognitive ability did not predict any Buoyancy, but did predict the Science Statement, task among Chinese participants, while science education did not associate with all executive function components and general cognitive ability among Chinese participants.

As mentioned above, the main differences between the relative contributions of executive function, general cognitive ability and science education on explanatory coexistence were related to three contributors: behavioural inhibition, working memory and science education. These factors are each in the following paragraphs.

First, behavioural inhibition did not predict performance in any naïve thinking tasks among Chinese participants but did predict performance in the Animism and Buoyancy tasks among United Kingdom participants in overall and congruent trials. A possible reason for this difference may be the varying strategies adopted by the participants from the two countries in addressing the speed-accuracy trade-off. As discussed in Section 6.7.1., Chinese participants tend to take longer response times in order to ensure their answers' accuracy. Brookman-Byrne et al. (2018) have noted that individuals with better behavioural inhibition may take longer times on answers for the sake of accuracy and found that behavioural inhibition impacted response accuracy more significant than response times. In terms of the United Kingdom participants, this study's results did not align with Ellefson et al. (2024), who used similar executive functions and naïve thinking tasks to find that behavioural inhibition does not directly predict naïve thinking tasks, either in terms of response time or accuracy. These inconsistent results on the impact of behavioural inhibition may have been caused by the data quality collected from the two countries, as well as differences in sample size.

Second, working memory correlated in this study with performance in the Science Statement task among Chinese participants, but this correlation was not significant among United Kingdom participants. It is possible that, as both working memory and the Science Statement task were complex, Chinese participants spent more time answering these two tasks, resulting in lower efficiency. Consequently, the response pattern of the Chinese participants could better capture the role of working memory in processing the Science Statement task, as compared to participants from the United Kingdom, who responded faster.

Third, science education among Chinese participants did not predict performance in any naïve thinking tasks, nor did it correlate with other contributing factors. This finding differed from the results of United Kingdom participants, and contrasted with previous

research conducted with Chinese participants. Jin et al. (2019) reported that brain areas related to inhibition found in science experts were more active when compared to those among novices during the completion of naïve thinking tasks, a finding similar to other neuroimaging studies conducted in Western countries (i.e., Masson et al., 2014). As the impact of science education reflects more on a participant's accuracy than response time (Shtulman & Harrington, 2016), the non-significant link between science education and all other variables might be caused by the extended response time used by Chinese participants. This longer response time might influence efficiency and diminish the apparent impact of varying levels of science education.

In addition to examining whether significant links exist between the measured tasks, the effect sizes of these links were also compared, and are presented in Table 6.5. Effect sizes of most links among participants from the United Kingdom were similar, predominantly small to medium. Overall, results replicated those of Chapter 5, indicating that executive functions have the strongest predictive power for explanatory coexistence when compared with science education and general cognitive ability. Among the executive functions, switching demonstrated the most robust and strongest predictive power.

However, it is worth noting that the effect sizes of the links between switching and the three naïve thinking tasks were higher among United Kingdom participants (medium effect) than Chinese participants (medium effect). Similarly, the link between interference inhibition and the Animism task was stronger among United Kingdom participants (medium effect) than Chinese (small effect) participants. As previously discussed, this discrepancy may be attributed to a trade-off between accuracy and response time observed in the Chinese participants.

Table 6.5

Summary of effect size of significant links in United Kingdom and Chinese Participants

	Overall			Congruent			Incongruent		
	Science	Animism	Buoyancy	Science	Animism	Buoyancy	Science	Animism	Buoyancy
	Statements			Statements			Statements		
<i>United Kingdom</i>									
General Cognitive Ability	.21*					.08*			
Behavioural Inhibition		.10*	.16**		.11*	.18**			
Interference Inhibition		.42***	.20**		.42***	.15*		.36***	.16*
Switching	.40**	.35***	.36***	.33***	.32***	.40***	.38***	.35***	.35***
<i>China</i>									
General Cognitive Ability	.24***			.22***			.26**		
Behavioural Inhibition									
Interference Inhibition		.31***	.27**		.33***	.23*		.28***	.23*
Switching	.25***	.31***	.28***	.30***	.27***	.27***	.19*	.32***	.29***

Overall, these results from the correlation and path analysis indicate that the predicting role of switching, interference inhibition and general cognitive ability on explanatory coexistence remain relatively constant across the United Kingdom and China. These results align with previous studies suggesting a robust link between executive functions and general cognitive ability in cross-cultural contexts (Ellefson et al., 2020; Georgiou et al., 2020; Lan et al., 2011). Moreover, consistent with previous research (Ellefson et al., 2020; Georgiou et al., 2020), all executive functions have been found to be closely intercorrelated in the current study, collectively predicting performance on explanatory coexistence. The robustness of executive functions' predicting role to explanatory coexistences across the United Kingdom and China highlights that executive functions might be the key contributing factor for explanatory coexistence in science learning within diverse culture contexts.

Similar Findings When Considering Task Modalities Between the United Kingdom and China. As to whether task modality influences contributing factors involved in explanatory coexistence, Study 2 finds the impact of task modality to be consistent among participants in both China and the United Kingdom. In addition to the differences in behavioural inhibition, working memory and science education between China and the United Kingdom, both switching and inhibition were found to contribute to tasks that included pictorial stimuli, such as the Animism and Buoyancy tasks. In contrast, for the Science Statement task, which involved text stimuli, the main contributors are general cognitive ability and switching. Working memory and science education were not found to predict the performance of naïve thinking tasks directly.

Although results for executive functions, general cognitive ability and science education in the pictorial task were expected to be consistent across both countries due to pictorial stimuli's intuitive nature, correlation findings for the text-based task diverged slightly from previous research. This may be because Chinese children have

a stronger working memory capacity for retrieving sentences written in Chinese characters (Hanley, 2005), but it is worth noting that science statements presented in the task were relatively simple (in both English and Chinese) in meaning and short in length, and constructed with similar grammar (Shtulman & Valcarcel, 2012), rendering the role of working memory in semantic understanding less critical when compared with its role in processing conflicts between naïve and science thinking. Additionally, since participants in the current project were adults with mature executive functions, unlike children, this fact could help them process linguistic information more rapidly, likely diminishing working memory's contribution. Thus, the contribution of working memory in explanatory coexistence still requires further investigation.

Different Multigroup Analysis Results Among the Two Trial Types Between the United Kingdom and China. When separating the overall trials into congruent and incongruent categories, a significant difference was observed during path analysis regarding the congruent trials among Chinese and United Kingdom participants, while for incongruent trials there was measurement invariance across the two countries. These results suggest that, in situations with conflicts between naïve and science thinking, cognitive skills used by participants from both countries were consistent in incongruent trials only. In Section 6.7.1, it was stated that participants from both countries showed similar efficiency in incongruent trials, a finding that may indicate a similarity in the underlying mechanism between United Kingdom and Chinese participants.

Among Chinese participants, the predicting role of behavioural inhibition in explanatory coexistence was weaker, while the contributions of general cognitive ability and science education were unstable across trial types, although interference inhibition and switching consistently acted as contributing factors. In addition to the reasons related to response strategies, as discussed in Section 5.7.2, different results may also be due to the different sample sizes of Chinese and United Kingdom

participants, or, when combined, affect statistical outcomes. Notably, when combining data from the two countries, the impact of science education on explanatory coexistence was observed for the first time, suggesting that the lack of science education's observed impact in other analyses could be due to smaller data size. However, unlike previous findings showing science experts to be more accurate than novices in incongruent trials (Jin et al., 2019; Kelemen et al., 2012; Shtulman & Harrington, 2016), this study indicates that science education negatively influences the efficiency of incongruent trials. This finding may be due to the fact that science experts demonstrate greater activity in brain regions associated with inhibition (Jin et al., 2019), and thus spend more time arriving at the correct information. Nonetheless, this area of research requires further exploration.

When comparing the effect sizes of congruent and incongruent trials, effect sizes from Chinese and United Kingdom participants exhibited patterns consistent with those observed in the overall trials (see Table 6.5 and the discussion in Section 6.7.2). However, the associations between general cognitive ability and naïve thinking tasks differed between the two countries: while general cognitive ability demonstrated a very weak association with the Buoyancy task among United Kingdom participants when predicting congruent trials, it consistently predicted performance on the Science Statement task among Chinese participants. These findings from the Chinese participants align with those reported in previous studies (e.g., Chen et al., 2010; Ellefson et al., 2017; Oh & Lewis, 2008; Xu et al., 2020), which could be attributed to these participants' response patterns: Chinese participants may spend more time reflecting on the text- and reasoning-intensive science statements, an engagement aligned with cognitive demands typically associated with general cognitive ability. In contrast, results from the United Kingdom participants may reflect a mismatch in efficiency when completing the cognitive processes required for general cognitive ability assessments and those involved in the Science Statement task. Alternatively, the weak association between the Buoyancy task and United Kingdom participants'

general cognitive ability might explain these differences. Interestingly, no correlation was found between general cognitive ability and the Buoyancy task among United Kingdom participants, suggesting that discrepancies in the path analysis results might stem from factors such as sample size or non-normality in the data (Fritz et al., 2022). These findings highlight the need for further research to validate the observed patterns.

Chapter 7 General Discussion and Conclusion

The current PhD dissertation makes several key contributions to research on explanatory coexistence. First, the studies measure the roles of different executive function components in explanatory coexistence, as well as these contributing factors' interconnections. Second, by incorporating general cognitive ability and science education into its analysis, this research gains a deeper understanding of how these potential variables, pertinent to cognitive development and academic achievement in science, interact with executive functions in a way that impacts explanatory coexistence. Third, by utilising different modalities of naïve thinking tasks, this research confirms that task modality generally influences the involvement of different contributing factors in explanatory coexistence. Fourth, by separately exploring congruent and incongruent trials in naïve thinking tasks, this research provides a deeper understanding of the diverse involvement of the different contributing factors in tasks for both congruent and incongruent trials. Finally, by adopting a cross-cultural design and accessing data from both the United Kingdom and China, this research robustly identifies the underlying mechanisms of explanatory coexistence within a cross-cultural context, further revealing the potential impact of response patterns and strategies used by participants from different countries.

Before proceeding to a general discussion of this dissertation's key research findings, I first briefly summarise the research questions and results in Table 7.1. Additional exploratory research findings are also presented to explain the comprehensive use of efficiency in this research.

Table 7.1

Summary of This Dissertation's Studies, Research Questions, Main Research Results and Key Research Findings

Study and Research Questions	Main Research Results	Key Research Findings
Study 1: Research Question 1: What are the relative contributions of executive functions, science education and general cognitive ability to explanatory coexistence?	- Executive functions are the most significant contributors to explanatory coexistence when compared with general cognitive ability and science education -Switching is the most important contributor to all naïve thinking tasks. -Inhibition abilities (behavioural and interference) are also important as they contribute to some naïve thinking tasks (the Animism and Buoyancy tasks). -Working memory correlates with switching and inhibition but does not contribute to performance in any naïve thinking tasks. - General cognitive ability is involved in explanatory coexistence, but only in a certain task (the Science Statement task). - Science education does not significantly contribute to any naïve thinking task but correlates with some components of executive function.	Key research finding 1: Dynamics of underlying mechanism of explanatory coexistence
Study 1: Research Question 2: Do task modalities (i.e., pictorial versus text tasks) affect the involvement of executive functions, science education and general cognitive ability in explanatory coexistence?	- Modality influences the involvement of cognitive skills in explanatory coexistence - Switching contributes to tasks with both pictorial and text stimuli - Inhibition (behavioural and interference) contributes to tasks with only pictorial stimuli (the Animism and Buoyancy tasks) - General cognitive ability contributes to tasks with only text stimuli (the Science Statement task) - Working memory and science education do not directly contribute to tasks, regardless of task modality	Key research finding 2: Influences of modality on the reasoning of explanatory coexistence

Table 7.1 continued

Study 1: Research Question 3: Do executive functions, science education and general cognitive ability contribute differently to congruent and incongruent trials in explanatory coexistence?	1. Switching and inhibition consistently contribute to explanatory coexistence across congruent and incongruent trials. 2. Working memory and science education do not directly contribute to explanatory coexistence. 3. General cognitive ability only contributes to explanatory coexistence in congruent trials.	Key research finding 1: Dynamics of the underlying mechanism of explanatory coexistence
Study 2: Research Question 4: Are the response patterns of executive functions, general cognitive ability and explanatory coexistence similar among adult participants from the United Kingdom and China?	1. For both executive function and naïve thinking tasks, Chinese participants tend to have higher accuracy but longer response times when compared with United Kingdom participants.	Key research finding 3: Explanatory Coexistence Within a Cross-Cultural Perspective
Study 2: Research Question 5: Is the relationship between science education, executive functions, general cognitive ability and explanatory coexistence robust across the United Kingdom and China?	1. Significant differences in the multigroup analysis is observed in terms of the link between all variables among United Kingdom participants when compared to Chinese participants, except in incongruent trials. 2. There is a cross-cultural difference in terms of behavioural inhibition, which significantly predicts among United Kingdom participants but not among Chinese participants.	Key research finding 3: Explanatory Coexistence Within a Cross-Cultural Perspective

7.1 Key Finding 1: The Underlying Mechanism of Explanatory Coexistence

7.1.1 *The Involvement of Different Contributing Factors in Explanatory Coexistence*

Throughout Studies 1 and 2, the significant contribution of executive functions to explanatory coexistence was found when efficiency was used as a measurement metric in both the United Kingdom and China. While each executive function component played a different role in explanatory coexistence, their contributions were interconnected. Executive functions were more important than general cognitive ability and science education in terms involvement in the explanatory coexistence process. Although the contributions of general cognitive ability and science education were observed in some specific naïve thinking tasks, their involvements were not consistently stable. The following sections integrate the results of Studies 1 and 2 to provide a comprehensive discussion on these contributing factors' different roles.

Inhibition. As the most extensively studied executive function component in explanatory coexistence, inhibition had long been viewed as the most significant contributor to explanatory coexistence (see review by Mason & Zaccoletti, 2021). Results from both Studies 1 and 2 align with previous findings, indicating that inhibition positively contributes to explanatory coexistence (Brault Foisy et al., 2015; Brookman-Byrne et al., 2018; Dumontheil et al., 2023; Potvin et al., 2015; Vosniadou et al., 2018). Given the mutual links between inhibition and explanatory coexistence in correlation analyses, both interference and behavioural inhibition correlate positively with most three naïve thinking tasks (with small to large effect sizes) in the data collected from both the United Kingdom and China. These findings are similar to those found in behavioural studies among children and adolescents investigating inhibition and explanatory coexistence (Brookman-Byrne et al., 2018; Butterfuss & Kendeou, 2020; Vosniadou et al., 2018), as well as neuroimaging studies

demonstrating that brain regions relevant to inhibition are activated in trials necessitating the use of explanatory coexistence (Brault Foisy et al., 2015; Jin et al., 2019; Potvin et al., 2020; Zhu et al., 2019). The current research findings indicate that, even among adults, inhibition is required to help suppress naïve thinking, which is in line with what has been proposed about the explanatory coexistence process, that naïve thinking does not disappear but is instead suppressed (Shtulman & Valcarcel, 2012).

That said, path analysis showed slightly different findings, in terms of data from both countries. When correlating all executive functions, general cognitive ability and science education, both behavioural and interference inhibitions did not predict Science Statement task performance in either country. Moreover, only interference inhibition predicted Chinese participants' performance on the Animism and Buoyancy tasks, while both interference and behavioural inhibitions predicted the performance on the two tasks among participants from the United Kingdom. The absence of inhibition's contribution to the Science Statement task has also been found in Kelemen et al. (2012) and Ellefson et al. (2024), both of which were conducted among adults. A possible explanation that may account for this difference is that inhibition's role in solving complex explanatory coexistence reasoning may be less pronounced among adults. Brault Foisy et al. (2021) has found that inhibition plays a relatively minor role in adulthood when solving naïve thinking tasks when compared to children and adolescents. Adults usually have a richer science knowledge base when compared to children and adolescents, making inhibition's role less obvious among explanatory coexistence concepts that require complex reasoning. Moreover, inhibition may not be the sole executive function component contributing to explanatory coexistence, as studies have found it to be less involved when considered alongside other executive function components among young children (Grenell & Carlson, 2021), children (Vosniadou et al., 2018) and adults (Ellefson et al., 2024).

Results from the current study did not find differences regarding the significant contribution of behavioural (with small effect sizes) and interference inhibitions (with small to medium effect sizes) to explanatory coexistence in the United Kingdom dataset, but differences were observed in participants from China, as only interference inhibitions (with small to medium effect sizes) significantly contribute to explanatory coexistence. Some studies note that these two inhibitions play different roles in explanatory coexistence, according to analyses using response time and accuracy metrics. Brookman-Byrne et al. (2018) note that better interference inhibition predicts shorter response times, while better behavioural inhibition predicts higher accuracy among participants from the United Kingdom. Ellefson et al. (2024) have also found that interference inhibition only directly predicts Animism task response time, while behavioural inhibition does not directly predict the accuracy or response time of any naïve thinking tasks among United Kingdom participants.

In the current study, efficiency scored showed variances between countries that led to different outcomes. It is possible that both forms of inhibition might contribute to more efficient responses to naïve thinking tasks, albeit gauged using different metrics (i.e., higher accuracy for Chinese participants, quicker responses from United Kingdom participants). Therefore, when calculating efficiency, the contributions of the two types of inhibition varies between countries. Nevertheless, interpretations of the current findings may be diverse, and will thus be interpreted from modality and cross-cultural perspectives in later sections.

Switching. While switching has been less researched than inhibition, some research indicates that switching may also be involved in explanatory coexistence (see Chapter 2). This study's research results, from both countries, aligned with previous findings and found switching to be involved in explanatory coexistence, both through correlation and path analyses.

Most studies supporting the connections between switching and explanatory coexistence have been conducted among children and via correlation analysis (Bascandziev et al., 2018; Grenell & Carlson, 2021; Vosniadou et al., 2018; Zaitchik et al., 2014). For example, Vosniadou et al. (2018) conducted the first study separating switching and inhibition, and found that both skills contributed differently to explanatory coexistence among children. Although their results mainly relied on correlation and regression analyses, they found that inhibition linked more with naïve thinking task accuracy, whereas switching linked with both accuracy and response times in tasks. This association between switching and explanatory coexistence was found to hold true irrespective of whether switching was included within the overall scores for executive functions (Bascandziev et al., 2018; Grenell & Carlson, 2021; Zaitchik et al., 2014), or in separate switching tasks (Vosniadou et al., 2018).

More recently, Ellefson et al. (2024) have found some distinctions in the contribution of inhibition and switching through path analysis, identifying switching as a key contributing factor in naïve thinking tasks. These researchers have therefore proposed that switching may contribute not only to the suppression of naïve thinking, but also to the activation of later learnt science thinking. Moreover, they note that, in comparison to the direct effects of switching, both behavioural and interference inhibitions seem to play an indirect role in explanatory coexistence, as these correlate with switching and thus contribute to explanatory coexistence through their link to switching. Consequently, the researchers argue that inhibition and switching are not independent of each other but work together when needed (i.e., to solve naïve thinking problems).

The current research results have replicated those of Ellefson et al. (2024), noting the significant role of switching in explanatory coexistence in all naïve thinking tasks among both countries (with small to medium effect sizes) and further supporting their view of the cognitive process under explanatory coexistence. This view denotes a distinction between inhibition and switching, with task switching

requiring inhibition of irrelevant information and the activation of new tasks (Ellefson et al., 2024; Monsell, 2003). Thus, it is likely that merely inhibiting naïve thinking is insufficient during the process of explanatory coexistence, and that participants must also activate the scientific thinking they acquire later in life. While inhibition aids in the inhibition process, activating science thinking makes switching relatively more important. Therefore, inhibition assists switching in explanatory coexistence, which may also explain why, when correlating switching with inhibition, switching is a stronger predictor of naïve thinking task performance. Participants with better switching ability could consequently reject naïve thinking and activate science thinking more rapidly.

Working Memory. Working memory has received less attention than inhibition and switching in previous studies. In this research, correlation results from both the United Kingdom and China indicated that working memory positively correlated with the Animism and Buoyancy tasks. While working memory correlated with the Science Statement task among Chinese participants, this link was not significant among United Kingdom participants. These results were similar to those found in studies showing a significant correlation between working memory and naïve thinking tasks among children, even if these studies used combined executive function scores that did not distinguish working memory's role (Zaitick et al., 2014; Grenell & Carlson, 2021). Thus, the current study extends upon previous studies, revealing the mutual links between working memory and the performance of some naïve thinking tasks.

Path analysis, however, found that working memory did not predict any naïve thinking tasks directly in either country, although it correlated with all other executive functions. These results suggest that working memory may influence performance on naïve thinking tasks through other executive functions. Given this study's use of the Corsi Block task, a visuospatial working memory task, its results aligned with

Dumontheil et al.'s (2023) findings via multiple regression. Using a similar visuospatial working memory task (the Frankie task, Morris et al., 2019), Dumontheil et al. (2023) suggested that visuospatial working memory has limited impact on explanatory coexistence when compared with verbal working memory. Moreover, when accounting for both visuospatial and verbal working memory alongside other executive functions, both working memory tasks made no significant contribution to explanatory coexistence. Such results might indicate that, although working memory is involved in the explanatory coexistence processes, its influence is weaker when compared with other executive function components. Individuals with stronger working memory may better hold and manipulate both naïve and science thinking more effectively, supporting the switching between these two types of thinking more robustly. Thus, the weaker role of working memory may be due to its strong link with other executive function components. However, due to limited research on working memory within the context of explanatory coexistence, its exact role in this process cannot be identified, and future studies are needed to explore this factor's different effects.

General Cognitive Ability. Results from correlation and path analyses indicated that general cognitive ability mainly associated with participants' performance on the Science Statement task (with small effect sizes) in both the United Kingdom and China, and might weakly associated with congruent trials in Buoyancy task (with weak effect size) in the United Kingdom. One possible reason why general cognitive ability did not correlate with Animism and Buoyancy task performance is the current study's efficiency metrics, unlike most existing research on this topic, which has tended to use accuracy scores of general cognitive ability tests as their main measurement (i.e., Vosniadou et al., 2018), especially in research conducted on young children (i.e., Zaitchik et al., 2014). Dumontheil et al. (2022) incorporated response time into their analysis of the Science Statement task and found that, while

the accuracy of the general cognitive ability test positively predicted performance on naïve thinking tasks, response time negatively predicted performance on naïve thinking tasks. Given that the Animism task (animism thinking) and the Buoyancy task (buoyancy thinking) each involved only one type of naïve thinking, trials in both tasks required only a short response time to make a correct answer, which thus yielded a high accuracy rate. Consequently, the lack of significant link between general cognitive ability and simple naïve thinking tasks may also be due to the use of efficiency scores combining accuracy and response time.

Another finding from this study's path analysis from both countries indicated that general cognitive ability did not correlate with any components of executive functions. This finding is different from what was reported in previous studies (Ackerman et al., 2005; Alloway & Alloway, 2009; Ardila, 2018; Blair, 2006), which may be due to the fact that this research's participants tended to take greater amounts of time giving more accurate answers on general cognitive ability test. This strategy may have made variations in response time greater when compared to executive function tasks requiring less time to answer. This significant difference in response time could impact the relationship between the efficiency of the general cognitive ability task and that of all other measurements.

Appendix O supplements this study's path analysis results by presenting the accuracy scores of general cognitive ability and the efficiency scores of all other tasks. Results show that although general cognitive ability only predicts performance on the Science Statement task among United Kingdom participants, it predicts on all naïve thinking tasks among Chinese participants. Moreover, consistent with other literature on the relationship between general cognitive ability and executive functions (Ackerman et al., 2005; Alloway & Alloway, 2009; Ardila, 2018; Blair, 2006), when accuracy was used for general cognitive ability, it correlated with the efficiency of all executive functions in both countries. These results indicate the need to further explore the impact of measurement metrics on task response.

Science Education. Science education, though found to be correlated to performances in some naïve thinking tasks, was not found to directly contribute to any naïve thinking task in this study's path analysis. One possible reason for such results is that the current study's naïve thinking tasks were relatively simple and straightforward, and participants only needed to use basic cognitive abilities, like executive functions and general cognitive ability, to suppress naïve thinking, rather than have a significant accumulation of scientific learning experience. On the other hand, results also indicated that science education might influence explanatory coexistence by improving executive functions. As shown in some studies, the brain regions of science experts related to inhibition have been found to be more active when compared with those of science novices when completing naïve thinking tasks, suggesting that science education might not influence explanatory coexistence in a direct way (Brault Foisy et al., 2015; Jin et al., 2019; Potvin et al., 2020; Zhu et al., 2019). Furthermore, Shtulman and Harrington (2016) have found that the impact of science education is more reflective on accuracy than response time when completing naïve thinking tasks. Thus, the effect of science education on explanatory coexistence may not be apparent in terms of efficiency.

Notably, path analysis of the United Kingdom participants in this study showed science education to positively correlate with inhibition and working memory. Such a correlation might suggest that scientific education allows individuals to more easily activate knowledge via working memory, helping them inhibit naïve thinking more effectively. However, no such correlation was found in Chinese subjects, further indicating that research treating science education as a continuous variable requires more exploration.

7.1.2 Overview the Underlying Mechanism of Explanatory Coexistence

Overall, when considering all contributing factors in this research, results confirm the contributing role of executive functions in explanatory coexistence when

using efficiency as metric, especially that of switching and inhibition (Bascandziev et al., 2018; Ellefson et al., 2024; Grenell & Carlson, 2021; Vosniadou et al., 2018). The research also fills a research gap regarding the role of working memory. Switching appears to be the most important contributor to explanatory coexistence when compared with the other executive function components, as it robustly predicts explanatory coexistence across all naïve thinking tasks (the Animism, Buoyancy and Science Statement tasks), different trial types (congruent and incongruent) and different countries (the United Kingdom and China). Inhibition also predicts certain forms of explanatory coexistence, although the involvement of behavioural and interference inhibitions is different across countries.

General cognitive ability seems to predict science learning within explanatory coexistence in congruent trials, rather than convert naïve thinking to science thinking in complex naïve thinking task. Moreover, working memory and science education appear to support the other two executive functions in predicting explanatory coexistence. Thus, general cognitive ability seems to be required to facilitate the initial learning of scientific concepts, then afterwards executive functions can undertake comparison and selection between naïve thinking and science thinking. Science education may also be necessary to ensure the continuous reinforcement of correct science thinking, but its role requires further investigation. However, the most crucial underlying mechanism helping individuals solve conflicts in explanatory coexistence is executive functions.

Previous studies on the distinction between inhibition and switching indicate that switching involves not just the inclusion of inhibition but also the activation of newer tasks. This implies that switching also encompasses the capability of inhibition (Monsell, 2003; Rogers & Monsell, 1995; Yeung & Monsell, 2003a), especially in explanatory coexistence (Ellefson et al., 2024). Moreover, when switching between two different task sets, a higher load on working memory is required, as it needs to maintain two types of thinking in an active state, rather than just one (Los, 1996). The

uncertainty of which thinking type will be selected next also contributes to the cognitive load (Poljac et al., 2009).

Based on this research's findings, I propose a dynamic underlying mechanism that slightly differs from previously suggested patterns of executive function involvement. Previous studies, such as those by Vosniadou et al. (2018b), indicate that working memory is required to retain both naïve and science thinking and use switching when comparing these types of thinking, meaning that inhibition plays a key role in rejecting naïve thinking. This study proposes an underlying cognitive mechanism to explanatory coexistence in which individuals with better switching abilities not only inhibit unnecessary information more quickly but also activate newer science thinking more efficiently. Moreover, while switching functions during the processing of explanatory coexistence, working memory helps to keep both naïve thinking and science thinking in current attention, while inhibition limits unnecessary naïve thinking, both of which contribute to more efficient switching between naïve and science thinking. Therefore, in situations requiring the resolution of conflicts between naïve and science thinking, switching benefits from and coordinates both inhibition and working memory. This mechanism could also explain why inhibition and working memory, when measured alone, significantly correlate with explanatory coexistence, but lack a significant predictive role when measured alongside switching.

In sum, the two studies in my PhD provide some explanations for why executive functions are crucial for explanatory coexistence and scientific learning.

7.2 Key Finding 2: Influences of Modality on the Reasoning of Explanatory Coexistence

7.2.1 *Links Between Different Contribution Factors and Modality*

This research's second key finding reiterated that when task modality varied, the contributing factors involved differed as well. However, results were consistent in terms of the links between different contributing factors and task modality across both countries, as discussed in previous studies about task modality (Ellefson et al., 2024; Stricker et al., 2021).

Switching Involved in All Task Modalities. Results found switching to be involved in both picture-based and text-based tasks, across different trial types and countries, a finding consistent with previous research (Ellefson et al., 2024; Vosniadou et al., 2018). This suggests that switching contributes significantly to task performance, irrespective of various task modalities.

Inhibition Involved in Picture-Based Tasks. Inhibition was only involved in picture-based tasks. Such results differed from previous studies examining child and adolescent participants, which have found significant relationships between inhibition and text-based naïve thinking tasks (Brookman-Byrne et al., 2018; Vosniadou et al., 2018; Wilkinson et al., 2019). One possible explanation for this difference may be the varying age groups involved in these studies: studies reporting a significant relationship between inhibition and explanatory coexistence with text stimuli have been mostly conducted among children (Vosniadou et al., 2018; Wilkinson et al., 2019) and adolescents (Brookman-Byrne et al., 2018), and their Science Statement tasks have incorporated both pictorial and text stimuli to help younger participants better understand these trials' science concepts.

Another possible reason may be the similarity of modalities between pictorial stimuli in the interference inhibition task (the Flanker task); pictorial and audio stimuli in the behavioural inhibition task (the Stop Signal task); and the pictorial stimulus in naïve thinking tasks (the Animism and Buoyancy tasks). Such similarity in task modalities may evoke

similar cognitive functions, as all involve pictorial stimuli and thus require less time to respond. In contrast, naïve thinking tasks with text stimuli requiring reading are more complex, and more cognitive abilities (i.e., switching, general cognitive ability) may be involved during processing texts.

General Cognitive Ability Involved in Text-Based Tasks. General cognitive ability was only involved in text-based tasks. Although general cognitive ability was assessed via picture-based tasks, it required the consideration of more reasoning processes. A possible explanation for the lack of a direct link between general cognitive ability and naïve thinking tasks may be that picture-based naïve thinking tasks were simple, so participants did not need to reason about different scientific concepts; rather, these participants only had to perform simple inhibition and switching within the same concept. The Science Statement task, on the other hand, contains scientific information from different domains (Shtulman & Valcarcel, 2012), requiring participants to have a wider understanding of scientific knowledge and make decisions in different science conflicts. Consequently, answering the Science Statement task requires both switching (to suppress naïve and activate science thinking) and general cognitive ability (to understand complex concepts and sentences). Additionally, when considering the association between general cognitive and reading abilities (i.e., Ahr et al., 2016; Lynn & Mikk, 2009), it is possible that general cognitive ability mainly affect tasks requiring reading skills.

Science Education and Working Memory Not Involved. Science education and working memory, however, were found not to directly contribute to explanatory coexistence when considering other contributing factors in this PhD research. As mentioned in the previous key findings, this might be because the contributions of these two factors are not as significant as others. For example, in the Science Statement task, which requires reading skills, individuals with slower reading speeds might experience reduced efficiency, which can impede their ability to demonstrate their strength in science during task performance.

Working memory may be more involved in different forms of explanatory coexistence through its contributions to inhibition and switching.

7.2.2 Views from the Modality Perspective

From the modality perspective, pictorial stimuli provide more complete information. As discussed in Chapter 2, presenting tasks through pictorial stimuli is a more direct method of information delivery and easier for participants to respond and reason (Bourisseau et al., 1965; Ward, 1994). Therefore, when delivering explanatory coexistence concepts via pictorial stimuli, both the suppression of naïve thinking (inhibition) and the activation of science thinking (switching) are easier to measure. In contrast, presenting tasks through text stimuli allows participants to spend more time understanding the semantics of science statements (Bourisseau et al., 1965; Lehmann & Seufert, 2020). Thus, when delivering explanatory coexistence concepts via text stimuli, the extended reading time supports switching between naïve and science thinking and requires general cognitive ability to reason through the abstract concepts presented in text statements. Therefore, while the processing of both pictorial and text stimuli requires prior knowledge, pictorial stimuli could provide more intensive information when compared to text (Adams & Campbell, 1999; Carney & Levin, 2002; Mayer et al., 2001; Miller & Stine-Morrow, 1998; Moreno & Mayer, 1999). As the Animism and Buoyancy tasks contain only pictorial stimuli with one concept presented repetitively, it is easier for participants to respond to these two tasks. In contrast, the Science Statement task encompasses various scientific concepts expressed through text, which lacks the ability to provide comprehensive information, making it harder for participants to respond to this task. These differences may be the reason why different skills are involved across different task modalities.

In sum, this study has found that across different countries (the United Kingdom and China) and in various trials (congruent vs. incongruent), task modality influences the underlying mechanisms of explanatory coexistence. Various contributing factors are involved when explanatory coexistence is presented via pictorial or text formats. From a modality

perspective, the inherent differences in information perception enabled by text and picture stimuli may be the reason for this varying involvement. Thus, apart from simply comparing the task modality of cognitive ability and naïve thinking tasks, as well as similarities or differences in response patterns, it is also important to consider how modality helps convey messages.

7.3 Key Finding 3: Explanatory Coexistence Within a Cross-Cultural Perspective

7.3.1 Response Patterns in Cross Cultural Research

Study 2 of the current research provides a cross-cultural perspective on the underlying mechanisms influencing explanatory coexistence. Its results indicate that Chinese participants exhibit different response patterns in executive functions, general cognitive ability and explanatory coexistence from United Kingdom participants, as Chinese participants spend a longer amount of time on tasks in order to make correct responses. This differences may result from different speed-accuracy trade-off strategies adopted by participants from the two countries. This finding offers a new angle for considering the development of cognitive abilities and science learning across cultures.

For a long time, some large-scale international assessments (i.e., PISA, TIMSS) have shown that East Asian children and adolescents perform differently than their global counterparts in terms of achievement in certain academic domains (OECD, 2022; TIMSS, 2019). Various studies have offered explanations for this finding, both from cultural and cognitive perspectives (see Chapter 3). For example, some studies argue that Asian children develop their executive functions earlier than Western children due to the higher demands for self-regulation within collectivist cultures, as opposed to the emphasis on self-expression found in individualistic cultures (Ellefson et al., 2017; Oh & Lewis, 2008; Xu et al., 2020). While some studies on adults have found similarities in cognitive abilities across countries, the understanding of cognitive skills in cross-cultural context remained limited (Ellefson et al., 2017; Kelkar et al., 2013).

Observing from different metrics, this study further demonstrates how adults' cognitive skills and science measurement outcomes are a matter of different response patterns. Chinese participants in this study were to be more willing to spend time

ensuring accuracy, which was not the case for United Kingdom participants. These results align with expectations from Eastern countries that students will have a greater focus on academic achievement and a stronger work ethic (Hau & Ho, 2012; Hsin & Wu, 2011; Lau, 1996). Chinese children are often aware of their parents' expectations of academic excellence from a young age (Zhang et al., 2019), which makes them more likely to follow task instructions carefully. In contrast, Western students may not be as stringent in pursuing academic outcomes, instead focusing more on independent thinking and self-achievement, with personal feelings becoming a more important factor in education (Aikenhead & Jegede, 1999; Greenfield et al., 2006; Tobin & Hayashi, 2011). Thus, participants from the United Kingdom may not have been willing to spend as long a response time on their tasks ensuring accuracy as their Chinese counterparts.

Taken together, the different response patterns between the two countries indicate that, as results interpretation can vary when viewed from different metrics, no absolute conclusion can be made about superior performance. When judging solely by accuracy, Chinese participants seem to perform better on tasks, but this view overlooks the significant effort and deliberation Chinese participants invest in these tasks; in situations requiring quick decision-making, this excessive focus on accuracy could lead to bias. Similarly, when viewed solely from the perspective of response time, participants from the United Kingdom appear to outperform their Chinese counterparts, but this overlooks their choice of a more effortless decision-making approach. In contexts where accuracy is more important, rapid selection may lead to bias. An overall observation of response patterns suggests that, when comparing cognitive abilities in cross-cultural contexts, it may be more appropriate to focus on cultural influences rather than comparing single scores.

7.3.2 Robust Links Between Cognitive Skills and Explanatory Coexistence

Although this study finds that certain executive function components (i.e., behavioural inhibition) contribute differently to explanatory coexistence, the overall involvement of executive functions, general cognitive ability and science education in explanatory coexistence is robust in both the United Kingdom and China, especially in incongruent trials with conflicts between naïve and science thinking. Despite differences in cognitive task response patterns, results from both studies showed the link between cognitive skills and explanatory coexistence to be similar across cultures. These results also align with previous findings exploring the link between cognitive skills and academic performance in cross-cultural contexts (Ellefson et al., 2017; Georgiou et al., 2020; Lan et al., 2011; Thorell et al., 2013).

Certain executive functions, particularly switching and interference inhibition, have been identified in this research as crucial predictors for explanatory coexistence, and are robust across different countries and task modalities. Beyond an understanding of scientific knowledge, the unique contribution of executive functions to science learning may be partly reflected in the process of explanatory coexistence (Carey, 2000). Specifically, when naïve thinking conflicts with science thinking, the suppression of naïve thinking and activation of science thinking could represent a common aspect of human cognitive thought when dealing with explanatory coexistence (Vosniadou, 2019). Moreover, consistent with previous research that found a close link among different executive function components across cultures (Ellefson et al., 2020; Georgiou et al., 2020), all executive functions are closely intercorrelated in the current study, either directly or indirectly, predicting the performance of explanatory coexistence. This predicting role may be the key underlying mechanism of explanatory coexistence in science learning across countries.

7.4 Exploratory Finding: Story of Efficiency

The previous sections discussed several results of this research that differed from previous literature, potentially due to the use of efficiency in this PhD research. Research on efficiency in studies relating to explanatory coexistence has emerged in recent years, yet remains relatively limited (Young & Shtulman, 2020). Therefore, although using efficiency was not in my original research plan, I discuss it as part of my exploratory findings in this section.

First, efficiency appeared to be a more comprehensive indicator of the relationship between cognitive skills and explanatory coexistence than either accuracy or response time alone. Given the varying modality of tasks, accuracy or response time measurements alone may not accurately reflect an individual's performance. For example, the Science Statement task requires reading sentences, which may require a broad knowledge base, and accuracy for this task tends to be lower than that of the Animism and Buoyancy tasks; response times for the Science Statement task also tend to be longer. Similarly, tasks that require more reasoning, such as the Matrix Reasoning and Corsi Block tasks might also lead to differences in accuracy and response times when compared to other tasks. Thus, considering inherent task differences, adopting efficiency can considerably address these variations, better reflecting participant responses.

Second, efficiency may better demonstrate similarity in response patterns across tasks. Efficiency offers advantages over the sole use of accuracy or response time alone, as accuracy may not reflect the precise reasoning processes involved in addressing problems, while response time may not reflect outcomes after the reasoning processes (Bruyer & Brysbaert, 2011; Miyake et al., 2000; Rogers & Monsell, 1995). Thus, efficiency may better integrate both metrics, producing an indicator that more accurately reflects the correct thinking process. Especially in cases of simpler executive function tasks prone to ceiling effects (Ellefson et al., 2020), efficiency can better reflect speed-accuracy trade-offs in adult participants, offering a

more comprehensive view of the test results.

Third, efficiency provides an alternative angle for discussing cross-cultural research, bridging biases and interpreting culture's influence on cognition and learning through response patterns. Currently, much cross-cultural research on executive functions or science learning outcomes tends to focus on accuracy comparisons (Cho et al., 2023; OECD, 2022; Oh & Lewis, 2008; Thorell et al., 2013), perhaps because these studies often involve child participants whose response times may be unpredictable due to their age; consequently, accuracy is a more reliable indicator for reflecting these children's cognitive processes and reactions (Ackerman & Koriat, 2011). Although previous studies on adults have noted differences in accuracy or response time across different countries (i.e., Kelkar et al., 2013), such research remains limited. This PhD research indicates that efficiency score may capture some of the commonly used response patterns of participants from different countries when completing cognitive and naïve thinking tasks. Efficiency can integrate both the time taken to make strategic decisions and the accuracy of final choices, so as to better reflect differences in thinking patterns and response strategies for participants from different cultures. Therefore, efficiency may be considered more extensively in future cross-cultural research.

Finally, although using efficiency offers advantages, the current method of calculating efficiency (Ellefson et al., 2017), based on executive function tests, may be highly influenced by the speed-accuracy trade-off. For complex tasks that require more from the thinking process, such as the Corsi Block task (for working memory) or the Matrix Reasoning task (for general cognitive ability), the trade-off effect might influence whether participants focus more on accuracy or speed, biasing efficiency results. However, overall, this study's research findings are still broadly consistent with previous studies analysing findings using response time and accuracy.

7.5 Limitations and Future Directions

This PhD research has explored various contributing factors to explanatory coexistence within a cross-cultural context, yet its novelty may result in numerous unanswered questions. This section outlines some of this research's potential limitations and discusses possible future directions for this field of research.

7.5.1 Limitations on Age Representation

This research's findings have been limited to adult participants. As mentioned in Chapter 2, the framework proposed in both studies was relatively new, so adults participants whose executive function and general cognitive abilities were in a stable, mature phase of development were selected (Bedard et al., 2002; Best & Miller, 2010; Diamond, 2013; Miyake et al., 2000). Previous studies have found that inhibition plays a more prominent role in explanatory coexistence during childhood and adolescence, then stabilises in adulthood (Brault Foisy et al., 2021). Apart from inhibition, working memory, switching and general cognitive ability have also been found to be involved in explanatory coexistence among children (Bascandziev et al., 2018; Vosniadou et al., 2018; Zaitchik et al., 2014). As executive function and general cognitive ability develop with age (Best & Miller, 2010), and most fundamental scientific learning occurs before adulthood (Carey, 1985, 2000; Duit & Treagust, 2003), it is difficult to claim that this research's results could be replicated among children and adolescents due to differences in participants' cognitive development. Future research that include findings from adolescents and children would therefore offer valuable insights into the developmental dynamics of the mechanisms underlying explanatory coexistence.

7.5.2 Overlaps Between Executive Function Measurements

The measurement of distinct executive function components may be influenced by impurity problems due to substantial overlaps among the different executive functions skills in executive function tasks, as well as the questionable ecological validity of laboratory-based cognitive tasks (Chaytor et al., 2006; Moissala et al., 2018; Wang et al., 2021). For example, some of the tasks used in previous research on explanatory coexistence, such as the Stroop (Brookman-Byrne et al., 2018; Vosniadou, Pnevmatikos, Makris, et al., 2018) and Wisconsin Card Sorting (Thibault & Potvin, 2018) tasks, have been used to measure inhibition, while elsewhere they have measured switching (Miyake et al., 2000; Monsell, 2003). During the study design for this research, although the potential overlap between different executive function measurements was considered in order to select comparatively ‘purer’ measurements for each executive function component (see Appendix A.3.2), the current research’s tasks could still involve the influence of overlapping executive function components.

Therefore, given the variety of tasks used to measure executive functions (i.e., the Stop Signal task, the Flanker task for measuring inhibition), and the fact that some tasks may apply to different executive function components (i.e., the Stroop task used for both inhibition and switching), one cannot conclude that results from a particular task exclusively measure or represent a single executive function (Chan et al., 2008). Future research may consider using a more diverse array of executive function tasks, such as extending upon this research’s finding that switching plays the most important role in explanatory coexistence. For my next research project, the current research experiment will be replicated using multiple tasks to measure switching, which will help uncover whether the findings on switching from this study are robust, or are present only when using the Figure Matching task (used in the current study).

7.5.3 Bias During Cross-Cultural Data Collection

Certain unavoidable factors may have impacted this research's results during cross-cultural data collection, despite the fact that previously verified tasks were carefully selected (Ellefson et al., 2017) for the current research, and necessary steps taken to ensure the fidelity of translation (see Appendix A.4). First, as mentioned in Appendix A.5, the online PsychoPy platform occasionally experienced download delays in China due to the country's firewall. This problem may have impacted the response time of Chinese participants, influencing their efficiency scores. However, following the completion of this research's data collection, which was completed between 2021 and 2022, some local online PsychoPy platforms in China (i.e., Nao-Dao) were developed and became compatible with local networks (<https://www.naodao.com/>). Future research could use more suitable platforms such as Nao-Dao to access more accurate measurements on response times.

Apart from the possible influence of internet problems, the varying reading speeds of different languages should also be considered. The reading speeds of English and Chinese are different (Lü & Zhang, 1999), and may influence the response time and efficiency of text-based tasks (i.e., the Science Statement task). Although z scores could potentially resolve this problem, the use of more diverse measurements, including response time and accuracy, to collectively interpret explanatory coexistence's underlying mechanism may be necessary in cross-cultural research.

Finally, some scholars argue that cultural differences mean that the measurements of executive functions and general cognitive abilities cannot be generalised (see Chapter 3.2.1 & 3.2.2). In the current study, participants from different countries might not be equally familiar with the tasks used in this research. While some of these tasks have been used in cross-cultural contexts (Ellefson et al., 2017), culture's impact on this research's experimental results are hard to measure. Thus, future research should consider the possible impact of culture difference by

utilising more diverse measurement tools, and including a broader range of countries and regions to verify these research results.

7.5.4 Limitations of the Form of Task Modality

This research found that task modality effects interpretations of the involvement of different cognitive abilities in explanatory coexistence (Ellefson et al., in preparation; Stricker et al., 2021). Due to various constraints such as experimental materials, research funding and PhD study completion time, the current research only used tasks consisting of picture- and text-based stimuli. These stimuli were intuitive and simple, but naïve thinking may manifest in various ways in real life (Carey, 1985; Chi, 2005; Inagaki, 2001), particularly with the advancement of the internet and artificial intelligence, where different modalities such as audio, images, videos and virtual reality are involved in educational environments (Boulton et al., 2018). Future research might expand upon the measurement formats of both cognitive abilities and naïve thinking tasks, while using tasks with wider modalities (i.e., including audio) to enhance the understanding of the dynamic modality of science reasoning in real-life learning and teaching situations.

7.5.5 Limitations on Science Educations Measurements

The current research has found that science education does not directly predict explanatory coexistence across both countries when measured through task efficiency, yet these results may have been affected by several limitations in experimental design. First, science education was assessed by asking participants of their experiences receiving science education. As some participants who studied humanities subjects in university might not have received science education after high school, their memory of scientific knowledge may be more vague than university science major students (Bjork & Bjork, 2019). Previous research on experts and novices has often involved

undergraduate students (Jin et al., 2019) or professors (Potvin et al., 2020), presenting a more direct contrast in science exposure. Additionally, the science education questionnaire used in the current study lacked a measurement for science education duration, as participants had not received science education. This oversight might be a reason why the current research results are inconsistent with previous findings on the role of science education in explanatory coexistence.

7.5.6 Limitations on Non-verbal Reasoning Tasks

Most of the executive function tasks used in the current research were non-verbal, as reducing the influence of verbal or numerical demand appears to be a better fit for cross-cultural research (see Appendix A.3.2). Differences in familiarity with and preference for non-verbal tasks among individuals with different majors may be associated with their ability or motivation to solve abstract information, which relates to math (Tsigaridis et al., 2022). Thus, some participants may prefer non-verbal tasks, and therefore be more willing to spend time achieving high accuracy, whereas others might struggle with non-verbal tasks, leading to less response time spent and lower accuracy. However, as the connection between executive functions and science learning remains unclear, future research designs should consider balancing task selection.

7.6 Practical Implications

This research's main focus is on cognitive science, not educational practice *per se*, meaning that while its findings may have some practical implications, these implications are far removed from the possibility of changes to the classroom in the near term. That said, these results may inform real-life science education by 1) providing education practitioners with a better understanding of concepts related to the learning process; 2) assisting in the development of intervention programmes through instructional design; and 3) offering a deeper cognitive science background for understanding the formation of misconceptions, thereby guiding public outreach.

First, both Studies 1 and 2 demonstrated that, cross-culturally, the transformation of naïve to science thinking requires various cognitive abilities, not only simply suppressing naïve thinking but also activating science thinking. Education practitioners may therefore acknowledge the difficulties faced by students in science reasoning when confronted with conflicts between naïve and science thinking (Chi, 2005; Li et al., 2023; Meier et al., 2022). These results may be used to promote some amount of support for students, for example providing them with more time for doing science reasoning, and helping them realise it is normal to make mistakes. These steps would help students overcome learning difficulties and increase their motivation to learn.

Second, the current research stresses the importance of executive functions, so developing intervention programmes with a focus on enhancing executive functions may help individuals learn science knowledge with explanatory coexistence concepts. However, to date, most studies have not found far-transfer effects of executive function training on academic performance (Klingberg, 2010; Niebaum & Munakata, 2023; Schwaighofer et al., 2015; Willis et al., 2006). Consequently, consideration as to how executive function knowledge and science learning contents can be integrated might be another way to improve performance on solving naïve thinking tasks. As

inhibition is the most studied executive function component, there are already intervention programmes integrating inhibition into teaching science. For example, the *Stop and Think* programme asks students in primary school to delay responding to questions requiring explanatory coexistence, thereby giving these students more time for the inhibition of naïve thinking (Wilkinson et al., 2020). Results found that this intervention significantly improved primary school students' performance on solving science questions. The importance of switching may indicate further knowledge-related interventions that consider both inhibiting naïve thinking and activating science thinking, such as using refutation texts that not only reject wrong naïve thinking, but also provide accurate science explanations (Asterhan & Resnick, 2020; Kendeou et al., 2014).

Finally, the current research findings affirm the importance of disseminating accurate scientific knowledge. The proliferation of smartphones and different software applications, including social media platforms such as Instagram and TikTok, has exposed children to an unprecedented amount incorrect naïve thinking (i.e., online misinformation), as well as correct science thinking (see the review by Suarez-Lledo & Alvarez-Galvez, 2021). Moreover, given the diversity of media modalities currently circulating, children may establish incorrect naïve thinking through more intuitive means, such as images and videos (Muhammed & Mathew, 2022). Without proper scientific guidance, children may be easily misled by inaccurate information on social media surrounding issues such as vaccines or climate change, solidifying scientific misunderstandings (Shtulman, 2023; van der Linden et al., 2017). These research results, when aligned with studies on misconceptions, emphasise the need for policymakers to pay closer attention to the filtering and distribution of accurate scientific information to the public, such as providing interventions or training to help the public resist misinformation (Roozenbeek & van der Linden, 2019).

Critically, the results of cognitive science research cannot directly inform educational practice, but can offer deeper insights into the factors that influence learning (Chang et al., 2021; Howard-Jones et al., 2016). Moreover, this research requires further replication among children and adolescents. Thus, although this study's findings may have some implications for real-world teaching and learning, its experimental results are perhaps still confined to the psychological laboratory.

7.7 Conclusion

This PhD research study concurrently measures the relative contributions of executive functions, general cognitive ability and science education on explanatory coexistence among adult participants in a cross-cultural context. Through Study 1 (the United Kingdom dataset) and Study 2 (China dataset), its results suggest that executive functions may be the most significant factor contributing to explanatory coexistence. More specifically, inhibition and working memory seem to support switching during the process of suppressing naïve thinking and activating science thinking. General cognitive ability, meanwhile, contributes more to the general learning of scientific knowledge. Consistent with previous explorations into task modality, these results show that only switching is involved in both picture- and text-based tasks, while inhibition seems to be engaged only in picture-based tasks, and general cognitive ability only in text-based tasks.

Inconsistent with previous studies, this research does not find science education to contribute to explanatory coexistence. However, this finding may be due to the different measurement methods used in this study. Moreover, cross-cultural results indicate that, apart from the different response patterns used by both countries' participants in cognitive and explanatory coexistence-related tasks, the factors contributing to explanatory coexistence, apart from behavioural inhibition, are similar across the two countries. This finding further illustrates that the underlying mechanisms of explanatory coexistence revealed in the current research may be robust.

This PhD research has some limitations that further studies may address by expanding its sample size to different age groups, employing more diverse executive function tasks to measure the robustness of findings, considering more regions for cross-cultural data collection, refining the measurements of science education and using tasks with more modalities. While there is a long way from this research to real-

world implications, its results do generate some potential practical implications, particularly revealing from a cognitive perspective why transitioning from naïve to science thinking is so challenging. In so doing, this research highlights the importance of accurate science communication, not only in the classroom but for whole social contexts.

References

- Ackerman, P. L. (2007). New developments in understanding skilled performance. *Current Directions in Psychological Science*, 16(5), 235–239.
<https://doi.org/10.1111/j.1467-8721.2007.00511.x>
- Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin*, 131(1), 30–60. <http://dx.doi.org.pitt.idm.oclc.org/10.1037/0033-2909.131.1.30>
- Ackerman, R., & Koriath, A. (2011). Response latency as a predictor of the accuracy of children's reports. *Journal of Experimental Psychology: Applied*, 17, 406–417.
<https://doi.org/10.1037/a0025129>
- Adams, F., & Campbell, K. (1999). Modality and abstract concepts. *Behavioral and Brain Sciences*, 22(4), 610–610. <https://doi.org/10.1017/S0140525X99222145>
- Ahr, E., Houdé, O., & Borst, G. (2016). Inhibition of the mirror generalization process in reading in school-aged children. *Journal of Experimental Child Psychology*, 145, 157–165. <https://doi.org/10.1016/j.jecp.2015.12.009>
- Aikenhead, G. S., & Jegede, O. J. (1999). Cross-cultural science education: A cognitive explanation of a cultural phenomenon. *Journal of Research in Science Teaching*, 36(3), 269–287. [https://doi.org/10.1002/\(SICI\)1098-2736\(199903\)36:3<269::AID-TEA3>3.0.CO;2-T](https://doi.org/10.1002/(SICI)1098-2736(199903)36:3<269::AID-TEA3>3.0.CO;2-T)
- Akaike, H. (1998). A new look at the statistical model identification. In E. Parzen, K. Tanabe, & G. Kitagawa (Eds.), *Selected papers of Hirotugu Akaike* (pp. 215–222). Springer. https://doi.org/10.1007/978-1-4612-1694-0_16
- Allaire-Duquette, G., Babai, R., & Stavy, R. (2019). Interventions aimed at overcoming intuitive interference: Insights from brain-imaging and behavioral studies. *Cognitive Processing*, 20(1), 1–9. <https://doi.org/10.1007/s10339-018-0893-2>

- Allaire-Duquette, G., Bélanger, M., Grabner, R. H., Koschutnig, K., & Masson, S. (2019). Individual differences in science competence among students are associated with ventrolateral prefrontal cortex activation. *Journal of Neuroscience Research*, 97(9), 1163–1178. <https://doi.org/10.1002/jnr.24435>
- Allen, M. (2014). *Misconceptions in primary science*. McGraw-Hill Education (UK).
- Allison, P. (2012). *Handling missing data by maximum likelihood*. Retrieved from <https://www.semanticscholar.org/paper/Handling-Missing-Data-by-Maximum-Likelihood-Allison/f2de0d551978821a6ac64f85d084fd19e0c1947d>
- Allison, P. D. (2003). Missing data techniques for structural equation modeling. *Journal of Abnormal Psychology*, 112(4), 545–557. <https://doi.org/10.1037/0021-843X.112.4.545>
- Alloway, T., & Alloway, R. (2009). The efficacy of working memory training in improving crystallized intelligence. *Nature Precedings*. <https://doi.org/10.1038/npre.2009.3697.1>
- Allport, A., & Wylie, G. (1999). Task-switching: Positive and negative priming of task-set. In *Attention, space, and action: Studies in cognitive neuroscience* (pp. 273–296). Oxford University Press.
- Altman, D. G., & Royston, P. (2006). The cost of dichotomising continuous variables. *British Medical Journal*, 332(7549), 1080.
- Alvermann, D. E., & Hague, S. A. (1989). Comprehension of counterintuitive science text: Effects of prior knowledge and text structure. *The Journal of Educational Research*, 82(4), 197–202. <https://doi.org/10.1080/00220671.1989.10885893>
- Anderson, P. (2002). Assessment and development of executive function (EF) during childhood. *Child Neuropsychology*, 8(2), 71–82. <https://doi.org/10.1076/chin.8.2.71.8724>
- Ardila, A. (2018). Is intelligence equivalent to executive functions? *Psicothema*, 30(2), 159–164. <https://doi.org/10.7334/psicothema2017.329>

- Artuso, C., Carretti, B., & Palladino, P. (2019). Short-term training on working memory updating and metacognition in primary school: The effect on reading comprehension. *School Psychology International*, 40(6), 641–657.
<https://doi.org/10.1177/0143034319881671>
- Asterhan, C. S. C., & Resnick, M. S. (2020). Refutation texts and argumentation for conceptual change: A winning or a redundant combination? *Learning and Instruction*, 65. Scopus. <https://doi.org/10.1016/j.learninstruc.2019.101265>
- Babai, R., & Amsterdamer, A. (2008). The persistence of solid and liquid naïve conceptions: A reaction time study. *Journal of Science Education and Technology*, 17(6), 553–559. <https://doi.org/10.1007/s10956-008-9122-6>
- Babai, R., Sekal, R., & Stavy, R. (2010). Persistence of the intuitive conception of living things in adolescence. *Journal of Science Education and Technology*, 19(1), 20–26. <https://doi.org/10.1007/s10956-009-9174-2>
- Baddeley, A. (2003). Working memory and language: An overview. *Journal of Communication Disorders*, 36(3), 189–208. [https://doi.org/10.1016/S0021-9924\(03\)00019-4](https://doi.org/10.1016/S0021-9924(03)00019-4)
- Baggetta, P., & Alexander, P. A. (2016). Conceptualization and operationalization of executive function. *Mind, Brain, and Education*, 10(1), 10–33.
<https://doi.org/10.1111/mbe.12100>
- Bandalos, D. L. (2014). Relative performance of categorical diagonally weighted least squares and robust maximum likelihood estimation. *Structural Equation Modeling: A Multidisciplinary Journal*, 21(1), 102–116.
<https://doi.org/10.1080/10705511.2014.859510>
- Bartlett, F. C. (1995). *Remembering: A study in experimental and social psychology*. Cambridge University Press.
- Barkley, R. A. (1997). Behavioral inhibition, sustained attention, and executive functions: Constructing a unifying theory of ADHD. *Psychological Bulletin*, 121(1), 65–94. <https://doi.org/10.1037/0033-2909.121.1.65>

- Barrouillet, P., Gavens, N., Vergauwe, E., Gaillard, V., & Camos, V. (2009). Working memory span development: A time-based resource-sharing model account. *Developmental Psychology*, 45(2), 477–490. <https://doi.org/10.1037/a0014615>
- Bascandziev, I., Tardiff, N., Zaitchik, D., & Carey, S. (2018a). The role of domain-general cognitive resources in children’s construction of a vitalist theory of biology. *Cognitive Psychology*, 104, 1–28. <https://doi.org/10.1016/j.cogpsych.2018.03.002>
- Bascandziev, I., Tardiff, N., Zaitchik, D., & Carey, S. (2018b). The role of domain-general cognitive resources in children’s construction of a vitalist theory of biology. *Cognitive Psychology*, 104, 1–28. <https://doi.org/10.1016/j.cogpsych.2018.03.002>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Bayanova, L., Chichinina, E., & Aslanova, M. (2024). The association between music training and executive function in 6–7-year-old children. *Frontiers in Education*, 9. <https://doi.org/10.3389/educ.2024.1333580>
- Bedard, A.-C., Nichols, S., Barbosa, J. A., Schachar, R., Logan, G. D., & Tannock, R. (2002). The development of selective inhibitory control across the life span. *Developmental Neuropsychology*, 21(1), 93–111. https://doi.org/10.1207/S15326942DN2101_5
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107(2), 238–246. <https://doi.org/10.1037/0033-2909.107.2.238>
- Bentler, P. M., & Chou, C.-P. (1987). Practical issues in structural modelling. *Sociological Methods & Research*, 16(1), 78–117. <https://doi.org/10.1177/0049124187016001004>

- Bélanger, É., Brault Foisy, L.-M., & Masson, S. (2024). What insights can response times provide for education research? *International Journal of Research & Method in Education*, 1-16. <https://doi.org/10.1080/1743727X.2024.2336146>
- Berch, D., Krikorian, R., & Huha, E. (1999). The corsi block-tapping task: Methodological and theoretical considerations. *Brain and Cognition*, 38, 317–338. <https://doi.org/10.1006/brcg.1998.1039>
- Best, J. R., & Miller, P. H. (2010). A developmental perspective on executive function. *Child Development*, 81(6), 1641–1660. <https://doi.org/10.1111/j.1467-8624.2010.01499.x>
- Bird-David, N. (1999). "Animism" revisited: Personhood, environment, and relational epistemology. *Current Anthropology*, 40(S1), S67–S91. <https://doi.org/10.1086/200061>
- Bishara, A. J., & Hittner, J. B. (2012). Testing the significance of a correlation with nonnormal data: Comparison of Pearson, Spearman, transformation, and resampling approaches. *Psychological Methods*, 17(3), 399–417. <https://doi.org/10.1037/a0028087>
- Bjork, R. A., & Bjork, E. L. (2019). Forgetting as the friend of learning: Implications for teaching and self-regulated learning. *Advances in Physiology Education*, 43(2), 164–167. <https://doi.org/10.1152/advan.00001.2019>
- Blair, C. (2006). How similar are fluid cognition and general intelligence? A developmental neuroscience perspective on fluid cognition as an aspect of human cognitive ability. *The Behavioral and Brain Sciences*, 29(2), 109–125; discussion 125-160. <https://doi.org/10.1017/S0140525X06009034>
- Blair, C., & Raver, C. C. (2016). Poverty, stress, and brain development: New directions for prevention and intervention. *Academic Pediatrics*, 16(3 Suppl), S30–S36. <https://doi.org/10.1016/j.acap.2016.01.010>

- Borgstede, M., & Scholz, M. (2021). Quantitative and qualitative approaches to generalization and replication—a representationalist view. *Frontiers in Psychology, 12*, 605191. <https://doi.org/10.3389/fpsyg.2021.605191>
- Bourisseau, W., Davis, O. L., & Yamamoto, K. (1965). Sense-impression responses to differing pictorial and verbal stimuli. *AV Communication Review, 13*(3), 249–258.
- Boyle, G. J., Wongsri, N., Bahr, M., Macayan, J. V., & Bentler, P. M. (2020). Cross-cultural differences in personality, motivation and cognition in Asian vs. Western societies. *Personality and Individual Differences, 159*, 109834. <https://doi.org/10.1016/j.paid.2020.109834>
- Braasch, J. L., Goldman, S. R., & Wiley, J. (2013). The influences of text and reader characteristics on learning from refutations in science texts. *Journal of Educational Psychology, 105*(3), 561.
- Brault Foisy, L.-M., Ahr, E., Blanchette Sarrasin, J., Potvin, P., Houdé, O., Masson, S., & Borst, G. (2021). Inhibitory control and the understanding of buoyancy from childhood to adulthood. *Journal of Experimental Child Psychology, 208*. Scopus. <https://doi.org/10.1016/j.jecp.2021.105155>
- Brault Foisy, L.-M., Potvin, P., Riopel, M., & Masson, S. (2015). Is inhibition involved in overcoming a common physics misconception in mechanics? *Trends in Neuroscience and Education, 4*(1–2), 26–36. <https://doi.org/10.1016/j.tine.2015.03.001>
- Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica, 47*(5), 1287–1294. <https://doi.org/10.2307/1911963>
- Bring, J. (1994). How to standardize regression coefficients. *The American Statistician, 48*(3), 209–213. <https://doi.org/10.2307/2684719>

- Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology, 1*(3), 185–216.
<https://doi.org/10.1177/135910457000100301>
- Brockmole, J. R., & Logie, R. H. (2013). Age-related change in visual working memory: A study of 55,753 participants aged 8–75. *Frontiers in Psychology, 4*. <https://doi.org/10.3389/fpsyg.2013.00012>
- Brod, G. (2021). Toward an understanding of when prior knowledge helps or hinders learning. *Npj Science of Learning, 6*(1), Article 1.
<https://doi.org/10.1038/s41539-021-00103-w>
- Brookman-Byrne, A., Mareschal, D., Tolmie, A. K., & Dumontheil, I. (2018). Inhibitory control and counterintuitive science and maths reasoning in adolescence. *PLOS ONE, 13*(6), e0198973.
<https://doi.org/10.1371/journal.pone.0198973>
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research, 2nd ed* (pp. xvii, 462). The Guilford Press.
- Bruyer, R., & Brysbaert, M. (2011). Combining speed and accuracy in cognitive psychology: Is the inverse efficiency score (IES) a better dependent variable than the mean reaction time (RT) and the percentage of errors (PE)? *Psychologica Belgica, 51*(1), Article 1. <https://doi.org/10.5334/pb-51-1-5>
- Brydges, C. R., Clunies-Ross, K., Clohessy, M., Lo, Z. L., Nguyen, A., Rousset, C., Whitelaw, P., Yeap, Y. J., & Fox, A. M. (2012). Dissociable components of cognitive control: An event-related potential (erp) study of response inhibition and interference suppression. *PLOS ONE, 7*(3), e34482.
<https://doi.org/10.1371/journal.pone.0034482>
- Brydges, C. R., Reid, C. L., Fox, A. M., & Anderson, M. (2012). A unitary executive function predicts intelligence in children. *Intelligence, 40*(5), 458–469.
<https://doi.org/10.1016/j.intell.2012.05.006>

- Bull, R., Espy, K. A., & Wiebe, S. A. (2008). Short-term memory, working memory, and executive functioning in preschoolers: Longitudinal predictors of mathematical achievement at age 7 years. *Developmental Neuropsychology*, 33(3), 205–228. <https://doi.org/10.1080/87565640801982312>
- Bull, R., & Lee, K. (2014). Executive functioning and mathematics achievement. *Child Development Perspectives*, 8(1), 36–41. <https://doi.org/10.1111/cdep.12059>
- Burgess, P., & Simons, J. (2005). Theories of frontal lobe executive function: Clinical applications. *Effectiveness of Rehabilitation for Cognitive Deficits*, 211–231. <https://doi.org/10.1093/acprof:oso/9780198526544.003.0018>
- Burkart, J. M., Schubiger, M. N., & Schaik, C. P. van. (2017). The evolution of general intelligence. *Behavioral and Brain Sciences*, 40, e195. <https://doi.org/10.1017/S0140525X16000959>
- Butterfuss, R., & Kendeou, P. (2018). The role of executive functions in reading comprehension. *Educational Psychology Review*, 30(3), 801–826. <https://doi.org/10.1007/s10648-017-9422-6>
- Butterfuss, R., & Kendeou, P. (2020). Reducing interference from misconceptions: The role of inhibition in knowledge revision. *Journal of Educational Psychology*, 112(4), 782–794. <https://doi.org/10.1037/edu0000385>
- Byrne, B. M. (2016). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (3rd ed.). Routledge. <https://doi.org/10.4324/9781315757421>
- Cahan, S., & Cohen, N. (1989). Age versus schooling effects on intelligence development. *Child Development*, 60(5), 1239–1249.
- Carey, S. (1985). *Conceptual change in childhood*. MIT Press.
- Carey, S. (2000). Science education as conceptual change. *Journal of Applied Developmental Psychology*, 21, 13–19. [https://doi.org/10.1016/S0193-3973\(99\)00046-5](https://doi.org/10.1016/S0193-3973(99)00046-5)

- Carney, R. N., & Levin, J. R. (2002). Pictorial illustrations still improve students' learning from text. *Educational Psychology Review*, 14(1), 5–26.
<https://doi.org/10.1023/A:1013176309260>
- Carter, R. L. (2006). Solutions for missing data in structural equation modeling. *Research & Practice in Assessment*, 1, 20-27.
- Case, R., Kurland, D. M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology*, 33(3), 386–404. [https://doi.org/10.1016/0022-0965\(82\)90054-6](https://doi.org/10.1016/0022-0965(82)90054-6)
- Cavanaugh, J. E., & Neath, A. A. (2019). The Akaike information criterion: Background, derivation, properties, application, interpretation, and refinements. *WIREs Computational Statistics*, 11(3), e1460.
<https://doi.org/10.1002/wics.1460>
- Çepni, S., & Keleş, E. (2006). Turkish students' conceptions about the simple electric circuits. *International Journal of Science and Mathematics Education*, 4(2), 269–291. <https://doi.org/10.1007/s10763-005-9001-z>
- Chai, W. J., Abd Hamid, A. I., & Abdullah, J. M. (2018). Working memory from the psychological and neurosciences perspectives: A review. *Frontiers in Psychology*, 9. <https://www.frontiersin.org/articles/10.3389/fpsyg.2018.00401>
- Chakravartty, A. (2022). Scientific knowledge vs. Knowledge of science. *Science & Education*. <https://doi.org/10.1007/s11191-022-00376-6>
- Chan, M. S., Jones, C. R., Hall Jamieson, K., & Albarracín, D. (2017). Debunking: A meta-analysis of the psychological efficacy of messages countering misinformation. *Psychological Science*, 28(11), 1531–1546.
<https://doi.org/10.1177/0956797617714579>
- Chan, R. C. K., Shum, D., Touloupoulou, T., & Chen, E. Y. H. (2008). Assessment of executive functions: Review of instruments and identification of critical issues. *Archives of Clinical Neuropsychology*, 23(2), 201–216.
<https://doi.org/10.1016/j.acn.2007.08.010>

- Chang, Z., Schwartz, M. S., Hinesley, V., & Dubinsky, J. M. (2021). Neuroscience concepts changed teachers' views of pedagogy and students. *Frontiers in Psychology, 12*.
<https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2021.685856>
- Chaytor, N., Schmitter-Edgecombe, M., & Burr, R. (2006). Improving the ecological validity of executive functioning assessment. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists, 21*(3), 217–227.
<https://doi.org/10.1016/j.acn.2005.12.002>
- Chen, C., Sonnert, G., Sadler, P. M., Sasselov, D., & Fredericks, C. (2020). The impact of student misconceptions on student persistence in a MOOC. *Journal of Research in Science Teaching, 57*(6), 879–910.
<https://doi.org/10.1002/tea.21616>
- Chen, H., Keith, T. Z., Weiss, L., Zhu, J., & Li, Y. (2010). Testing for multigroup invariance of second-order WISC-IV structure across China, Hong Kong, Macau, and Taiwan. *Personality and Individual Differences, 49*(7), 677–682.
<https://doi.org/10.1016/j.paid.2010.06.004>
- Cheung, A. M., Mitsis, E. M., & Halperin, J. M. (2004). The relationship of behavioral inhibition to executive functions in young adults. *Journal of Clinical and Experimental Neuropsychology, 26*(3), 393–404.
<https://doi.org/10.1080/13803390490510103>
- Chi, M. T. H. (2005). Commonsense conceptions of emergent processes: Why some misconceptions are robust. *Journal of the Learning Sciences, 14*(2), 161–199.
https://doi.org/10.1207/s15327809jls1402_1
- Chi, M. T. H., De Leeuw, N., Chiu, M.-H., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*(3), 439–477.
[https://doi.org/10.1016/0364-0213\(94\)90016-7](https://doi.org/10.1016/0364-0213(94)90016-7)

- Chierchia, G., Fuhrmann, D., Knoll, L. J., Pi-Sunyer, B. P., Sakhardande, A. L., & Blakemore, S.-J. (n.d.). The matrix reasoning item bank (MaRs-IB): Novel, open-access abstract reasoning items for adolescents and adults. *Royal Society Open Science*, 6(10), 190232. <https://doi.org/10.1098/rsos.190232>
- Chmielewski, A. K. (2019). The global increase in the socioeconomic achievement gap, 1964 to 2015. *American Sociological Review*, 84(3), 517–544. <https://doi.org/10.1177/0003122419847165>
- Cho, I., Hosseini-Kamkar, N., Song, H., & Morton, J. B. (2023). Culture, executive functions, and academic achievement. *Frontiers in Psychology*, 14. <https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2023.1100537>
- Christ, B. U., Combrinck, M. I., & Thomas, K. G. F. (2018). Both reaction time and accuracy measures of intraindividual variability predict cognitive performance in alzheimer’s disease. *Frontiers in Human Neuroscience*, 12. <https://doi.org/10.3389/fnhum.2018.00124>
- Chua, H. F., Boland, J. E., & Nisbett, R. E. (2005). Cultural variation in eye movements during scene perception. *Proceedings of the National Academy of Sciences of the United States of America*, 102(35), 12629–12633. <https://doi.org/10.1073/pnas.0506162102>
- Clement, J. (1993). Using bridging analogies and anchoring intuitions to deal with students’ preconceptions in physics. *Journal of Research in Science Teaching*, 30(10), 1241–1257. <https://doi.org/10.1002/tea.3660301007>
- Carter, R. L. (2006). Solutions for missing data in structural equation modeling. *Research & Practice in Assessment*, 1, 20-27.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*, 3rd ed (pp. xxviii, 703). Lawrence Erlbaum Associates Publishers.

- Collins, A., & Koechlin, E. (2012). Reasoning, learning, and creativity: Frontal lobe function and human decision-making. *PLOS Biology*, *10*(3), e1001293.
<https://doi.org/10.1371/journal.pbio.1001293>
- Conway, A. R. A., Cowan, N., Bunting, M. F., Theriault, D. J., & Minkoff, S. R. B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, *30*(2), 163–183. [https://doi.org/10.1016/S0160-2896\(01\)00096-4](https://doi.org/10.1016/S0160-2896(01)00096-4)
- Costa, A., & Faria, L. (2018). Implicit theories of intelligence and academic achievement: A meta-analytic review. *Frontiers in Psychology*, *9*.
<https://www.frontiersin.org/articles/10.3389/fpsyg.2018.00829>
- Costa, A., & Santesteban, M. (2004). Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language*, *50*(4), 491–511.
<https://doi.org/10.1016/j.jml.2004.02.002>
- Cowan, N. (1999). The differential maturation of two processing rates related to digit span. *Journal of Experimental Child Psychology*, *72*(3), 193–209.
<https://doi.org/10.1006/jecp.1998.2486>
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114.
<https://doi.org/10.1017/S0140525X01003922>
- Cowan, N. (2010). Multiple concurrent thoughts: The meaning and developmental neuropsychology of working memory. *Developmental Neuropsychology*, *35*(5), 447–474. <https://doi.org/10.1080/87565641.2010.494985>
- Cowan, N. (2014). Working memory underpins cognitive development, learning, and education. *Educational Psychology Review*, *26*(2), 197–223.
<https://doi.org/10.1007/s10648-013-9246-y>
- Crone, E. A., Bunge, S. A., van der Molen, M. W., & Ridderinkhof, K. R. (2006). Switching between tasks and responses: A developmental study.

- Developmental Science*, 9(3), 278–287. <https://doi.org/10.1111/j.1467-7687.2006.00490.x>
- Dajani, D. R., & Uddin, L. Q. (2015). Demystifying cognitive flexibility: Implications for clinical and developmental neuroscience. *Trends in Neurosciences*, 38(9), 571–578. <https://doi.org/10.1016/j.tins.2015.07.003>
- Davidson, D. (2006). *The essential Davidson*. New York: Oxford University Press.
- Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia*, 44(11), 2037–2078. <https://doi.org/10.1016/j.neuropsychologia.2006.02.006>
- De Boeck, P., & Jeon, M. (2019). An overview of models for response times and processes in cognitive tests. *Frontiers in Psychology*, 10. <https://doi.org/10.3389/fpsyg.2019.00102>
- de Oliveira, C. R., Pedron, A. C., Gurgel, L. G., Reppold, C. T., & Fonseca, R. P. (2012). Executive functions and sustained attention: Comparison between age groups of 19-39 and 40-59 years old. *Dementia & Neuropsychologia*, 6(1), 29–34. <https://doi.org/10.1590/S1980-57642012DN06010005>
- Deary, I. J., & Johnson, W. (2010). Intelligence and education: Causal perceptions drive analytic processes and therefore conclusions. *International Journal of Epidemiology*, 39(5), 1362–1369. <https://doi.org/10.1093/ije/dyq072>
- Denes, G., & Pizzamiglio, L. (Eds.). (2020). *Handbook of Clinical and Experimental Neuropsychology*. Psychology Press. <https://doi.org/10.4324/9781315791272>
- Deng, L., Yang, M., & Marcoulides, K. M. (2018). Structural equation modeling with many variables: A systematic review of issues and developments. *Frontiers in Psychology*, 9. <https://www.frontiersin.org/articles/10.3389/fpsyg.2018.00580>
- Denis, D. J., & Legerski McCormick, J. (2006). Causal modeling and the origins of path analysis. *Theory and Science*, 7.

- Diamond, A. (2002). Normal development of prefrontal cortex from birth to young adulthood: Cognitive functions, anatomy, and biochemistry. In *Principles of frontal lobe function* (pp. 466–503). Oxford University Press.
<https://doi.org/10.1093/acprof:oso/9780195134971.003.0029>
- Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, 64(1), 135–168. <https://doi.org/10.1146/annurev-psych-113011-143750>
- Diamond, A., & Wright, A. (2014). An effect of inhibitory load in children while keeping working memory load constant. *Frontiers in Psychology*, 5.
<https://doi.org/10.3389/fpsyg.2014.00213>
- diSessa, A. A. (1993). Toward an epistemology of physics. *Cognition and Instruction*, 10(2–3), 105–225. <https://doi.org/10.1080/07370008.1985.9649008>
- Dong, Y., & Peng, C.-Y. J. (2013). Principled missing data methods for researchers. *SpringerPlus*, 2, 222. <https://doi.org/10.1186/2193-1801-2-222>
- Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., & Lautenbach, S. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36(1), 27–46.
<https://doi.org/10.1111/j.1600-0587.2012.07348.x>
- Draheim, C., Hicks, K. L., & Engle, R. W. (2016). Combining reaction time and accuracy: The relationship between working memory capacity and task switching as a case example. *Perspectives on Psychological Science*, 11(1), 133–155. <https://doi.org/10.1177/1745691615596990>
- Duit, R., & Treagust, D. (2012). How can conceptual change contribute to theory and practice in science education? In *Second International Handbook of Science Education* (pp. 107–118). https://doi.org/10.1007/978-1-4020-9041-7_9

- Duit, R., & Treagust, D. F. (2003). Conceptual change: A powerful framework for improving science teaching and learning. *International Journal of Science Education*, 25(6), 671–688. <https://doi.org/10.1080/09500690305016>
- Dumontheil, I., Brookman-Byrne, A., Tolmie, A. K., & Mareschal, D. (2022). Neural and cognitive underpinnings of counterintuitive science and math reasoning in adolescence. *Journal of Cognitive Neuroscience*, 34(7), 1205–1229. https://doi.org/10.1162/jocn_a_01854
- Dumontheil, I., & Klingberg, T. (2012). Brain activity during a visuospatial working memory task predicts arithmetical performance 2 years later. *Cerebral Cortex*, 22(5), 1078–1085. <https://doi.org/10.1093/cercor/bhr175>
- Dumontheil, I., Wilkinson, H. R., Farran, E. K., Smid, C., Modhvia, R., Mareschal, D., Bell, D., Brookman-Byrne, A., Gauthier, A., de Jong, R., Holmes, W., Mayer, S., Morris, S., Porayska-Pomsta, K., Sumanapala, D., Thomas, M., Tolmie, A., & the UnLocke team. (2023). How do executive functions influence children’s reasoning about counterintuitive concepts in mathematics and science? *Journal of Cognitive Enhancement*. <https://doi.org/10.1007/s41465-023-00271-0>
- Duncan, P. W., Weiner, D. K., Chandler, J., & Studenski, S. (1990). Functional reach: A new clinical measure of balance. *Journal of Gerontology*, 45(6), M192–M197. <https://doi.org/10.1093/geronj/45.6.M192>
- Dykstra JR., D. I., Boyle, C. F., & Monarch, I. A. (1992). Studying conceptual change in learning physics. *Science Education*, 76(6), 615–652. <https://doi.org/10.1002/sce.3730760605>
- Egner, T., & Hirsch, J. (2005). Cognitive control mechanisms resolve conflict through cortical amplification of task-relevant information. *Nature Neuroscience*, 8(12), 1784–1790. <https://doi.org/10.1038/nn1594>
- Elfil, M., & Negida, A. (2017). Sampling methods in clinical research; an educational review. *Emergency*, 5(1), e52.

- Ellefson, M. R., Baker, S. T., & Gibson, J. L. (2019). Lessons for successful cognitive developmental science in educational settings: The case of executive functions. *Journal of Cognition and Development*, 20(2), 253–277.
<https://doi.org/10.1080/15248372.2018.1551219>
- Ellefson, M. R., Ng, F. F.-Y., Wang, Q., & Hughes, C. (2017). Efficiency of executive function: A two-generation cross-cultural comparison of samples from Hong Kong and the United Kingdom. *Psychological Science*, 28(5), 555–566. <https://doi.org/10.1177/0956797616687812>
- Ellefson, M. R., Shapiro, L. R., & Chater, N. (2006). Asymmetrical switch costs in children. *Cognitive Development*, 21(2), 108–130.
<https://doi.org/10.1016/j.cogdev.2006.01.002>
- Ellefson, M. R., Wang, R., Britton, A., Godfrey, H., & Feeney, A. (2024). *Inhibition and switching make different contributions to counterintuitive science thinking in adults*. PsychArXiv. <https://doi.org/10.31234/osf.io/cm6kg>
- Ellefson, M. R., Zachariou, A., Ng, F. F.-Y., Wang, Q., & Hughes, C. (2020). Do executive functions mediate the link between socioeconomic status and numeracy skills? A cross-site comparison of Hong Kong and the United Kingdom. *Journal of Experimental Child Psychology*, 194, 104734.
<https://doi.org/10.1016/j.jecp.2019.104734>
- Enders, C. K. (2008). A note on the use of missing auxiliary variables in full information maximum likelihood-based structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, 15(3), 434–448.
<https://doi.org/10.1080/10705510802154307>
- Engel de Abreu, P. M. J., Conway, A. R. A., & Gathercole, S. E. (2010). Working memory and fluid intelligence in young children. *Intelligence*, 38(6), 552–561.
<https://doi.org/10.1016/j.intell.2010.07.003>
- Engelhardt, L. E., Mann, F. D., Briley, D. A., Church, J. A., Harden, K. P., & Tucker-Drob, E. M. (2016). Strong genetic overlap between executive functions and

- intelligence. *Journal of Experimental Psychology. General*, 145(9), 1141–1159. <https://doi.org/10.1037/xge0000195>
- Engelhardt, P. E., Nigg, J. T., & Ferreira, F. (2017). Executive function and intelligence in the resolution of temporary syntactic ambiguity: An individual differences investigation. *Quarterly Journal of Experimental Psychology*, 70(7), 1263–1281. <https://doi.org/10.1080/17470218.2016.1178785>
- Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, 128(3), 309–331.
- Eriksen, B. A., & 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. <https://doi.org/10.3758/BF03203267>
- Evermann, J. (2010). Multiple-group analysis using the sem package in the r system. *Structural Equation Modeling: A Multidisciplinary Journal*, 17(4), 677–702. <https://doi.org/10.1080/10705511.2010.510070>
- Falch, T., & Sandgren Massih, S. (2011). The effect of education on cognitive ability. *Economic Inquiry*, 49(3), 838–856. <https://doi.org/10.1111/j.1465-7295.2010.00312.x>
- Fan, Y., Chen, J., Shirkey, G., John, R., Wu, S. R., Park, H., & Shao, C. (2016). Applications of structural equation modeling (SEM) in ecological studies: An updated review. *Ecological Processes*, 5(1), 19. <https://doi.org/10.1186/s13717-016-0063-3>
- Feng, C., Wang, H., Lu, N., Chen, T., He, H., Lu, Y., & Tu, X. M. (2014). Log-transformation and its implications for data analysis. *Shanghai Archives of Psychiatry*, 26(2), 105–109. <https://doi.org/10.3969/j.issn.1002-0829.2014.02.009>

- Ferguson, H. J., Brunsdon, V. E. A., & Bradford, E. E. F. (2021). The developmental trajectories of executive function from adolescence to old age. *Scientific Reports*, 11(1), Article 1. <https://doi.org/10.1038/s41598-020-80866-1>
- Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, 141(1), 2–18. <https://doi.org/10.1037/a0024338>
- Field, A. P. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). Sage, Newbury Park.
- Flege, J. E., Yeni-Komshian, G., & Liu, S. (1999). Age of learning affects the learning of L2 pronunciation but not morphosyntax. *The Journal of the Acoustical Society of America*, 105(2), 1033–1033. <https://doi.org/10.1121/1.424934>
- Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). Sage. <https://socialsciences.mcmaster.ca/jfox/Books/Companion/>
- Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: A latent-variable analysis. *Journal of Experimental Psychology: General*, 133(1), 101–135. <https://doi.org/10.1037/0096-3445.133.1.101>
- Friedman, N. P., & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, 86, 186–204. <https://doi.org/10.1016/j.cortex.2016.04.023>
- Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., Defries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science*, 17(2), 172–179. <https://doi.org/10.1111/j.1467-9280.2006.01681.x>
- Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual differences in executive functions are almost entirely

- genetic in origin. *Journal of Experimental Psychology. General*, 137(2), 201–225. <https://doi.org/10.1037/0096-3445.137.2.201>
- Fugelsang, J., & Mareschal, D. (2013). The development and application of scientific reasoning. In *Educational neuroscience* (pp. 237–267). Wiley-Blackwell.
- Garnett, P. J., Garnett, P. J., & Hackling, M. W. (1995). Students' alternative conceptions in chemistry: A review of research and implications for teaching and learning. *Studies in Science Education*, 25(1), 69–96. <https://doi.org/10.1080/03057269508560050>
- Gaskins, S., & Alcalá, L. (2023). Studying executive function in culturally meaningful ways. *Journal of Cognition and Development*, 24(2), 260–279. <https://doi.org/10.1080/15248372.2022.2160722>
- Gebauer, M. M., McElvany, N., Köller, O., & Schöber, C. (2021). Cross-cultural differences in academic self-efficacy and its sources across socialization contexts. *Social Psychology of Education*, 24(6), 1407–1432. <https://doi.org/10.1007/s11218-021-09658-3>
- Georgiou, G. K., Wei, W., Inoue, T., Das, J. P., & Deng, C. (2020). Cultural influences on the relation between executive functions and academic achievement. *Reading and Writing*, 33(4), 991–1013. <https://doi.org/10.1007/s11145-019-09961-8>
- Ger, E., & Roebers, C. M. (2023). The relationship between executive functions, working memory, and intelligence in kindergarten children. *Journal of Intelligence*, 11(4), Article 4. <https://doi.org/10.3390/jintelligence11040064>
- Gernsbacher, M. A., & Faust, M. E. (1991). The mechanism of suppression: A component of general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17(2), 245–262. <http://dx.doi.org.pitt.idm.oclc.org/10.1037/0278-7393.17.2.245>

- Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: A guide for non-statisticians. *International Journal of Endocrinology and Metabolism*, 10(2), 486–489. <https://doi.org/10.5812/ijem.3505>
- Goldberg, R. F., & Thompson-Schill, S. L. (2009). Developmental ‘roots’ in mature biological knowledge. *Psychological Science*, 20(4), 480–487. <https://doi.org/10.1111/j.1467-9280.2009.02320.x>
- Goldfeld, S. M., & Quandt, R. E. (1965). Some tests for homoscedasticity. *Journal of the American Statistical Association*, 60(310), 539–547. <https://doi.org/10.2307/2282689>
- Goretzko, D., Siemund, K., & Sterner, P. (2024). Evaluating model fit of measurement models in confirmatory factor analysis. *Educational and Psychological Measurement*, 84(1), 123–144. <https://doi.org/10.1177/00131644231163813>
- Goris, T. V., & Dyrenfurth, M. J. (2010). Students’ misconceptions in science, technology, and engineering. In *ASEE Illinois/Indiana Section Conference proceedings*, Purdue University, West Lafayette, IN. Retrieved from <http://ilin.asee.org/Conference2010/Papers2010.html>
- Gottfredson, L. S. (2002). g: Highly general and highly practical. In *The general factor of intelligence: How general is it?* (pp. 331–380). Lawrence Erlbaum Associates Publishers. <https://doi.org/10.4324/9781410613165>
- GOV.UK. (2024, April 1). *The national curriculum*. GOV.UK. <https://www.gov.uk/national-curriculum>
- Grace, J., & Jutila, H. (1999). The relationship between species density and community biomass in grazed and ungrazed coastal meadows. *Oikos*, 85. <https://doi.org/10.2307/3546689>
- Grant, D. A., & Berg, E. (1948). A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem.

- Journal of Experimental Psychology*, 38(4), 404–411.
<https://doi.org/10.1037/h0059831>
- Gray, J. R., Chabris, C. F., & Braver, T. S. (2003). Neural mechanisms of general fluid intelligence. *Nature Neuroscience*, 6(3), 316–322.
<https://doi.org/10.1038/nn1014>
- Greenfield, P. M., Trumbull, E., Keller, H., Rothstein-Fisch, C., Suzuki, L., & Quiroz, B. (2006). Cultural conceptions of learning and development. In *Handbook of educational psychology* (pp. 675–692). Lawrence Erlbaum Associates Publishers.
- Greenland, S., Maclure, M., Schlesselman, J. J., Poole, C., & Morgenstern, H. (1991). Standardized regression coefficients: A further critique and review of some alternatives. *Epidemiology*, 2(5), 387. <https://www.jstor.org/stable/20065707>
- Grenell, A., & Carlson, S. M. (2021). Individual differences in executive function and learning: The role of knowledge type and conflict with prior knowledge. *Journal of Experimental Child Psychology*, 206, 105079.
<https://doi.org/10.1016/j.jecp.2020.105079>
- Gropen, J., Clark-Chiarelli, N., Hoisington, C., & Ehrlich, S. (2011). The importance of executive function in early science education. *Child Development Perspectives*, 5, 298–304. <https://doi.org/10.1111/j.1750-8606.2011.00201.x>
- Guzzetti, B. J. (2000). Learning counter-intuitive science concepts: What have we learned from over a decade of research? *Reading & Writing Quarterly: Overcoming Learning Difficulties*, 16(2), 89–98.
<https://doi.org/10.1080/105735600277971>
- Haig, B. D. (2013). The philosophy of quantitative methods. In *The Oxford handbook of quantitative methods in psychology: Vol. 1*.
<https://doi.org/10.1093/oxfordhb/9780199934874.013.0002>

- Hailikari, T., Katajavuori, N., & Lindblom-Ylänne, S. (2008). The relevance of prior knowledge in learning and instructional design. *American Journal of Pharmaceutical Education*, 72(5), 113.
- Hanania, R., & Smith, L. B. (2010). Selective attention and attention switching: Towards a unified developmental approach. *Developmental Science*, 13(4), 622–635. <https://doi.org/10.1111/j.1467-7687.2009.00921.x>
- Hanley, J. R. (2005). Learning to Read in Chinese. In *The science of reading: A handbook* (pp. 316–335). Blackwell Publishing.
<https://doi.org/10.1002/9780470757642.ch17>
- Harrell, Jr., F. (2023). *Hmisc: Harrell miscellaneous* (R package version 5.1-2). Retrieved from <https://hbiostat.org/R/Hmisc/>.
- Hasher, L., Lustig, C., & Zacks, R. (2008). Inhibitory mechanisms and the control of attention. In A. Conway, C. Jarrold, M. Kane, A. Miyake, & J. Towse (Eds.), *Variation in Working Memory* (pp. 227–249). Oxford University Press.
<https://doi.org/10.1093/acprof:oso/9780195168648.003.0009>
- Hasher, L., & Zacks, R. T. (1984). Automatic processing of fundamental information: The case of frequency of occurrence. *American Psychologist*, 39(12), 1372–1388. <http://dx.doi.org.pitt.idm.oclc.org/10.1037/0003-066X.39.12.1372>
- Hau, K.-T., & Ho, I. (2012). Chinese students' motivation and achievement. *The Oxford Handbook of Chinese Psychology*, 187–204.
<https://doi.org/10.1093/oxfordhb/9780199541850.013.0014>
- Hauf, P., Paulus, M., & Baillargeon, R. (2012). Infants use compression information to infer objects' weights: Examining cognition, exploration, and prospective action in a preferential-reaching task. *Child Development*, 83(6), 1978–1995.
<https://doi.org/10.1111/j.1467-8624.2012.01824.x>
- Hebbali, A. (2024). *olsrr: Tools for building OLS regression models* (R package version 0.6.0.9000). Retrieved from <https://github.com/rsquaredacademy/olsrr>,
<https://olsrr.rsquaredacademy.com/>.

- Hegarty, M., & Waller, D. (2004). A dissociation between mental rotation and perspective-taking spatial abilities. *Intelligence*, 32(2), 175–191.
<https://doi.org/10.1016/j.intell.2003.12.001>
- Heitz, R. P. (2014). The speed-accuracy trade-off: History, physiology, methodology, and behavior. *Frontiers in Neuroscience*, 8, 150.
<https://doi.org/10.3389/fnins.2014.00150>
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature*, 466(7302), Article 7302. <https://doi.org/10.1038/466029a>
- Hilbert, S., Nakagawa, T. T., Puci, P., Zech, A., & Bühner, M. (2015). The digit span backwards task: Verbal and visual cognitive strategies in working memory assessment. *European Journal of Psychological Assessment*, 31(3), 174–180.
<https://doi.org/10.1027/1015-5759/a000223>
- Howard-Jones, P. A., Varma, S., Ansari, D., Butterworth, B., De Smedt, B., Goswami, U., Laurillard, D., & Thomas, M. S. C. (2016). The principles and practices of educational neuroscience: Comment on Bowers (2016). *Psychological Review*, 123(5), 620–627. <https://doi.org/10.1037/rev0000036>
- Hox, J. J., & Bechger, T. M. (1998). An introduction to structural equation modeling. *Family Science Review*, 11, 354–373.
<https://dspace.library.uu.nl/handle/1874/23738>
- Hoyle, R., & St, C. (2022). *Handbook of structural equation modeling*. New York: The Guilford Press.
- Hsin, A., & Xie, Y. (2014). Explaining Asian Americans’ academic advantage over Whites. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, 111(23), 8416–8421. <https://doi.org/10.1073/pnas.1406402111>
- Hsin, C. T., & Wu, H. K. (2011). Using scaffolding strategies to promote young children’s scientific understandings of floating and sinking. *Journal of Science*

- Education and Technology*, 20(5), 656–666. <https://doi.org/10.1007/s10956-011-9310-7>
- Hu, L., & Bentler, P. M. (1999). Cut off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Hübner, R., Futterer, T., & Steinhauser, M. (2001). On attentional control as a source of residual shift costs: Evidence from two-component task shifts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(3), 640–653. <https://doi.org/10.1037/0278-7393.27.3.640>
- Hughes, C. (2023). Executive functions: Going places at pace. *Journal of Cognition and Development*, 24(2), 296–306. <https://doi.org/10.1080/15248372.2023.2187636>
- Huizinga, M., Dolan, C. V., & van der Molen, M. W. (2006). Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia*, 44(11), 2017–2036. <https://doi.org/10.1016/j.neuropsychologia.2006.01.010>
- Huizinga, M., & Molen, M. W. van der. (2011). Task switching and shifting between stopping and going: Developmental change in between-trial control adjustments. *Journal of Experimental Child Psychology*, 108(3), 484–503. <https://doi.org/10.1016/j.jecp.2010.10.004>
- Huntington, S. P. (1993). The clash of civilizations? *Foreign Affairs*, 72(3), 22–49. <https://doi.org/10.2307/20045621>
- Hutchison, K. A., Balota, D. A., & Duchek, J. M. (2010). The utility of stroop task switching as a marker for early stage alzheimer’s disease. *Psychology and Aging*, 25(3), 545–559. <https://doi.org/10.1037/a0018498>
- Inagaki, K., & Hatano, G. (1996). Young children’s recognition of commonalities between animals and plants. *Child Development*, 67(6), 2823–2840. <https://doi.org/10.2307/1131754>

- Isbell, E., Fukuda, K., Neville, H. J., & Vogel, E. K. (2015). Visual working memory continues to develop through adolescence. *Frontiers in Psychology, 6*.
<https://doi.org/10.3389/fpsyg.2015.00696>
- Jaeggi, S. M., Buschkuhl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences of the United States of America, 105*(19), 6829–6833.
<https://doi.org/10.1073/pnas.0801268105>
- Jaschke, A. C., Honing, H., & Scherder, E. J. A. (2018). Longitudinal analysis of music education on executive functions in primary school children. *Frontiers in Neuroscience, 12*.
<https://www.frontiersin.org/articles/10.3389/fnins.2018.00103>
- Jersild, A. T. (1927). Mental set and shift. *Archives of Psychology, 14*, 89, 81–81.
- Jin, L., Jia, H., Li, H., & Yu, D. (2019). Differences in brain signal complexity between experts and novices when solving conceptual science problem: A functional near-infrared spectroscopy study. *Neuroscience Letters, 699*, 172–176. <https://doi.org/10.1016/j.neulet.2019.02.015>
- Kane, M., Hambrick, Z., Tuholski, S., Wilhelm, O., Payne, T., & Engle, R. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology. General, 133*, 189–217. <https://doi.org/10.1037/0096-3445.133.2.189>
- Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2005). Working memory capacity and fluid intelligence are strongly related constructs: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin, 131*(1), 66–71.
<http://dx.doi.org.pitt.idm.oclc.org/10.1037/0033-2909.131.1.66>
- Kang, H. (2013). The prevention and handling of the missing data. *Korean Journal of Anesthesiology, 64*(5), 402–406. <https://doi.org/10.4097/kjae.2013.64.5.402>

- Kelemen, D., Rottman, J., & Seston, R. (2012). Professional physical scientists display tenacious teleological tendencies: Purpose-based reasoning as a cognitive default. *Journal of Experimental Psychology. General*, 142. <https://doi.org/10.1037/a0030399>
- Kelkar, A. S., Hough, M. S., & Fang, X. (2013). Do we think alike? A cross-cultural study of executive functioning. *Culture and Brain*, 1(2), 118–137. <https://doi.org/10.1007/s40167-013-0010-4>
- Kell, H. J., Lubinski, D., & Benbow, C. P. (2013). Who rises to the top? Early indicators. *Psychological Science*, 24(5), 648–659. <https://doi.org/10.1177/0956797612457784>
- Kendeou, P., & O'Brien, E. J. (2014). The knowledge revision components (KReC) framework: Processes and mechanisms. In *Processing inaccurate information: Theoretical and applied perspectives from cognitive science and the educational sciences* (pp. 353–377). The MIT Press. <https://doi.org/10.7551/mitpress/9737.001.0001>
- Kendeou, P., & van den Broek, P. (2007). The effects of prior knowledge and text structure on comprehension processes during reading of scientific texts. *Memory & Cognition*, 35(7), 1567–1577. <https://doi.org/10.3758/BF03193491>
- Kendeou, P., Walsh, E. K., Smith, E. R., & O'Brien, E. J. (2014). Knowledge revision processes in refutation texts. *Discourse Processes*, 51(5–6), 374–397. <https://doi.org/10.1080/0163853X.2014.913961>
- Kim, J. H. (2019). Multicollinearity and misleading statistical results. *Korean Journal of Anesthesiology*, 72(6), 558–569. <https://doi.org/10.4097/kja.19087>
- Kim, S. (2015). ppcor: An R package for a fast calculation to semi-partial correlation coefficients. *Communications for Statistical Applications and Methods*, 22(6), 665–674. <https://doi.org/10.5351/CSAM.2015.22.6.665>
- Kim, S. won. (2019). Is socioeconomic status less predictive of achievement in East Asian countries? A systematic and meta-analytic review. *International*

- Journal of Educational Research*, 97, 29–42.
<https://doi.org/10.1016/j.ijer.2019.05.009>
- Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology*, 55(4), 352–358.
<https://doi.org/10.1037/h0043688>
- Kline, R. B. (2016). *Principles and practice of structural equation modeling, 4th ed* (pp. xvii, 534). Guilford Press.
- Knief, U., & Forstmeier, W. (2021). Violating the normality assumption may be the lesser of two evils. *Behavior Research Methods*, 53(6), 2576–2590.
<https://doi.org/10.3758/s13428-021-01587-5>
- Koch, I., Gade, M., Schuch, S., & Philipp, A. M. (2010). The role of inhibition in task switching: A review. *Psychonomic Bulletin & Review*, 17(1), 1–14.
<https://doi.org/10.3758/PBR.17.1.1>
- Konishi, S., & Kitagawa, G. (Eds.). (2008). Bayesian information criteria. In *Information criteria and statistical modeling* (pp. 211–237). Springer.
https://doi.org/10.1007/978-0-387-71887-3_9
- Kopp, B., Lange, F., & Steinke, A. (2021). The reliability of the wisconsin card sorting test in clinical practice. *Assessment*, 28(1), 248–263.
<https://doi.org/10.1177/1073191119866257>
- Kovacs, K., & Conway, A. R. A. (2019). What is IQ? Life beyond “general intelligence”. *Current Directions in Psychological Science*, 28, 189–194.
<https://doi.org/10.1177/0963721419827275>
- Kraut, R., Olson, J., Banaji, M., Bruckman, A., Cohen, J., & Couper, M. (2004). Psychological research online: Report of board of scientific affairs’ advisory group on the conduct of research on the internet. *American Psychologist*, 59, 105–117. <https://doi.org/10.1037/0003-066X.59.2.105>
- Krawczyk, D. C. (2012). The cognition and neuroscience of relational reasoning. *Brain Research*, 1428, 13–23. <https://doi.org/10.1016/j.brainres.2010.11.080>

- Kuhl, J. (2000). The volitional basis of Personality Systems Interaction Theory: Applications in learning and treatment contexts. *International Journal of Educational Research*, 33(7), 665–703. [https://doi.org/10.1016/S0883-0355\(00\)00045-8](https://doi.org/10.1016/S0883-0355(00)00045-8)
- Kuhn, T. (1970). The structure of scientific revolutions. Chicago, IL: University of Chicago Press.
- Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2004). Academic performance, career potential, creativity, and job performance: Can one construct predict them all? *Journal of Personality and Social Psychology*, 86(1), 148–161. <https://doi.org/10.1037/0022-3514.86.1.148>
- Kuwabara, M., & Smith, L. B. (2012). Cross-cultural differences in cognitive development: Attention to relations and objects. *Journal of Experimental Child Psychology*, 113(1), 20–35. <https://doi.org/10.1016/j.jecp.2012.04.009>
- Kwan, J. L. Y., & Chan, W. (2011). Comparing standardized coefficients in structural equation modeling: A model reparameterization approach. *Behavior Research Methods*, 43(3), 730–745. <https://doi.org/10.3758/s13428-011-0088-6>
- Lamm, B., Keller, H., Teiser, J., Gudi, H., Yovsi, R. D., Freitag, C., Poloczek, S., Fassbender, I., Suhrke, J., Teubert, M., Vöhringer, I., Knopf, M., Schwarzer, G., & Lohaus, A. (2018). Waiting for the second treat: Developing culture-specific modes of self-regulation. *Child Development*, 89(3), e261–e277. <https://doi.org/10.1111/cdev.12847>
- Lan, X., Legare, C. H., Ponitz, C. C., Li, S., & Morrison, F. J. (2011). Investigating the links between the subcomponents of executive function and academic achievement: A cross-cultural analysis of Chinese and American preschoolers. *Journal of Experimental Child Psychology*, 108(3), 677–692. <https://doi.org/10.1016/j.jecp.2010.11.001>
- Lau, S. (1996). *Growing up the Chinese way: Chinese child and adolescent development*. Chinese University Press.

- Lawson, G. M., & Farah, M. J. (2017). Executive function as a mediator between SES and academic achievement throughout childhood. *International Journal of Behavioral Development*, 41(1), 94–104.
<https://doi.org/10.1177/0165025415603489>
- Lazar, A. S., Panin, F., Goodman, A. O. G., Lazic, S. E., Lazar, Z. I., Mason, S. L., Rogers, L., Murgatroyd, P. R., Watson, L. P. E., Singh, P., Borowsky, B., Shneerson, J. M., & Barker, R. A. (2015). Sleep deficits but no metabolic deficits in premanifest Huntington's disease. *Annals of Neurology*, 78(4), 630–648. <https://doi.org/10.1002/ana.24495>
- Lee, D. K. (2020). Data transformation: A focus on the interpretation. *Korean Journal of Anesthesiology*, 73(6), 503–508. <https://doi.org/10.4097/kja.20137>
- Lehmann, J., & Seufert, T. (2020). The interaction between text modality and the learner's modality preference influences comprehension and cognitive load. *Frontiers in Psychology*, 10, 2820. <https://doi.org/10.3389/fpsyg.2019.02820>
- Leung, L. (2015). Validity, reliability, and generalizability in qualitative research. *Journal of Family Medicine and Primary Care*, 4(3), 324–327.
<https://doi.org/10.4103/2249-4863.161306>
- Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest: A Journal of the American Psychological Society*, 13(3), 106–131.
<https://doi.org/10.1177/1529100612451018>
- Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, 49(4), 764–766.
<https://doi.org/10.1016/j.jesp.2013.03.013>
- Li, J. (2012). Cultural foundations of learning: East and west. In *Cambridge University Press*. Cambridge University Press.

- Li, X., Li, Y., & Wang, W. (2023). Long-lasting conceptual change in science education. *Science & Education*, 32(1), 123–168.
<https://doi.org/10.1007/s11191-021-00288-x>
- Liesefeld, H. R., & Janczyk, M. (2019). Combining speed and accuracy to control for speed-accuracy trade-offs. *Behavior Research Methods*, 51(1), 40–60.
<https://doi.org/10.3758/s13428-018-1076-x>
- Limón, M. (2002). Conceptual Change in History. In M. Limón & L. Mason (Eds.), *Reconsidering Conceptual Change: Issues in Theory and Practice* (pp. 259–289). Springer Netherlands. https://doi.org/10.1007/0-306-47637-1_14
- Lipina, S. J., & Evers, K. (2017). Neuroscience of childhood poverty: Evidence of impacts and mechanisms as vehicles of dialog with ethics. *Frontiers in Psychology*, 8. <https://www.frontiersin.org/articles/10.3389/fpsyg.2017.00061>
- Little, R. J., & Rubin, D. B. (2019). *Statistical analysis with missing data* (Vol. 793). John Wiley & Sons. <https://doi.org/10.1002/9781119482260>
- Liu, C. C., & Watanabe, T. (2012). Accounting for speed–accuracy tradeoff in perceptual learning. *Vision Research*, 61, 107–114.
<https://doi.org/10.1016/j.visres.2011.09.007>
- Lobo, M. P. (2019). *An open essay on open science*. OSF Preprints.
<https://doi.org/10.31219/osf.io/42jsg>
- Logan, G. D. (1994). On the ability to inhibit thought and action: A users’ guide to the stop signal paradigm. In *Inhibitory processes in attention, memory, and language* (pp. 189–239). Academic Press.
- Los, S. A. (1996). On the origin of mixing costs: Exploring information processing in pure and mixed blocks of trials. *Acta Psychologica*, 94(2), 145–188.
[https://doi.org/10.1016/0001-6918\(95\)00050-X](https://doi.org/10.1016/0001-6918(95)00050-X)
- Lövdén, M., Fratiglioni, L., Glymour, M. M., Lindenberger, U., & Tucker-Drob, E. M. (2020). Education and cognitive functioning across the life span.

- Psychological Science in the Public Interest*, 21(1), 6–41.
<https://doi.org/10.1177/1529100620920576>
- Lozano-Blasco, R., Quílez-Robres, A., Usán, P., Salavera, C., & Casanovas-López, R. (2022). Types of intelligence and academic performance: A systematic review and meta-analysis. *Journal of Intelligence*, 10(4), 123.
<https://doi.org/10.3390/jintelligence10040123>
- Lü, X., & Zhang, J. (1999). Reading efficiency: A comparative study of English and Chinese orthographies. *Reading Research and Instruction*, 38(4), 301–317.
<https://doi.org/10.1080/19388079909558298>
- Lubczyk, T., Lukács, G., & Ansorge, U. (2022). Speed versus accuracy instructions in the response time concealed information test. *Cognitive Research: Principles and Implications*, 7(1), 3. <https://doi.org/10.1186/s41235-021-00352-8>
- Lynn, R., & Mikk, J. (2009). National IQs predict educational attainment in math, reading and science across 56 nations. *Intelligence*, 37(3), 305–310.
<https://doi.org/10.1016/j.intell.2009.01.002>
- Magalhães, S., Carneiro, L., Limpo, T., & Filipe, M. (2020). Executive functions predict literacy and mathematics achievements: The unique contribution of cognitive flexibility in grades 2, 4, and 6. *Child Neuropsychology: A Journal on Normal and Abnormal Development in Childhood and Adolescence*, 26(7), 934–952. <https://doi.org/10.1080/09297049.2020.1740188>
- Mann, E. M., Ikeda, Y., Mueller, C. W., Takahashi, A., Tao, K. T., Humris, E., Li, B. L., & Chin, D. (1992). Cross-cultural differences in rating hyperactive-disruptive behaviors in children. *The American Journal of Psychiatry*, 149(11), 1539–1542. <https://doi.org/10.1176/ajp.149.11.1539>
- Marcoulides, K. M., & Raykov, T. (2019). Evaluation of variance inflation factors in regression models using latent variable modeling methods. *Educational and Psychological Measurement*, 79(5), 874–882.
<https://doi.org/10.1177/0013164418817803>

- Maris, G., & Maas, H. (2012). Speed-accuracy response models: Scoring rules based on response time and accuracy. *Psychometrika*, 4.
<https://doi.org/10.1007/s11336-012-9288-y>
- Mason, L., & Zaccoletti, S. (2021). Inhibition and conceptual learning in science: A review of studies. *Educational Psychology Review*, 33(1), 181–212.
<https://doi.org/10.1007/s10648-020-09529-x>
- Mason, L., Zaccoletti, S., Carretti, B., Scrimin, S., & Diakidoy, I.-A. N. (2019). The role of inhibition in conceptual learning from refutation and standard expository texts. *International Journal of Science and Mathematics Education*, 17(3), 483–501. Scopus. <https://doi.org/10.1007/s10763-017-9874-7>
- Masson, S., Potvin, P., Riopel, M., & Foisy, L.-M. B. (2014a). Differences in brain activation between novices and experts in science during a task involving a common misconception in electricity. *Mind, Brain, and Education*, 8(1), 44–55. <https://doi.org/10.1111/mbe.12043>
- Mathy, F., Chekaf, M., & Cowan, N. (n.d.). Simple and complex working memory tasks allow similar benefits of information compression. *Journal of Cognition*, 1(1). <https://doi.org/10.5334/joc.31>
- Mayer, R. E., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology*, 93(1), 187–198. <https://doi.org/10.1037/0022-0663.93.1.187>
- McDermott, J. M., Pérez-Edgar, K., & Fox, N. A. (2007). Variations of the flanker paradigm: Assessing selective attention in young children. *Behavior Research Methods*, 39(1), 62–70. <https://doi.org/10.3758/BF03192844>
- McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, 37(1), 1–10. <https://doi.org/10.1016/j.intell.2008.08.004>

- McNeil, N. M., & Alibali, M. W. (2005). Why won't you change your mind? Knowledge of operational patterns hinders learning and performance on equations. *Child Development*, 76(4), 883–899.
<https://www.jstor.org/stable/3696735>.
- McNeish, D., & Kelley, K. (2019). Fixed effects models versus mixed effects models for clustered data: Reviewing the approaches, disentangling the differences, and making recommendations. *Psychological Methods*, 24(1), 20–35.
<https://doi.org/10.1037/met0000182>
- Meier, M. A., Wambacher, D., Vogel, S. E., & Grabner, R. H. (2022). Interference between naïve and scientific theories in mathematics and science: An fMRI study comparing mathematicians and non-mathematicians. *Trends in Neuroscience and Education*, 29, 100194.
<https://doi.org/10.1016/j.tine.2022.100194>
- Menz, C., Spinath, B., & Seifried, E. (2021). Misconceptions die hard: Prevalence and reduction of wrong beliefs in topics from educational psychology among preservice teachers. *European Journal of Psychology of Education*, 36(2), 477–494. <https://doi.org/10.1007/s10212-020-00474-5>
- Meuter, R. F. I., & Allport, A. (1999). Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language*, 40(1), 25–40. <https://doi.org/10.1006/jmla.1998.2602>
- Miller, L. M. S., & Stine-Morrow, E. A. L. (1998). Aging and the effects of knowledge on on-line reading strategies. *The Journals of Gerontology: Series B: Psychological Sciences and Social Sciences*, 53(4), P223–P233.
<https://doi.org/10.1093/geronb/53B.4.P223>
- Miller-Cotto, D., Smith, L. V., Wang, A. H., & Ribner, A. D. (2022). Changing the conversation: A culturally responsive perspective on executive functions, minoritized children and their families. *Infant and Child Development*, 31(1).
<https://doi.org/10.1002/icd.2286>

- Ministry of Education of the People's Republic of China. (2022). Curriculum Plan for Compulsory Education (Version 2022) [PDF]. Retrieved from <http://www.moe.gov.cn/srcsite/A26/s8001/202204/W020220420582343217634.pdf>
- Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science*, 21(1), 8–14. <https://doi.org/10.1177/0963721411429458>
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex ‘Frontal Lobe’ tasks: A latent variable analysis. *Cognitive Psychology*, 41(1), 49–100. <https://doi.org/10.1006/cogp.1999.0734>
- Moisala, M., Salmela, V., Carlson, S., Salmela-Aro, K., Lonka, K., Hakkarainen, K., & Alho, K. (2018). Neural activity patterns between different executive tasks are more similar in adulthood than in adolescence. *Brain and Behavior*, 8(9), e01063. <https://doi.org/10.1002/brb3.1063>
- Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences*, 7(3), 134–140. [https://doi.org/10.1016/s1364-6613\(03\)00028-7](https://doi.org/10.1016/s1364-6613(03)00028-7)
- Moreno, R., & Mayer, R. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology*, 91, 358–368. <https://doi.org/10.1037/0022-0663.91.2.358>
- Muhammed T, S., & Mathew, S. K. (2022). The disaster of misinformation: A review of research in social media. *International Journal of Data Science and Analytics*, 13(4), 271–285. <https://doi.org/10.1007/s41060-022-00311-6>
- Munakata, Y., Herd, S. A., Chatham, C. H., Depue, B. E., Banich, M. T., & O'Reilly, R. C. (2011). A unified framework for inhibitory control. *Trends in Cognitive Sciences*, 15(10), 453–459. <https://doi.org/10.1016/j.tics.2011.07.011>

- Munakata, Y., Yanaoka, K., Doebel, S., Guild, R. M., Michaelson, L. E., & Saito, S. (2020). Group influences on children's delay of gratification: Testing the roles of culture and personal connections. *Collabra: Psychology*, 6(1), 1. <https://doi.org/10.1525/collabra.265>
- Murphy, F. C., Sahakian, B. J., Rubinsztein, J. S., Michael, A., Rogers, R. D., Robbins, T. W., & Paykel, E. S. (1999). Emotional bias and inhibitory control processes in mania and depression. *Psychological Medicine*, 29(6), 1307–1321. <https://doi.org/10.1017/s0033291799001233>
- Murphy, P. K., & Mason, L. (2006). Changing Knowledge and Beliefs. In *Handbook of educational psychology* (pp. 305–324). Lawrence Erlbaum Associates Publishers.
- Nadelson, L. S., Heddy, B. C., Jones, S., Taasobshirazi, G., & Johnson, M. (2018). Conceptual change in science teaching and learning: Introducing the dynamic model of conceptual change. *International Journal of Educational Psychology*, 7(2), 151–195.
- National Center for Education Statistics. (2021). *TIMSS 2019 U.S. Highlights Web Report (NCES 2021-021)*. Institute of Education Sciences, U.S. Department of Education. <https://nces.ed.gov/timss/results19/index.asp>
- Newell, T. (2014). The Empirical Paradigm. In T. Newell (Ed.), *Five Paradigms for Education: Foundational Views and Key Issues* (pp. 74–99). Palgrave Macmillan US. https://doi.org/10.1057/9781137391803_4
- Newman, D. A. (2014). Missing data: Five practical guidelines. *Organizational Research Methods*, 17(4), 372–411. <https://doi.org/10.1177/1094428114548590>
- Ng, F. F.-Y., & Wei, J. (2020). Delving into the minds of Chinese parents: What beliefs motivate their learning-related practices? *Child Development Perspectives*, 14(1), 61–67. <https://doi.org/10.1111/cdep.12358>

- Niebaum, J. C., & Munakata, Y. (2023). Why doesn't executive function training improve academic achievement? Rethinking individual differences, relevance, and engagement from a contextual framework. *Journal of Cognition and Development, 24*(2), 241–259.
<https://doi.org/10.1080/15248372.2022.2160723>
- Nieuwenhuis, S. & Nick Yeung. (2005). Neural mechanisms of attention and control: Losing our inhibitions? *Nature Neuroscience, 8*(12), 1631–1633.
<https://doi.org/10.1038/nn1205-1631>
- Nigg, J. T. (2000). On inhibition/disinhibition in developmental psychopathology: Views from cognitive and personality psychology and a working inhibition taxonomy. *Psychological Bulletin, 126*(2), 220–246.
<https://doi.org/10.1037/0033-2909.126.2.220>
- Nikitenko, T., Chowdhury, N., Puri, R., & He, J. L. (2019). Response inhibition in humans: A whistle stop review. *Journal of Neurophysiology, 123*(3), 861–864.
<https://doi.org/10.1152/jn.00572.2019>
- Nozari, N., Trueswell, J. C., & Thompson-Schill, S. L. (2016). The interplay of local attraction, context and domain-general cognitive control in activation and suppression of semantic distractors during sentence comprehension. *Psychonomic Bulletin & Review, 23*(6), 1942–1953.
<https://doi.org/10.3758/s13423-016-1068-8>
- Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior, 32*(2), 303–330.
<https://doi.org/10.1007/s11109-010-9112-2>
- Nzinga, K., Rapp, D. N., Leatherwood, C., Easterday, M., Rogers, L. O., Gallagher, N., & Medin, D. L. (2018). Should social scientists be distanced from or engaged with the people they study? *Proceedings of the National Academy of Sciences of the United States of America, 115*(45), 11435–11441.
<https://doi.org/10.1073/pnas.1721167115>

- O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity: International Journal of Methodology*, 41(5), 673–690. <https://doi.org/10.1007/s11135-006-9018-6>
- OECD. (2022). *OECD guidelines for citizen participation processes*. OECD Public Governance Reviews. OECD Publishing. <https://doi.org/10.1787/f765caf6-en>
- Oh, S., & Lewis, C. (2008). Korean preschoolers' advanced inhibitory control and its relation to other executive skills and mental state understanding. *Child Development*, 79(1), 80–99. <https://doi.org/10.1111/j.1467-8624.2007.01112.x>
- Oliveira, G. S. de, & Baggs, E. (2023). Psychology's WEIRD Problems. *Elements in Psychology and Culture*. <https://doi.org/10.1017/9781009303538>
- Osborne, J. W., & Overbay, A. (n.d.). *The power of outliers (and why researchers should ALWAYS check for them)*. <https://doi.org/10.7275/QF69-7K43>
- Osborne, R., & Freyberg, P. (1985). *Learning in Science. The Implications of Children's Science*. Heinemann Educational Books, Inc.
- Park, C., & Cho, B. (2003). Development of robust design under contaminated and non-normal data. *Quality Engineering*, 15, 463–469. <https://doi.org/10.1081/QEN-120018045>
- Peng, A., Kirkham, N. Z., & Mareschal, D. (2018). Task switching costs in preschool children and adults. *Journal of Experimental Child Psychology*, 172, 59–72. <https://doi.org/10.1016/j.jecp.2018.01.019>
- Pérez-Arce, P. (1999). The influence of culture on cognition. *Archives of Clinical Neuropsychology*, 14(7), 581–592. [https://doi.org/10.1016/S0887-6177\(99\)00007-4](https://doi.org/10.1016/S0887-6177(99)00007-4)
- Peters, M. A. (2015). Socrates and Confucius: The cultural foundations and ethics of learning. *Educational Philosophy and Theory*, 47(5), 423–427. <https://doi.org/10.1080/00131857.2014.930232>

- Petrides, M., & Milner, B. (1982). Deficits on subject-ordered tasks after frontal- and temporal-lobe lesions in man. *Neuropsychologia*, 20(3), 249–262.
[https://doi.org/10.1016/0028-3932\(82\)90100-2](https://doi.org/10.1016/0028-3932(82)90100-2)
- Piaget, J. (1929). *The child's conception of the world*. London: Kegan Paul, Trench & Trubner.
- Philipp, A. M., & Koch, I. (2009). Inhibition in language switching: What is inhibited when switching between languages in naming tasks? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(5), 1187–1195.
<https://doi.org/10.1037/a0016376>
- Pintrich, P. R., Marx, R. W., & Boyle, R. A. (1993). Beyond cold conceptual change: The role of motivational beliefs and classroom contextual factors in the process of conceptual change. *Review of Educational Research*, 63(2), 167–199. <https://doi.org/10.2307/1170472>
- Poling, A., Methot, L. L., & LeSage, M. G. (1995). Within-subject experimental designs. In A. Poling, L. L. Methot, & M. G. LeSage (Eds.), *Fundamentals of behavior analytic research* (pp. 79–102). Springer US.
https://doi.org/10.1007/978-1-4899-1436-1_5
- Poljac, E., Koch, I., & Bekkering, H. (2009). Dissociating restart cost and mixing cost in task switching. *Psychological Research*, 73(3), 407–416.
<https://doi.org/10.1007/s00426-008-0151-9>
- Pollack, I., Johnson, L. B., & Knaff, P. R. (1959). Running memory span. *Journal of Experimental Psychology*, 57(3), 137–146.
<http://dx.doi.org.pitt.idm.oclc.org/10.1037/h0046137>
- Popper, K. R. (1959). *The logic of scientific discovery* (p. 480). Basic Books.
- Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education*, 66(2), 211–227.
<https://doi.org/10.1002/sce.3730660207>

- Postle, B. R., Brush, L. N., & Nick, A. M. (2004). Prefrontal cortex and the mediation of proactive interference in working memory. *Cognitive, Affective & Behavioral Neuroscience*, 4(4), 600–608.
- Potvin, P., Chastenay, P., Thibault, F., Riopel, M., Ahr, E., & Brault Foisy, L.-M. (2023). An understanding of falling bodies across schooling and experience based on the conceptual prevalence framework. *Disciplinary and Interdisciplinary Science Education Research*, 5(1), 8.
<https://doi.org/10.1186/s43031-023-00075-4>
- Potvin, P., & Cyr, G. (2017). Toward a durable prevalence of scientific conceptions: Tracking the effects of two interfering misconceptions about buoyancy from preschoolers to science teachers. *Journal of Research in Science Teaching*, 54(9), 1121–1142. <https://doi.org/10.1002/tea.21396>
- Potvin, P., Malenfant-Robichaud, G., Cormier, C., & Masson, S. (2020). Coexistence of misconceptions and scientific conceptions in chemistry professors: A mental chronometry and fMRI study. *Frontiers in Education*, 5, 542458.
<https://doi.org/10.3389/educ.2020.542458>
- Potvin, P., Masson, S., Lafortune, S., & Cyr, G. (2015). Persistence of the intuitive conception that heavier objects sink more: A reaction time study with different levels of interference. *International Journal of Science and Mathematics Education*, 13(1), 21–43. <https://doi.org/10.1007/s10763-014-9520-6>
- Potvin, P., Turmel, É., & Masson, S. (2014). Linking neuroscientific research on decision making to the educational context of novice students assigned to a multiple-choice scientific task involving common misconceptions about electrical circuits. *Frontiers in Human Neuroscience*, 8.
<https://doi.org/10.3389/fnhum.2014.00014>
- Pribram, K. H. (1973). The primate frontal cortex: Executive of the brain. In *Psychophysiology of the frontal lobes* (pp. xii, 332–xii, 332). Academic Press.

- R Core Team (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Ragland, D. R. (1992). Dichotomizing continuous outcome variables: Dependence of the magnitude of association and statistical power on the cutpoint. *Epidemiology (Cambridge, Mass.)*, 3(5), 434–440. <https://doi.org/10.1097/00001648-199209000-00009>
- Raud, L., Westerhausen, R., Dooley, N., & Huster, R. J. (2020). Differences in unity: The go/no-go and stop signal tasks rely on different mechanisms. *NeuroImage*, 210, 116582. <https://doi.org/10.1016/j.neuroimage.2020.116582>
- Reimers, S., & Maylor, E. A. (2005). Task switching across the life span: Effects of age on general and specific switch costs. *Developmental Psychology*, 41(4), 661–671. <https://doi.org/10.1037/0012-1649.41.4.661>
- Reips, U.-D. (2002). Standards for internet-based experimenting. *Experimental Psychology*, 49(4), 243–256. <https://doi.org/10.1027//1618-3169.49.4.243>
- Rhodes, S. M., Booth, J. N., Palmer, L. E., Blythe, R. A., Delibegovic, M., & Wheate, N. J. (2016). Executive functions predict conceptual learning of science. *British Journal of Developmental Psychology*, 34(2), 261–275. <https://doi.org/10.1111/bjdp.12129>
- Ritchie, S. J., Bates, T. C., & Deary, I. J. (2015). Is education associated with improvements in general cognitive ability, or in specific skills? *Developmental Psychology*, 51(5), 573–582. <https://doi.org/10.1037/a0038981>
- Ritchie, S. J., Bates, T. C., Der, G., Starr, J. M., & Deary, I. J. (2013). Education is associated with higher later life IQ scores, but not with faster cognitive processing speed. *Psychology and Aging*, 28(2), 515–521. <https://doi.org/10.1037/a0030820>

- Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, 124(2), 207–231. <https://doi.org/10.1037/0096-3445.124.2.207>
- Roos, L., Beauchamp, K., Flannery, J., & Fisher, P. (2017). Cultural contributions to childhood executive function. *Journal of Cognition and Culture*.
- Roozenbeek, J., & Van der Linden, S. (2019). Fake news game confers psychological resistance against online misinformation. *Palgrave Communications*, 5(1), 1–10.
- Rosopa, P. J., Schaffer, M. M., & Schroeder, A. N. (2013). Managing heteroscedasticity in general linear models. *Psychological Methods*, 18(3), 335–351. <https://doi.org/10.1037/a0032553>
- Rubens, M. T., & Zanto, T. P. (2011). Parameterization of transcranial magnetic stimulation. *Journal of Neurophysiology*, 107(5), 1257–1259. <https://doi.org/10.1152/jn.00716.2011>
- Rubin, O., & Meiran, N. (2005). On the origins of the task mixing cost in the cuing task-switching paradigm. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 31(6), 1477–1491. <https://doi.org/10.1037/0278-7393.31.6.1477>
- Rueda, M. R., Fan, J., McCandliss, B. D., Halparin, J. D., Gruber, D. B., Lercari, L. P., & Posner, M. I. (2004). Development of attentional networks in childhood. *Neuropsychologia*, 42(8), 1029-1040. <https://doi.org/10.1016/j.neuropsychologia.2003.12.012>
- Rudd, J. R., O’Callaghan, L., & Williams, J. (2019). Physical education pedagogies built upon theories of movement learning: How can environmental constraints be manipulated to improve children’s executive function and self-regulation skills? *International Journal of Environmental Research and Public Health*, 16(9), Article 9. <https://doi.org/10.3390/ijerph16091630>

- Rutter, M. (1981). Stress, coping and development: Some issues and some questions*. *Journal of Child Psychology and Psychiatry*, 22(4), 323–356.
<https://doi.org/10.1111/j.1469-7610.1981.tb00560.x>
- Ryu, E. (2015). Multiple group analysis in multilevel structural equation model across level 1 groups. *Multivariate Behavioral Research*, 50(3), 300–315.
<https://doi.org/10.1080/00273171.2014.1003769>
- Sabbagh, M. A., Xu, F., Carlson, S. M., Moses, L. J., & Lee, K. (2006). The development of executive functioning and theory of mind. A comparison of Chinese and U.S. preschoolers. *Psychological Science*, 17(1), 74–81.
<https://doi.org/10.1111/j.1467-9280.2005.01667.x>
- Sainani, K. L. (2012). Dealing with non-normal data. *PM&R*, 4(12), 1001–1005.
<https://doi.org/10.1016/j.pmrj.2012.10.013>
- Sainani, K. L. (2016). The value of scatter plots. *PM&R*, 8(12), 1213–1217.
<https://doi.org/10.1016/j.pmrj.2016.10.018>
- Salthouse, T. A. (2005). Relations between cognitive abilities and measures of executive functioning. *Neuropsychology*, 19(4), 532–545.
<https://doi.org/10.1037/0894-4105.19.4.532>
- Schachar, R., Logan, G. D., Robaey, P., Chen, S., Ickowicz, A., & Barr, C. (2007). Restraint and cancellation: Multiple inhibition deficits in attention deficit hyperactivity disorder. *Journal of Abnormal Child Psychology*, 35(2), 229–238. <https://doi.org/10.1007/s10802-006-9075-2>
- Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7(2), 147–177. <https://doi.org/10.1037/1082-989X.7.2.147>
- Schauble, L. (1990). Belief revision in children: The role of prior knowledge and strategies for generating evidence. *Journal of Experimental Child Psychology*, 49(1), 31–57. [https://doi.org/10.1016/0022-0965\(90\)90048-D](https://doi.org/10.1016/0022-0965(90)90048-D)

- Scheidecker, G., Chaudhary, N., Oppong, S., Röttger-Rössler, B., & Keller, H. (2022). Different is not deficient: Respecting diversity in early childhood development. *The Lancet Child & Adolescent Health*, 6(12), e24–e25. [https://doi.org/10.1016/S2352-4642\(22\)00277-2](https://doi.org/10.1016/S2352-4642(22)00277-2)
- Schirmbeck, K., Rao, N., & Maehler, C. (2020). Similarities and differences across countries in the development of executive functions in children: A systematic review. *Infant and Child Development*, 29(1), e2164. <https://doi.org/10.1002/icd.2164>
- Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation coefficients: Appropriate use and interpretation. *Anesthesia and Analgesia*, 126(5), 1763–1768. <https://doi.org/10.1213/ANE.0000000000002864>
- Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research*, 99(6), 323–338. <https://doi.org/10.3200/JOER.99.6.323-338>
- Schumacker, R. E., & Lomax, R. G. (2016). *A beginner's guide to structural equation modeling* (4th ed.). Routledge.
- Shanmugan, S., & Satterthwaite, T. D. (2016). Neural markers of the development of executive function: Relevance for education. *Current Opinion in Behavioral Sciences*, 10, 7–13. <https://doi.org/10.1016/j.cobeha.2016.04.007>
- Sharpiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality. *Biometrika*, 52(3–4), 591–611. <https://doi.org/10.1093/biomet/52.3-4.591>
- Sharp, D., Cole, M., Lave, C., Ginsburg, H. P., Brown, A. L., & French, L. A. (1979). Education and cognitive development: The evidence from experimental research. *Monographs of the Society for Research in Child Development*, 44(1/2), 1–112. <https://doi.org/10.2307/3181586>

- Shipstead, Z., Redick, T. S., & Engle, R. W. (2010). Does working memory training generalize? *Psychologica Belgica*, 50(3–4), 245. <https://doi.org/10.5334/pb-50-3-4-245>
- Shtulman, A. (2023). Children’s susceptibility to online misinformation. *Current Opinion in Psychology*, 101753.
- Shtulman, A., & Harrington, K. (2016). Tensions between science and intuition across the lifespan. *Topics in Cognitive Science*, 8, 118–137. <https://doi.org/10.1111/tops.12174>
- Shtulman, A., & Legare, C. H. (2020). Competing explanations of competing explanations: Accounting for conflict between scientific and folk explanations. *Topics in Cognitive Science*, 12(4), 1337–1362. <https://doi.org/10.1111/tops.12483>
- Shtulman, A., & Lombrozo, T. (2016). Bundles of contradiction: A coexistence view of conceptual change. In *Core knowledge and conceptual change* (pp. 53–71). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780190467630.003.0004>
- Shtulman, A., & Valcarcel, J. (2012). Scientific knowledge suppresses but does not supplant earlier intuitions. *Cognition*, 124(2), 209–215. <https://doi.org/10.1016/j.cognition.2012.04.005>
- Simon, J. R., & Rudell, A. P. (1967). Auditory S-R compatibility: The effect of an irrelevant cue on information processing. *Journal of Applied Psychology*, 51(3), 300–304. <https://doi.org/10.1037/h0020586>
- Singh, S. (2003). Simple random sampling. In S. Singh (Ed.), *Advanced sampling theory with applications: How Michael 'selected' Amy* (Vol. 1, pp. 71–136). Springer Netherlands. https://doi.org/10.1007/978-94-007-0789-4_2
- Smith, C., Carey, S., & Wiser, M. (1985). On differentiation: A case study of the development of the concepts of size, weight, and density. *Cognition*, 21(3), 177–237. [https://doi.org/10.1016/0010-0277\(85\)90025-3](https://doi.org/10.1016/0010-0277(85)90025-3)

- Snyder, H. R., Banich, M. T., & Munakata, Y. (2011). Choosing our words: Retrieval and selection processes recruit shared neural substrates in left ventrolateral prefrontal cortex. *Journal of Cognitive Neuroscience*, 23(11), 3470–3482. https://doi.org/10.1162/jocn_a_00023
- Soper, D.S. (2024). *A-priori sample size calculator for structural equation models* [Software]. Available from <https://www.danielsoper.com/statcalc>
- Spector, A., & Biederman, I. (1976). Mental set and mental shift revisited. *The American Journal of Psychology*, 89(4), 669–679. <https://doi.org/10.2307/1421465>
- St Clair-Thompson, H., & Sykes, S. (2010). Scoring methods and the predictive ability of working memory tasks. *Behavior Research Methods*, 42(4), 969–975. <https://doi.org/10.3758/BRM.42.4.969>
- Stahl, C., Voss, A., Schmitz, F., Nuszbaum, M., Tüscher, O., Lieb, K., & Klauer, K. C. (2014). Behavioral components of impulsivity. *Journal of Experimental Psychology. General*, 143(2), 850–886. <https://doi.org/10.1037/a0033981>
- Staub, F. C., & Stern, E. (2002). The nature of teachers’ pedagogical content beliefs matters for students’ achievement gains: Quasi-experimental evidence from elementary mathematics. *Journal of Educational Psychology*, 94(2), 344–355. <https://doi.org/10.1037/0022-0663.94.2.344>
- Stavy, R., & Tirosh, D. (2000). *How students (mis-)understand science and mathematics: Intuitive rules*. Teachers College Press.
- Steinhauser, M., & Hübner, R. (2007). Automatic activation of task-related representations in task shifting. *Memory & Cognition*, 35(1), 138–155. <https://doi.org/10.3758/BF03195950>
- Sternberg, R. J. (2008). Increasing fluid intelligence is possible after all. *Proceedings of the National Academy of Sciences*, 105(19), 6791–6792. <https://doi.org/10.1073/pnas.0803396105>

- Sterne, J. A. C., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., Wood, A. M., & Carpenter, J. R. (2009). Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls. *BMJ*, 338, b2393. <https://doi.org/10.1136/bmj.b2393>
- Stins, J. F., Polderman, J. C. T., Boomsma, D. I., & de Geus, E. J. C. (2008). Conditional accuracy in response interference tasks: Evidence from the Eriksen flanker task and the spatial conflict task. *Advances in Cognitive Psychology*, 3(3), 409–417. <https://doi.org/10.2478/v10053-008-0005-4>
- Streiner, D. L. (2005). Finding our way: An introduction to path analysis. *The Canadian Journal of Psychiatry*, 50(2), 115–122. <https://doi.org/10.1177/070674370505000207>
- Streiner, D. L. (2006). Building a better model: An introduction to structural equation modelling. *The Canadian Journal of Psychiatry*, 51(5), 317–324. <https://doi.org/10.1177/070674370605100507>
- Stricker, J., Vogel, S. E., Schöneburg-Lehnert, S., Krohn, T., Dögnitz, S., Jud, N., Spirk, M., Windhaber, M.-C., Schneider, M., & Grabner, R. H. (2021). Interference between naïve and scientific theories occurs in mathematics and is related to mathematical achievement. *Cognition*, 214, 104789. <https://doi.org/10.1016/j.cognition.2021.104789>
- Stroop, J. R. (1992). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology: General*, 121(1), 15–23. <https://doi.org/10.1037/0096-3445.121.1.15>
- Suarez-Lledo, V., & Alvarez-Galvez, J. (2021). Prevalence of health misinformation on social media: Systematic review. *Journal of Medical Internet Research*, 23(1), e17187. <https://doi.org/10.2196/17187>
- Suchotzki, K., Verschuere, B., & Gamer, M. (2021). How vulnerable is the reaction time concealed information test to faking? *Journal of Applied Research in*

- Memory and Cognition*, 10(2), 268–277.
<https://doi.org/10.1016/j.jarmac.2020.10.003>
- Taherdoost, H. (2016). *Sampling methods in research methodology; How to choose a sampling technique for research* (SSRN Scholarly Paper ID 3205035). Social Science Research Network. <https://papers.ssrn.com/abstract=3205035>
- Tardiff, N., Bascandziev, I., Carey, S., & Zaitchik, D. (2020). Specifying the domain-general resources that contribute to conceptual construction: Evidence from the child's acquisition of vitalist biology. *Cognition*, 195. Scopus.
<https://doi.org/10.1016/j.cognition.2019.104090>
- Tardiff, N., Bascandziev, I., Sandor, K., Carey, S., & Zaitchik, D. (2017). Some consequences of normal aging for generating conceptual explanations: A case study of vitalist biology. *Cognitive Psychology*, 95, 145–163.
<https://doi.org/10.1016/j.cogpsych.2017.04.004>
- Tarka, P. (2018). An overview of structural equation modeling: Its beginnings, historical development, usefulness and controversies in the social sciences. *Quality & Quantity*, 52(1), 313–354. <https://doi.org/10.1007/s11135-017-0469-8>
- Taylor, A. K., & Kowalski, P. (2004). Naïve psychological science: The prevalence, strength, and sources of misconceptions. *The Psychological Record*, 54(1), 15–25. <https://doi.org/10.1007/BF03395459>
- Thibault, F., & Potvin, P. (2018). Executive function as a predictor of physics-related conceptual change. *Neuroeducation*, 5(2), 119–126.
<https://doi.org/10.24046/neuroed.20180502.119>
- Thorell, L. B., Veleiro, A., Siu, A. F. Y., & Mohammadi, H. (2013). Examining the relation between ratings of executive functioning and academic achievement: Findings from a cross-cultural study. *Child Neuropsychology*, 19(6), 630–638.
<https://doi.org/10.1080/09297049.2012.727792>

- Tiego, J., Testa, R., Bellgrove, M. A., Pantelis, C., & Whittle, S. (2018). A hierarchical model of inhibitory control. *Frontiers in Psychology*, 9. <https://doi.org/10.3389/fpsyg.2018.01339>
- Tippett, C. D. (2010). Refutation text in science education: A review of two decades of research. *International Journal of Science and Mathematics Education*, 8(6), 951–970. <https://doi.org/10.1007/s10763-010-9203-x>
- Tobin, J., Hsueh, Y., & Karasawa, M. (2009). *Preschool in Three Cultures Revisited: China, Japan, and the United States*. Chicago: University of Chicago Press.
- Toplak, M. E., West, R. F., & Stanovich, K. E. (2013). Practitioner review: Do performance-based measures and ratings of executive function assess the same construct? *Journal of Child Psychology and Psychiatry*, 54(2), 131–143. <https://doi.org/10.1111/jcpp.12001>
- Treagust, D. F., & Duit, R. (2008). Conceptual change: A discussion of theoretical, methodological and practical challenges for science education. *Cultural Studies of Science Education*, 3(2), 297–328. <https://doi.org/10.1007/s11422-008-9090-4>
- Triandis, H.C. (1996) The psychological measurement of cultural syndromes. *American Psychologist*, 51, 407-415. <https://doi.org/10.1037/0003-066X.51.4.407>
- Tsigaridis, K. G., Rui, W., & Ellefson, M. (2022). The intercorrelation between executive function, physics problem solving, mathematical, and matrix reasoning skills: Reflections from a small-scale experiment. *Cambridge Educational Research e-Journal*, 9, 210-230. <https://doi.org/10.17863/CAM.90562>
- Tweed, R. G., & Lehman, D. R. (2002). Learning considered within a cultural context. Confucian and Socratic approaches. *The American Psychologist*, 57(2), 89–99. <https://doi.org/10.1037//0003-066x.57.2.89>

- Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, *114*(1), 104–132. <https://doi.org/10.1037/0033-295X.114.1.104>
- Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., & Newcombe, N. S. (2013). The malleability of spatial skills: A meta-analysis of training studies. *Psychological Bulletin*, *139*(2), 352–402. <https://doi.org/10.1037/a0028446>
- van de Vijver, F., & Hambleton, R. K. (1996). Translating tests: Some practical guidelines. *European Psychologist*, *1*(2), 89–99. <https://doi.org/10.1027/1016-9040.1.2.89>
- van den Wildenberg, W. P. M., & van der Molen, M. W. (2004). Developmental trends in simple and selective inhibition of compatible and incompatible responses. *Journal of Experimental Child Psychology*, *87*(3), 201–220. <https://doi.org/10.1016/j.jecp.2003.11.003>
- van der Linden, S., Leiserowitz, A., Rosenthal, S., & Maibach, E. (2017). Inoculating the public against misinformation about climate change. *Global Challenges*, *1*(2), 1600008. <https://doi.org/10.1002/gch2.201600008>
- Van Der Maas, H. L. J., Dolan, C. V., Grasman, R. P. P. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. J. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review*, *113*(4), 842–861. <https://doi.org/10.1037/0033-295X.113.4.842>
- van der Sluis, S., de Jong, P. F., & van der Leij, A. (2007). Executive functioning in children, and its relations with reasoning, reading, and arithmetic. *Intelligence*, *35*(5), 427–449. <https://doi.org/10.1016/j.intell.2006.09.001>
- van der Ven, S. H. G., Kroesbergen, E. H., Boom, J., & Leseman, P. P. M. (2013). The structure of executive functions in children: A closer examination of

- inhibition, shifting, and updating. *The British Journal of Developmental Psychology*, 31(Pt 1), 70–87. <https://doi.org/10.1111/j.2044-835X.2012.02079.x>
- Vatcheva, K. P., Lee, M., McCormick, J. B., & Rahbar, M. H. (2016). Multicollinearity in regression analyses conducted in epidemiologic studies. *Epidemiology*, 6(2), 227. <https://doi.org/10.4172/2161-1165.1000227>
- Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. *Learning and Instruction*, 4(1), 45–69. [https://doi.org/10.1016/0959-4752\(94\)90018-3](https://doi.org/10.1016/0959-4752(94)90018-3)
- Vosniadou, S. (2013a). Conceptual change in learning and instruction: The framework theory approach. In *International Handbook of Research on Conceptual Change* (2nd ed.). Routledge.
- Vosniadou, S. (2013b). *International Handbook of Research on Conceptual Change*. Routledge.
- Vosniadou, S. (2019). The development of students' understanding of science. *Frontiers in Education*, 4. <https://doi.org/10.3389/feduc.2019.00032>
- Vosniadou, S., Pnevmatikos, D., & Makris, N. (2018). The role of executive function in the construction and employment of scientific and mathematical concepts that require conceptual change learning. *Neuroeducation*, 5(2), 62–72. <https://doi.org/10.24046/neuroed.20180502.62>
- Vosniadou, S., Pnevmatikos, D., Makris, N., Lepenioti, D., Eikospentaki, K., Chountala, A., & Kyrianakis, G. (2018). The recruitment of shifting and inhibition in on-line science and mathematics tasks. *Cognitive Science*. <https://doi.org/10.1111/cogs.12624>
- Wang, S., & Baillargeon, R. (2008). Detecting impossible changes in infancy: A three-system account. *Trends in Cognitive Sciences*, 12(1), 17–23. <https://doi.org/10.1016/j.tics.2007.10.012>

- Wang, T., Li, C., Ren, X., & Schweizer, K. (2021). How executive processes explain the overlap between working memory capacity and fluid intelligence: A test of process overlap theory. *Journal of Intelligence*, 9(2), 21.
<https://doi.org/10.3390/jintelligence9020021>
- Wang, Y. A., & Rhemtulla, M. (2021). Power analysis for parameter estimation in structural equation modeling: A discussion and tutorial. *Advances in Methods and Practices in Psychological Science*, 4(1).
<https://doi.org/10.1177/2515245920918253>
- Wang, Y., Zhou, X., Peng, X., & Hu, X. (2022). Task switching involves working memory: Evidence from neural representation. *Frontiers in Psychology*, 13, 1003298. <https://doi.org/10.3389/fpsyg.2022.1003298>
- Ward, G., Roberts, M. J., & Phillips, L. H. (2001). Task-switching costs, stroop-costs, and executive control: A correlational study. *The Quarterly Journal of Experimental Psychology Section A*, 54(2), 491–511.
<https://doi.org/10.1080/713755967>
- Ward, L. M. (1994). Supramodal and modality-specific mechanisms for stimulus-driven shifts of auditory and visual attention. *Canadian Journal of Experimental Psychology / Revue Canadienne de Psychologie Expérimentale*, 48(2), 242–259. <https://doi.org/10.1037/1196-1961.48.2.242>
- Wei, J., Carroll, R. J., Müller, U. U., Van Keilegom, I., & Chatterjee, N. (2013). Robust estimation for homoscedastic regression in the secondary analysis of case–control data. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, 75(1), 185–206. <https://doi.org/10.1111/j.1467-9868.2012.01052.x>
- Welsh, M. C. (2002). Developmental and clinical variations in executive functions. In *Developmental variations in learning: Applications to social, executive function, language, and reading skills* (pp. 139–185). Lawrence Erlbaum Associates Publishers. <https://doi.org/10.1076/chin.9.1.76.14495>

- White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48(4), 817–838.
<https://doi.org/10.2307/1912934>
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., ... Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686. <https://joss.theoj.org/papers/10.21105/joss.01686>
- Wickham, H., Pedersen, T., Seidel, D. (2023). *scales: Scale functions for visualization*. R package version 1.3.0, <https://github.com/r-lib/scales>, <https://scales.r-lib.org>.
- Wiebe, S. A., Espy, K. A., & Charak, D. (2008). Using confirmatory factor analysis to understand executive control in preschool children: I. Latent structure. *Developmental Psychology*, 44(2), 575–587. <https://doi.org/10.1037/0012-1649.44.2.575>
- Wiesen, J. P. (2006, October 27). *Benefits, drawbacks and pitfalls of z-score weighting*. Proceedings of the 30th Annual IPMAAC Conference, Las Vegas, NV.
- Wilkinson, H. R., Smid, C., Morris, S., Farran, E. K., Dumontheil, I., Mayer, S., Tolmie, A., Bell, D., Porayska-Pomsta, K., Holmes, W., Mareschal, D., & Thomas, M. S. C. (2020). Domain-specific inhibitory control training to improve children’s learning of counterintuitive concepts in mathematics and science. *Journal of Cognitive Enhancement*, 4(3), 296–314.
<https://doi.org/10.1007/s41465-019-00161-4>
- Wilson, C. J., Bowden, S. C., Byrne, L. K., Joshua, N. R., Marx, W., & Weiss, L. G. (2023). The cross-cultural generalizability of cognitive ability measures: A

- systematic literature review. *Intelligence*, 98, 101751.
<https://doi.org/10.1016/j.intell.2023.101751>
- Wong, P. T. P., & Cowden, R. G. (2022). Accelerating the science and practice of psychology beyond WEIRD biases: Enriching the landscape through Asian psychology. *Frontiers in Psychology*, 13.
<https://www.frontiersin.org/articles/10.3389/fpsyg.2022.1054519>
- Xia, Y., & Yang, Y. (2019). RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data: The story they tell depends on the estimation methods. *Behavior Research Methods*, 51(1), 409–428.
<https://doi.org/10.3758/s13428-018-1055-2>
- Xu, C., Ellefson, M. R., Ng, F. F.-Y., Wang, Q., & Hughes, C. (2020). An East–West contrast in executive function: Measurement invariance of computerized tasks in school-aged children and adolescents. *Journal of Experimental Child Psychology*, 199, 104929. <https://doi.org/10.1016/j.jecp.2020.104929>
- Xu, Y., Zhang, W., Zhang, H., Wang, L., Luo, Y., & Ni, G. (2022). Association between tennis training experience and executive function in children aged 8–12. *Frontiers in Human Neuroscience*, 16, 924809.
<https://doi.org/10.3389/fnhum.2022.924809>
- Yeung, N., & Monsell, S. (2003a). Switching between tasks of unequal familiarity: The role of stimulus-attribute and response-set selection. *Journal of Experimental Psychology: Human Perception and Performance*, 29(2), 455–469. <https://doi.org/10.1037/0096-1523.29.2.455>
- Yeung, N., & Monsell, S. (2003b). The effects of recent practice on task switching. *Journal of Experimental Psychology: Human Perception and Performance*, 29(5), 919–936. <https://doi.org/10.1037/0096-1523.29.5.919>
- Yiend, J., André, J., Smith, L., Chen, L. H., Touloupoulou, T., Chen, E., Sham, P., & Parkinson, B. (2019). Biased cognition in East Asian and Western cultures. *PLOS ONE*, 14(10), e0223358. <https://doi.org/10.1371/journal.pone.0223358>

- Young, A. G., & Shtulman, A. (2020). Children's cognitive reflection predicts conceptual understanding in science and mathematics. *Psychological Science*, 31(11), 1396–1408. <https://doi.org/10.1177/0956797620954449>
- Yuan, K.-H., & Bentler, P. M. (1998). Normal theory based test statistics in structural equation modeling. *British Journal of Mathematical and Statistical Psychology*, 51(2), 289–309. <https://doi.org/10.1111/j.2044-8317.1998.tb00682.x>
- Zaitchik, D., Iqbal, Y., & Carey, S. (2014a). The effect of executive function on biological reasoning in young children: An individual differences study. *Child Development*, 85(1), 160–175. <https://doi.org/10.1111/cdev.12145>
- Zaitchik, D., Iqbal, Y., & Carey, S. (2014b). The effect of executive function on biological reasoning in young children: An individual differences study. *Child Development*, 85(1), 160–175. <https://doi.org/10.1111/cdev.12145>
- Zaroff, C., D'Amato, R. C., & Bender, H. A. (2014). Understanding differences in cognition across the lifespan: Comparing eastern and western cultures. In J. M. Davis & R. C. D'Amato (Eds.), *Neuropsychology of Asians and Asian-Americans: Practical and Theoretical Considerations* (pp. 91–116). Springer. https://doi.org/10.1007/978-1-4614-8075-4_6
- Zeileis, A., & Hothorn, T. (2002). Diagnostic checking in regression relationships. *R News*, 2(3), 7–10. Retrieved from <https://CRAN.R-project.org/doc/Rnews/>.
- Zelazo, P. D., Blair, C. B., & Willoughby, M. T. (2016). *Executive function: Implications for education*. U. S. Department of Education (pp. 1–148). Retrieved from <https://ies.ed.gov/ncer/pubs/20172000/pdf/20172000.pdf>
- Zelazo, P. D., & Carlson, S. M. (2012). Hot and cool executive function in childhood and adolescence: Development and plasticity. *Child Development Perspectives*. <https://doi.org/10.1111/j.1750-8606.2012.00246.x>
- Zhang, X., Hu, B., Ren, L., Huo, S., & Wang, M. (2019). Young chinese children's academic skill development: Identifying child-, family-, and school-level

factors: Young chinese children's academic skill development. *New Directions for Child and Adolescent Development*.

<https://doi.org/10.1002/cad.20271>

Zhao, X., Chen, A., & West, R. (2010). The influence of working memory load on the Simon effect. *Psychonomic Bulletin & Review*, 17(5), 687–692.

<https://doi.org/10.3758/PBR.17.5.687>

Zhu, Y., Zhang, L., Leng, Y., Pang, R., & Wang, X. (2019). Event-related potential evidence for persistence of an intuitive misconception about electricity. *Mind, Brain, and Education*, 13(2), 80–91.<https://doi.org/10.1111/mbe.12188>

Zimmerman, M. E. (2011). Speed–accuracy tradeoff. In J. S. Kreutzer, J. DeLuca, & B. Caplan (Eds.), *Encyclopedia of clinical neuropsychology* (pp. 2344–2344). Springer. https://doi.org/10.1007/978-0-387-79948-3_1247

Zonneveld, A. K., Serpell, Z., Parr, T., & Ellefson, M. R. (2021). Executive function measurement in urban schools: Exploring the links between performance-based metrics and teacher rating. *Developmental Psychology*, 25(6), e13319. <https://doi.org/10.31234/osf.io/myr2s>

Appendices

Appendix A: Research Methodology

A.1 Theoretical Underpinnings of Quantitative Research

When selecting a research design in psychology and education, it is crucial to consider the epistemological approach (Haig, 2013). To explore the role of executive functions, general cognitive ability, and science education in explanatory coexistence, I chose the quantitative approach within the empiricist paradigm. The quantitative method emphasises obtaining knowledge through controlled observation (Newell, 2014), and it offers three key benefits. First, as quantitative research relies on numerical data, it minimises the possibility of researcher bias and subjectivity, thus enabling more trustworthy and consistent findings (Leung, 2015). Second, quantitative research allows the generalisation of findings to a broader population, providing insights into how relationships among variables extend to larger groups (Borgstede & Scholz, 2021). Third, while quantitative research cannot confirm the accuracy of a specific hypothesis, it can identify incorrect hypotheses, thus eliminating unreliable ones and offering a greater degree of certainty and reliability through deduction (Lee, 2020). Therefore, I used the quantitative method for my PhD project, as I aim to explore the quantified relationships between each variable across Study 1 and Study 2.

A.2 Within-Subject Research

With the aim of minimising variations in cognitive processing speed across participants, I selected a within-subject design (also known as a repeated measures design). By having each participant complete all measures, the within-subject design accounts for individual differences and effectively captures the variation between each trial of the experiment (Poling et al., 1995). There are three key considerations in within-subject design to reduce confounding variables (Taherdoost, 2016). In Table

A1, I summarise these three considerations, my decisions for addressing them, and the corresponding sections that explain how my research design meets these requirements.

Table A1

Within-Subject Research Design Requirements and Decision

Requirement for Within-Subject Design	Decision	Related Section for Detailed Explanation
Random selection of participants to ensure even representation of all demographics, minimal personal bias and sufficiency sample size to reduce sampling errors.	Randomly select participants in the United Kingdom and China	Table A14-I discussed how I recruit random data from the United Kingdom and China
Counteract sequence effects to arrange counterbalance trials to reduce order effects.	Using balanced trials and random order in task design	Appendix A.3-In task description, I describe how I address balanced trials and random order
Trail intervals to diminish effects from previous trials.	Adding fixation trials	Appendix A.3-In task description, I describe how I add fixation trials between experiment trials

A.3 Task Selection and Verification

In this section, I summarise the considerations made in selecting the materials for this PhD research. A detailed description of the specific tasks is provided in Chapter 4.

A.3.1 Measurement of Science Education

In previous research, as discussed in Chapter 2, the examination of science education was often illustrated by comparing science experts and novices. For example, comparisons were made between university students in science disciplines and those in the humanities (e.g., Foisy et al., 2015), between professors in science disciplines and those in the humanities (e.g., Shtulman & Harrington, 2016), or between high school students with high science competence and those with low competence (e.g., Allaire - Duquette et al., 2019). While such studies revealed group differences, they did not elucidate how the effects of scientific learning influence explanatory coexistence. Therefore, following the methodology of Ellefson et al. (2017), this PhD research asked participants to report their highest level of science education in the demographic questionnaire and assign scores corresponding to different educational levels (see Table A2). Since the naïve thinking tasks in this PhD research included only concepts from the biology and physics domains, I calculated the Science Education score by summing the scores for the highest levels of biology and physics education.

Table A2*Scores for Education Levels*

Education Level		Score
The United Kingdom	China	
None/prefer not to say	NA/不愿透露(prefer not to say)	0
	小学 (Primary school) ¹	4
Secondary / Middle School	中学 (Secondary / Middle School)	8
GCSE / O-Levels / Junior Cert	一些高中/中专、职高 (Some high school; Professional high school)	10
Some A-Levels / Leaving Cert		10
	高中 ² (Completed high school)	12
Completed A-Levels	一些大专/高职/高专课程 (Some professional college)	14
Some undergraduate	部分本科课程 (Some undergraduate)	14
	大专/高职/高专学位 ³ (Completed professional college)	15
Completed bachelor's degree	本科 (Completed bachelor's degree)	16
Some masters	部分硕士课程 (Some masters)	18
Completed masters	硕士 (Completed masters)	20
Some doctoral	部分博士课程 (Some doctoral)	22
Completed doctoral	博士 (Completed doctoral)	24

Notes. ¹ Due to educational reforms in China, some older participants may not have received education beyond the elementary level. ² In the Chinese educational system, completing high school requires three years of education; therefore, it is assigned two points higher than the A-levels and equivalent exams in the United Kingdom. ³ In China, professional colleges typically offer shorter periods of study compared to regular universities; therefore, they are assigned one point less than completing a bachelor's degree.

A.3.2 Selection of Computerised Tasks

The two most commonly used methods to assess executive functions are questionnaire-based measurements and performance-based measurements (Toplak et al., 2013). Although questionnaire-based measurements offer enhanced ecological validity and are suitable for clinical application, they are susceptible to task-impurity issues (Snyder et al., 2011). Therefore, for the current study, I decided to use performance-based tasks, which provide more precise and standardised measurements of executive functions (Burgess & Simons, 2005; Miyake et al., 2000). In performance-based tasks, all participants completed the same measurements, and their performance was evaluated using more reliable and objective metrics, including accuracy, response time, and efficiency (Zonneveld et al., 2021).

Inhibition. Considering the distinct roles that interference inhibition and behavioural inhibition play in explanatory coexistence (Brookman-Byrne et al., 2018; Ellefson et al., 2024), I selected separate tasks to evaluate these two forms of inhibition. For interference inhibition, the Flanker task focuses more on inhibition itself rather than incorporating other executive function components (e.g., switching; see Table A5), in contrast to other commonly used tasks for measuring interference inhibition (see Table A3). Although the revised version of the Flanker task, the Fish Flanker task, was originally designed to be child-friendly, Rueda (2004) found that the same mental processes were involved when this version was administered to adults. Furthermore, the Fish Flanker task has been used in other adult studies (e.g., Brydges et al., 2012; Nozari et al., 2016). However, unlike the Fish Flanker task used by Ellefson et al. (2024), the current version adjusted certain aspects of the task to avoid a ceiling effect on accuracy, particularly with regard to the difficulty level for adult participants. I included different forms of incongruent trials to increase the task's difficulty (for more details, see Appendix B). In later sections, I refer to the Fish Flanker task in this research simply as the Flanker task.

Table A3*Task Selection for Interference Inhibition*

Tasks	Description	Reason for selection or exclusion
Stroop Task (Stroop, 1992)	Participants identify the colour (e.g., red, blue, yellow) of the given colour word (e.g., red, blue, yellow) shown on the screen. The colours of the given word are sometimes incongruent.	The Stroop task can also be used to evaluate task-switching (e.g., Hutchison et al., 2010; Ward et al., 2001)
Simon Task (Simon & Rudell, 1967)	Participants need to respond to visual stimuli by making a rightward response to one stimulus (e.g., a circle) and a leftward response to another (e.g., a square).	The Simon task can also be used to measure short-term memory/working memory (e.g., Diamond & Wright, 2014; Zhao et al., 2010)
Flanker task; Fish Flanker task (Eriksen & Eriksen, 1974; Ellefson et al., 2024; Rueda, 2004)	Participants need to respond to a directional response (usually left or right) to a central target stimulus.	The Flanker task has less interference compared with other two tasks

For behavioural inhibition, the most common paradigms are the Go/No-Go task and the Stop Signal task (Diamond, 2013; Nikitenko et al., 2019; Raud et al., 2020). However, although both tasks assess the ability to inhibit intentional behaviour, there is ongoing debate regarding whether they activate different inhibition mechanisms (Nikitenko et al., 2019) (see Table A4 for a detailed comparison). In the context of suppressing naïve thinking in favour of later-learned scientific thinking in explanatory coexistence, inhibiting the tendency to select naïve responses is crucial. Therefore, the inhibition mechanism of the Stop Signal task is more appropriate for this research.

Table A4

Task Selection for Behavioural Inhibition

Tasks	Task Description	Reason for Selection or Exclusion
Go/No-Go Task (Murphy et al., 1999)	Part of the <i>go</i> section of the trial is replaced by a <i>no-go</i> section, wherein participants are requested not to respond to these trials. <i>No-go</i> sections instruct the participant not to respond.	The Go/No-Go task captures action restraint, which denotes making a decision regarding whether or not to respond (Raud et al., 2020).
Stop Signal Task (Logan, 1994)	The <i>go</i> stimulus is always shown first, though may then be followed by a <i>stop</i> stimulus after a short stop-signal delay. Stop stimuli instruct the participant not to respond.	The default decision is always to respond, so the successful reaction relies on action cancellation, which refers to the suppression of an already initiated response (Schachar et al., 2007). Thus, this task is more appropriate for assessing the inhibition of selecting naïve thinking.

Switching. With regard to switching tasks, participants are typically instructed to switch between different tasks at least once to facilitate the performance of specified mental operations or actions in response to stimulus input (Miyake et al., 2000). Although there is no consensus on the precise definition of what constitutes a task, Diamond (2013) concluded that a switching task generally involves two tasks and requires participants to respond to one feature of each task. The stimuli in these tasks are often bivalent, meaning they contain a feature relevant to both tasks, with the correct response for one task being incorrect for the other. Participants generally take longer to respond in mixed-task blocks that require switching between trials, compared to single-task blocks that require only one task (Hübner et al., 2001; Koch et al., 2010; Rubin & Meiran, 2005; Steinhauser & Hübner, 2007). However, tasks designed to assess switching often interfere with inhibition tasks, leading to a confound in assessing pure switching. Therefore, after comparing various switching tasks (see Table A5), I selected the Figure Matching task, which is relatively pure for assessing cognitive switching and has been validated in cross-cultural research (Ellefson, 2017).

Table A5*Task Selection for Switching Tasks*

Task	Task Description	Reason for selection or exclusion
Stroop Task (Hübner et al., 2001; Koch et al., 2005; Rubin & Meiran, 2005; Steinhauser & Hübner, 2007)	Participants need to indicate the colour of the word on the screen. The word and the colour are sometimes incongruent.	Most commonly used for inhibition (depends on the specific items used) (e.g., Rogers & Monsell, 1995).
Wisconsin Card Sort Test (Grant & Berg, 1948)	Participants need to sort the cards based on rules, such as: number, colour, or shape. The rule will be changed after the presentation of several trials.	Results of this task might be far from reaching the desirable standard of reliability (e.g., Kopp et al., 2021).
Hearts & Flowers Task (D. Davidson, 2006)	A red heart or a red flower appeared on either the left or right side of the screen. Participants need to press the button which corresponds to the side on which heart appears, though they must press the opposite button when the flower appears.	Also used to assess inhibition (like the Simon task), mostly used on measuring children (e.g., Wright & Diamond, 2014).
Figure Matching Task (Ellefson et al., 2017; 2024; Wang et al, 2021)	Participants need to correspond the small pattern on the left or right corner to the big pattern in the middle based on certain rules (shape or colour)	Less interference with inhibition tasks, it is used on both children and adults across different countries (Ellefson et al., 2017; 2024).

Working Memory. Working memory tasks often assess the amount of information an individual can hold in mind and react to at once (Bunting, 2006). Currently, three types of working memory tasks are widely used in research: simple span tasks (Mathy et al., 2018), complex span tasks (St Clair-Thompson & Sykes, 2010), and running span tasks (Gray et al., 2003; Jaeggi et al., 2008). As this research aims to assess working memory specifically (see Table A6), the Simple Span task has been selected. Various versions of the Corsi Block task exist, each varying in block placement, span length, and the end-task process. Taking this into account, I referred to both the original Corsi Block task and the Pattern Challenge task (a revised version of the Corsi Block task) from the Thinking Games, which has been used in both child and adult participants from the United Kingdom and China (Ellefson et al., 2017; 2021; Xu et al., 2020), to create the version used in this research. In later sections, I refer to this revised version simply as the Corsi Block task for ease of reference.

Table A6*Task Selection for Working Memory Tasks*

Tasks	Task Description	Reason for selection or exclusion
Simple Span task: Digit Span Task (Hilbert et al., 2015) Corsi Block Task (Berch et al., 1999)	Comprises sequences of verbal items (letters, words or numbers) or visuospatial items (locations on a grid). Once the final item has been presented, participants are required to recall the list in a forward or reverse sequence order.	Backward span of these tasks is deemed more appropriate as working-memory measurements, as the forward span tasks simply repeat remembered items (more similar to short-term memory) in the same order. Both tasks does not need much mental manipulation (St Clair-Thompson & Sykes, 2010), Corsi Block Task also exclude the confounding of math ability.
Complex Span task: Self-Ordered Pointing Task (Petrides & Milner, 1982) Counting Span Task (Case et al., 1982)	The procedure is akin to the Simple Span task, with the primary exception to the process being the introduction of a simple processing task, such as a mathematical operation or an evaluation of symmetry, which must be performed between the presentation of each item.	Complex span tasks involve the processing of more subcomponents of executive functions, rather than simply the retention and manipulation of information in the mind. This, it is debatable as to whether it measures working-memory alone (P. L. Ackerman et al., 2005; Unsworth & Engle, 2007).
Running Span tasks: N-Back Task (Kirchner, 1958) Running Memory Span Task (Pollack et al., 1959)	Participants need to heed the sequential list and recall the last three or four items (letters, spatial locations). For the N-Back task, participants are required to respond in defined ways, such as by pressing keys, whenever current items match those presented n ago. Contrastingly, in the running span task, subjects need to wait until all items are presented before recalling the final n items.	Although these two tasks are widely accepted as valid working-memory tasks with robust links to complex span tasks, there is no consensus as to which cognitive mechanisms are involved in both the Running Span task and the N-Back task (Shipstead et al., 2010).

General Cognitive Ability. Evaluating general cognitive ability typically involves the use of standardised tests. Some standardised tests are known to offer reliable and thoroughly validated abstract reasoning components, and are widely used in clinical, educational, occupational, and research settings (see Table A7). However, these tests are not freely accessible, and copyright restrictions often prevent the transformation of traditional pen-and-paper tests into digital formats. Chierchia et al. (2019) summarised that, although some online tests, such as the 20-item Hagen Matrices Test, are available at no cost, these free tests typically include a limited range of items. Moreover, the predominantly analogue method of administration means that the existing normative data for these tests usually only include accuracy figures, excluding response times. This presents a significant practical and theoretical limitation due to the established connection between abstract reasoning and processing speed, as well as the potential for speed–accuracy trade-offs. Consequently, I selected the Matrix Reasoning task, newly validated by Chierchia et al. (2019), which is an 80-item online assessment that evaluates both accuracy and response time, making it suitable for adult participants.

Table A7*Task Selection for General Cognitive Ability Task*

Task	Task Description	Reason for selection or exclusion
Wechsler Abbreviated Scale of Intelligence (Brookman-Byrne et al., 2018)	Consist of various subtests that measure different cognitive domains, yielding a Full-Scale IQ score along with scores in specific areas.	Not freely accessible and lack valid online assessing form. Popular tests, some participants might access these tests before.
Raven's Progressive MatricMRes (Vosniadou. et al, 2018)	A non-verbal group test typically used to measure abstract reasoning. It is considered a non-verbal estimate of fluid ability or "g."	Not freely accessible and lack valid online assessing form. Popular tests, some participants might access these tests before.
Matrix Reasoning Item Bank (Chierchia et al., 2019)	Raven like intelligence task. Contains 80 matrix reasoning trials.	Newly developed online assessment. Free-to-use verified assessment. Suitable for wide age range.

Explanatory Coexistence. Unlike cognitive abilities, there is no standard assessment for explanatory coexistence. The decision to use tasks assessing naïve thinking in relation to explanatory coexistence was based on the most commonly used paradigms suitable for adults. Table A8 outlines the measurement methods used in previous research among adolescents and adults. All these tasks include congruent trials (where the stimuli's naïve thinking aligns with science thinking) and incongruent trials (where the stimuli's naïve thinking does not align with science thinking). Additionally, in my previous literature review, I discussed how naïve thinking tasks with different modalities might involve different cognitive abilities. Overall, the biological concept of animism (picture-based task) and the physics-related concept of buoyancy (picture-based task) were most commonly used in previous tasks that involve explanatory coexistence concepts. The Science Statement task (text-based task) was also commonly used and includes a broader range of explanatory coexistence concepts. Therefore, I selected the Animism task, the Buoyancy task, and the Science Statement task for my PhD research to measure explanatory coexistence.

However, I made a few minor revisions to the Animism and Science Statement tasks. First, I reduced the number of trials in the living group of the Animism task to balance the number of trials across all groups (for a detailed analysis, see Appendix B). Second, I selected only 64 sentences for the Science Statement task (see task description in Chapter 6). Third, as found in my master's research, the Animism and Science Statement tasks showed no differences based on left- or right-hand responses (for detailed analysis, see Appendix B). Therefore, I used fixed buttons in my PhD research.

Table A8*Summary of Measurements of Naïve Scientific Concepts in Behaviour Studies*

Measurements (Modality)	Research	Age group	Domain
Buoyancy tasks (Picture)	Brault Foisy et al., 2021	Children; Adolescent Adults	Physics concept about buoyancy
Buoyancy task (Picture); Animism task (Picture)	Ahr et al. (2016)	Adolescents; Adolescents	Physics concept about the buoyancy and biological concepts of Animism
Science & Math Statement task (Picture & Text)	Brookman-Byrne et al. (2018)	Adolescents	Physics concepts related to gravity; mathematics
Buoyancy task (Picture)	Potvin et al. (2015)	Adolescents	Physics concept about buoyancy
Solid and Liquid Classification Test (Picture)	Vosniadou, (2019)	Adolescents	Physics concept about solids and liquids
Animism task (Picture)	Babai et al. (2010)	Adolescents	Biological concept of animism
Scientific Statement task (Text)	Shtulman & Valcarcel (2012)	Adults	General physics and biology knowledge
Scientific Statement task (Text); Animism task (Picture)	Ellefson et al., 2024	Adults	General physics and biology knowledge; Biological concept of animism

A.4 Chinese Translation

In my PhD project, all instructional pages within tasks, as well as the 64 sentence statements in the Science Statement task, needed to be translated from English into Chinese. Regarding language translation in psychology experiments, van de Vijver and Hambleton (1996) identified three types of bias: 1) construct bias, resulting from non-equivalence of constructs across cultural groups; 2) method bias, arising from issues in instrument administration; and 3) item bias, stemming from inadequate translations, such as incorrect word choice. To avoid construct bias, all scientific statements relating to basic knowledge were designed to be easily understood across countries. I also ensured that the tasks were identical for both English and Chinese versions, maintaining the same research procedure to avoid method bias. Lastly, to prevent item bias caused by inaccurate translations, I followed a three-step process for translating each English sentence: first, I translated the text into Chinese; second, my translations were reviewed by two native Chinese-speaking members of my research lab, who are familiar with the research content; and third, I conducted a back-translation, in which a bilingual translator translated the Chinese version back into English (Brislin, 1970). Through these steps, I ensured that the Chinese materials accurately conveyed the meaning and intent of the original text in a way that was accessible and understandable to the target participants.

A.5 Online Administration

Both my master's and PhD projects were conducted during the COVID-19 pandemic. Given the feasibility of data collection, I transitioned both projects from offline to online, using PsychoPy© for computerised experimental tasks and Qualtrics© for demographic information. Online data collection reduces geographical constraints, enabling researchers to reach a more diverse sample and conduct experiments more efficiently (Kraut et al., 2004). Although online data collection may not provide a controlled lab environment for participants and could introduce

extraneous variables that affect the results (Reips, 2002), the tasks I employed have demonstrated ecological validity across several studies (Ellefson et al., 2017, 2024; Vosniadou et al., 2018). Nevertheless, two major issues with online data collection needed to be addressed.

First, during my PhD project, PsychoPy© (web: pavlovio) was still a relatively new tool and lacked a questionnaire function. As a result, I used both Qualtrics and PsychoPy to collect behavioural data and demographic information. Due to PsychoPy's inability to display a Qualtrics link upon completion of the computerised tasks, participants were first directed to complete demographic information on Qualtrics. Afterward, they received a 5-digit participant ID, which they used to download the PsychoPy tasks. Participants were instructed to enter this ID into the PsychoPy tasks at the beginning. However, some participants made errors when entering their ID numbers, which resulted in unmatched data between PsychoPy and Qualtrics. Additionally, due to the structure of the PsychoPy website, data from participants who did not complete all experimental tasks could not be saved. As a result, only participants who completed all PsychoPy tasks were included in the study.

Second, the accessibility of the PsychoPy website in China was not as reliable as in the United Kingdom, likely due to China's Great Firewall. To address this, I tested various Chinese web browsers before the study and recommended browsers that provided easier access to the links (e.g., Web 360 in China) to our participants. However, even in the United Kingdom, the ongoing development of PsychoPy's online version led to compatibility issues with several web browsers, which interfered with the smooth conduct of the research. Furthermore, during the experiment, some participants were unable to proceed with the fourth task, the Corsi Block task. After consulting with the Pavlovio platform, I also recommended that UK participants use the Chrome browser for better compatibility.

A.6 Behaviour Research Measurement

In assessing cognitive abilities, multiple metrics are often employed to capture the diverse aspects of complex cognitive processes (Maris & Maas, 2012). Accuracy is one of the most commonly used metrics to quantify individual variability in cognitive tasks, as it measures whether a response is correct or incorrect (Christ et al., 2018). In younger participants (e.g., children) or in tasks presented in questionnaire form (such as some standardised tests), accuracy is often the primary measurement index (e.g., Grenell & Carlson, 2021; Zaitchik et al., 2014). In executive function research, accuracy is widely accepted as it directly assesses individuals' ability to perform tasks in accordance with task requirements and provides insight into the efficacy of their decision-making processes (Miyake et al., 2000). In explanatory coexistence, accuracy reflects participants' understanding of various explanatory coexistence questions presented in pictorial or textual stimuli. Congruent trials, where the stimuli align between naïve and science thinking, typically yield higher accuracy than incongruent trials, where the stimuli conflict between naïve and science thinking (Babai et al., 2010; Potvin et al., 2015a; Shtulman & Valcarcel, 2012).

In addition to accuracy, response time is another key metric that provides insight into cognitive processing (De Boeck & Jeon, 2019). Response time is defined as the interval between the presentation of a stimulus (visual, audio, etc.) and the execution of a response (e.g., pressing a button), and it reflects the complexity of reasoning required to resolve a task (Denes & Pizzamiglio, 2020). In executive function and explanatory coexistence research, response time is often specifically analysed for correct trials (e.g., Ellefson et al., 2017; 2024; Potvin et al., 2015). As response time is sensitive to individual differences in cognitive efficiency and may be influenced by participant attention, using response time for correct trials can reveal subtle distinctions in cognitive processing that may not be apparent through accuracy measures alone (Bélanger et al., 2024; De Boeck & Jeon, 2019). For simplicity, I collectively refer to response time for correct trials as "response time." In explanatory coexistence, response time also reflects how participants reason about different trial types. Congruent trials generally require less response time than incongruent trials, suggesting that congruent trials demand less cognitive effort to reach an answer (Brookman-Byrne et al., 2018; Potvin et al., 2015; Vosniadou, 2019).

Both response time and accuracy provide valuable insights into the underlying mechanisms of cognitive processes and highlight the intricate reasoning involved in task-solving (Bruyer & Brysbaert, 2011; De Boeck & Jeon, 2019; Denes & Pizzamiglio, 2020). These metrics are extensively used in studies related to executive functions and explanatory coexistence (Brookman-Byrne et al., 2018; Grenell & Carlson, 2021; Vosniadou et al., 2018). The well-established psychological understanding that response accuracy varies with the time allocated for responding has been explored for over a century (Heitz, 2014). Although the exact relationship is complex, the speed-accuracy trade-off suggests that participants often allocate more time to their responses to enhance accuracy, particularly in adults (Bruyer & Brysbaert, 2011; Liu & Watanabe, 2012; Zimmerman, 2011). Depending on their response patterns, participants may adopt either a liberal (rapid but error-prone) or conservative (emphasising accuracy over speed) approach. With their baseline information, participants may reach different decision thresholds, making more or less informed decisions (Lubczyk et al., 2022).

Building on the considerations of accuracy and response time, I have chosen efficiency as the third metric, following the methodology outlined by Ellefson et al. (2017). Efficiency accounts for the speed-accuracy trade-off in participants' responses, providing a more comprehensive interpretation of results beyond simply considering accuracy and response time separately. For example, as numerous studies have shown, executive function tasks often exhibit a ceiling effect for accuracy in adults, making it difficult to fully assess participants' performance based solely on accuracy (Bruyer & Brysbaert, 2011; Miyake et al., 2000; Rogers & Monsell, 1995). Moreover, some participants may deliberately slow their responses to improve accuracy, so using efficiency to represent accuracy within time constraints offers a more nuanced view of participants' response patterns (Suchotzki et al., 2021).

Overall, using efficiency offers several advantages. First, it provides a more normalised distribution that accounts for the ceiling effect on accuracy (Ellefson et al., 2020). Second, efficiency allows for a better understanding of the interaction between accuracy and processing speed, offering insights into how participants respond to different trial types and make optimal decisions under varying conditions (Liesefeld & Janczyk, 2019). Third, efficiency can differentiate individuals with similar overall performance levels but distinct cognitive processing styles, providing a more comprehensive view of an individual's ability to perform tasks effectively and

adaptively (Salthouse, 2005). Given these considerations, I used efficiency as the primary metric in the analysis reported in this research. The formula for calculating these metrics is provided in Table A9 (Ellefson et al., 2017; 2024).

Table A9

Formula Summary of Accuracy, Response Time and Efficiency

Parameter	Formula
Accuracy	Correct Trials/Total Trials
Response Time	Mean Response Time for Correct Trials
Efficiency	Accuracy/Response Time

A.7 Ethics Consideration

The framework and details of this PhD research were developed in accordance with the ethical standards set by the British Educational Research Association (2018). No deception was used in this study, and all responses from participants were kept anonymous. Data were handled with strict confidentiality and used solely for the purposes of this research. Participation was voluntary, and all individuals involved were informed of their right to discontinue participation at any point during the research process, without the obligation to explain their decision. The research protocol was reviewed by the ethics committee of the University of Cambridge, Faculty of Education.

A.8 Open Science and Pre-registration

Adopting the principles of open science promotes early-stage transparency and fosters collaboration in creating and sharing knowledge (Lobo, 2019). Both studies in this PhD research were pre-registered on the Open Science Framework (OSF). Table A10 summarises the relevant registrations and updates for the two studies. Additionally, some analysis codes are available through the following OSF links.

Table A10*Information of the Pre-registration*

Study	Time	Stage	Model	OSF links
Study 1	2022 April	Collection	Figure A1	https://osf.io/3a7yf
Study 1 ^{updated}	2023 Nov	Analysis	Figures A2, A3, A4	
Study 2	2023 Oct	Analysis	Figure A4	https://osf.io/bqw3k

The first pre-registration (for Study 1) was related to data from the United Kingdom. As mentioned in Chapter 1, my PhD research was inspired by my master's research. Initially, my plan was to explore only the relationship between different executive functions and explanatory coexistence. Therefore, when I pre-registered Study 1 in 2022 (see Figure A1), I did not consider including general cognitive ability and science education. However, in later discussions with my supervisors, Prof. Michelle Ellefson and Prof. Aidan Feeney, we realised that exploring general cognitive ability and science education might provide more valuable insights. Additionally, due to collaboration with Konstantinos Tsigaridis in our lab (see OSF: <https://osf.io/mwqk2/>), I had already collected data related to general cognitive ability and science education. As a result, I later incorporated these two variables into the research questions and updated the pre-registration for Study 1 in 2023.

With the update for Study 1, depending on whether it was possible to form latent variables and whether the results of the confirmatory factor analysis (CFA) indicated a good model fit, I proposed three models. If my data could form latent variables for executive functions and explanatory coexistence, I would use Models 1 and 2. In Model 1, I treated science education as a direct predictor of explanatory coexistence, while also considering it as a covariate of executive functions and general cognitive ability. In this model, executive functions served as the mediating variable in the link between general cognitive ability and explanatory coexistence (see Figure A2). In Model 2, I considered science education as a moderator, exploring whether it moderates the link between executive functions and explanatory coexistence, as well as the impact of general cognitive ability on explanatory

coexistence. This model also included a mediation relationship, where executive functions mediated the link between general cognitive ability and explanatory coexistence (see Figure A3).

If my data could not form latent variables for executive functions and explanatory coexistence, the CFA model might show poor fit. In this case, I planned to use path analysis to assess the causal relationships between all variables. In Model 3, I used science education, general cognitive ability (the Matrix Reasoning task), and all executive function tasks (the Figure Matching task, the Stop Signal task, the Flanker task, and the Corsi Block task) as predictors, with all naïve thinking tasks (the Animism task, the Buoyancy task, and the Science Statement task) as outcomes (see Figure A4).

The results of the CFA (see Chapters 5 and 6) indicated collinearity issues between the latent variables of executive functions and explanatory coexistence, and some models showed poor fit. Consequently, I decided to use path analysis as proposed in Model 3 for the analysis.

Additionally, for Study 2, I conducted path analysis based on the results from Study 1, which was pre-registered simultaneously. Path analysis was also performed on the data from China.

Figure A1

Theoretical Structure Equation Model Pre-registration for Study 1

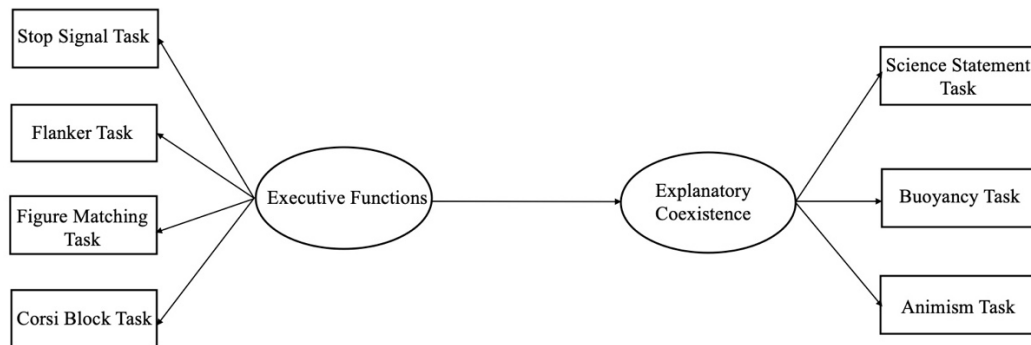


Figure A2

Theoretical Structure Equation Model 1 in Updated Pre-registration for Study 1

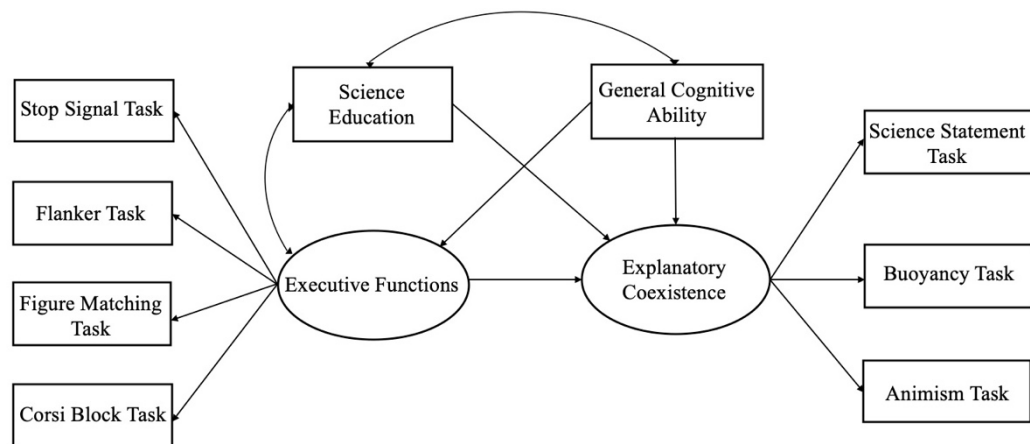


Figure A3

Theoretical Structure Equation Model 2 in Updated Pre-registration for Study 1

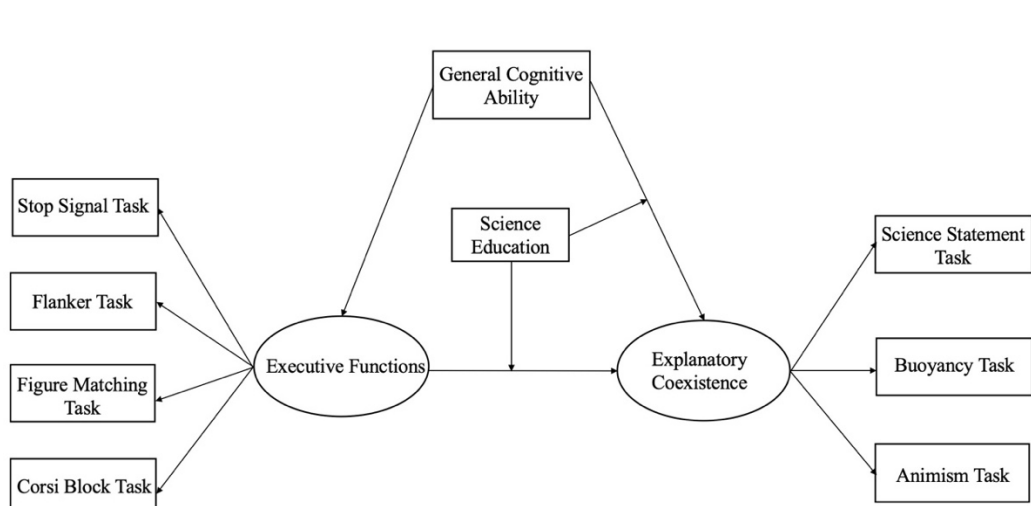
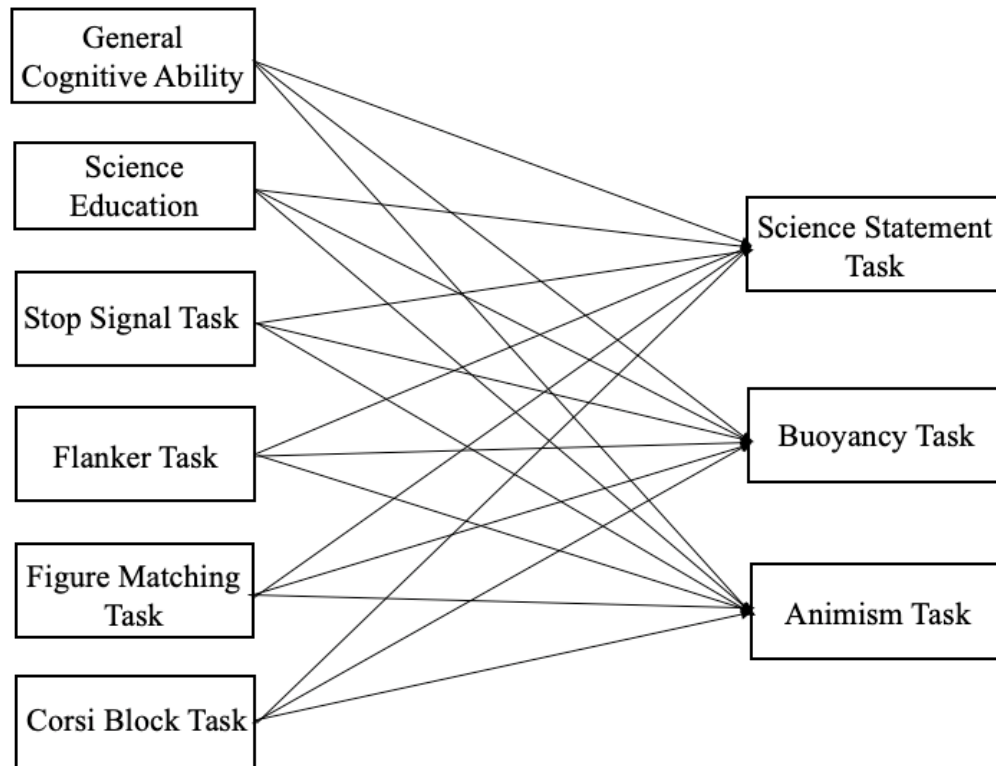


Figure A4

Theoretical Model 3 in Updated Pre-registration for Study 1



Notes. This path model is also used in Study 2 (China data).

A.9 Rationale and Justification of Data Analysis

Before conducting the data analysis, I first justified the data processing and analysis approaches based on the characteristics of the data and the research questions. In this section, I address issues that must be considered in the preliminary analysis, providing the basis for these considerations and the rationale for addressing them. I then introduce the key analytical methods and steps that were employed in the main analysis. This section includes conceptual explanations and logical reasoning, while the specific implementation steps and R packages used in the research are detailed in Chapter 4.

A.9.1 Preliminary analysis

In this section, I outline the assessments necessary before conducting the main analysis. These include processes for checking normality, identifying outliers, and handling missing data for each variable. I also explain the rationale for performing age adjustments and z-score conversions, as well as the steps for evaluating the naïve thinking tasks. Additionally, I address the examination of multicollinearity and homoscedasticity among the relevant data. Finally, I describe the logic and steps involved in using confirmatory factor analysis (CFA) to determine whether latent variables can be formed for subsequent analysis.

Normality and Outliers. My dataset included one categorical variable (nationality: the United Kingdom and China in Study 2) and nine continuous variables (science education and all computerised tasks). For continuous variables, normality testing is a crucial step in determining the appropriate measures of central tendency and statistical methods for various data analyses, such as correlation, regression, and path analysis (Ghasemi & Zahediasl, 2012). Normally distributed data allows the use of parametric tests; however, when the data does not follow a normal distribution, nonparametric tests or data adjustment methods must be considered (Field, 2018).

The Shapiro–Wilk test is commonly used for assessing normality, particularly with large sample sizes. A significant result from the Shapiro–Wilk test indicates a non-normal distribution (Shapiro & Wilk, 1965). In cases of non-normality, several methods can be applied to address the issue (see methods in Table A11). After considering the rationale and justification, I decided to use the Robust Maximum Likelihood method in the subsequent analysis to mitigate issues related to non-normal distribution.

Table A11*Approaches to Managing Non-Normal Data*

Methods	Rational	Justification	Decision
Transform original data (Sainani, 2012)	Square root (Lee, 2020) or Logarithm transformation (Feng et al., 2014)	Data transformation might mitigate skewness and kurtosis; however, if the non-normal distribution is moderate or minimal, data transformations might have limited impact and might not guarantee a multivariate normal distribution (Lee, 2020).	Consider data transformation together with other methods.
Remove outlier (Field, 2018)	Outliers influence the deviation of data set (Leys et al., 2013).	Although removing outliers might decrease skewness and kurtosis to preserve the linearity assumption, it could result in the loss of data points and reduce the predictive power of the model (Osborne & Overbay, 2019).	Keep outliers in my data analysis.
Using more robust data analysis methods (Park & Cho, 2003)	Using robust data analysis method (e.g., Robust Maximum Likelihood), adjusts the standard errors of the variables to account for non-normality and non-independence variables (Bandalos, 2014; Kline, 2016).	Using robust data analysis method could address the issues related to non-normal distribution in more complex models (Bandalos, 2014).	Use robust maximum likelihood in complex models such as confirmatory factor analysis and path analysis.

Missing Data. In my dataset, some data for the science education variable were missing due to matching issues between Qualtrics and PsychoPy data (as discussed in Appendix A.5). Missing data can lead to biased parameter estimates and standard errors (Little & Rubin, 2002), and it may also reduce statistical power, increasing the likelihood of Type I or Type II errors (Kang, 2013). Therefore, it was essential to apply appropriate techniques to address the missing data issue, ensuring it did not affect the conclusions drawn from the main analyses (Schafer & Graham, 2002). After evaluating various methods for handling missing data (see Table A12), I employed listwise deletion in correlation analysis and full information maximum likelihood in more complex analyses (confirmatory analysis and path analysis).

Table A12*Approaches to Managing Missing Data*

Methods	Rational	Justification	Decision
Listwise Deletion (Field, 2018)	In listwise deletion, cases with missing data in variables intended for inclusion in the analysis model are excluded (Newman, 2014).	Listwise deletion addresses skewness and kurtosis across all variables, however, if non-normality is moderate or minimal, these transformations have a limited impact and might not guarantee a multivariate normal distribution (Dong & Peng, 2013).	Use listwise deletion in correlation analysis (Ellefson et al., 2024)
Multiple Imputation (Field, 2018)	Multiple imputation generates plausible values for missing data based on observed variable distributions and relationships within the dataset (Sterne et al., 2009).	Multiple imputation reduces the bias resulting from data absence and allows for the use of all available data. However, repeated imputation enables the acquisition of accurate estimates of standard errors (Allison, 2012).	Not use multiple imputation.
Full Information Maximum Likelihood (Carter, 2006)	Full Information Maximum Likelihood operates on covariance structures within data subgroups and is typically used in more complex models (Allison, 2003)	Full Information maximum likelihood delivers unbiased estimates of model parameters for data that are missing completely at random and missing at random. Additionally, it utilises all available information in the dataset, thereby preserving the statistical power of the analysis (Enders, 2008).	Used full information maximum likelihood in confirmatory factor analysis and path analysis (Zonneveld et al., 2021).

Age Adjustment Testing and Z Score conversion. In Chapter 2, I reviewed the development of executive functions and general cognitive ability across the lifespan. These cognitive skills improve throughout childhood and adolescence, peak in early to middle adulthood, and then decline in late adulthood (e.g., Anderson, 2002; Best & Miller, 2010; de Oliveira et al., 2012). Although participants in my study were adults, their ages ranged from early to late adulthood (see Section 4.1). Therefore, it was necessary to consider the influence of age in my analysis to avoid potential bias associated with age differences (Reimers & Maylor, 2005). Given the complexity of the data analyses in my PhD research, including confirmatory factor analysis and path analysis, I performed an age adjustment prior to conducting the analyses to control for age-related bias (Lawson & Farah, 2017; Wiebe et al., 2008).

Following the approach used by Zonneveld et al. (2021), I controlled for age by removing the remaining variance through a hierarchical regression strategy (Lawson & Farah, 2017; Wiebe et al., 2008). In this approach, I examined the linear, parabolic, and curvilinear relationships between age and the outcomes of all my computerised tasks (Cohen et al., 2003). Significant results from these analyses indicated varying patterns of association between age and the measured variables.

After the age adjustment procedure, I converted all variables to z-scores. This step allows for the calculation of the likelihood of a specific score occurring within a normal distribution and facilitates the comparison of scores from distinct normal distributions (Wiesen, 2006).

Verifying Naïve Thinking Tasks. Building on the differences between congruent and incongruent trials discussed in the literature review, I conducted preliminary analyses before the main analysis to determine whether the three naïve thinking tasks in this PhD project were consistent with previous studies.

Previous research has shown that participants' accuracy in congruent trials was higher than in incongruent trials, and their response time in congruent trials was shorter than in incongruent trials (Babai et al., 2010; Potvin et al., 2015; Shtulman & Valcarcel, 2012). Additionally, since both congruent and incongruent trials belong to the same task with similar concepts, responses to these two trial types are likely to be correlated. Therefore, before proceeding with the main analysis, I performed preliminary analyses to assess whether the three naïve thinking tasks in this PhD project aligned with previous studies.

To evaluate the differences between congruent and incongruent trials, I used a mixed-effects model to estimate both between- and within-subject variability for accuracy and response time, and a t-test to examine the differences in efficiency scores. I also conducted correlation analysis to assess the relationships between congruent and incongruent trials across accuracy, response time, and efficiency. Table A13 provides further details on the rationale behind my decisions for checking the three naïve thinking tasks.

Table A13*Decision Making for Checking Naïve Thinking Tasks*

Measure	Method	Rationale	Justification
Compare individual response of accuracy and response time between congruent and incongruent trials	Mixed effects models	Mixed effects model includes random effects for the individual response and allows the modelling of both between-group and within-group variance (McNeish & Kelley, 2019).	Accuracy in congruent trials should be higher than that in incongruent trials, and response time in congruent trials should be shorter than that in incongruent trials (Babai et al., 2010; Potvin et al., 2015a; Shtulman & Valcarcel, 2012).
Compare efficiency score between congruent and incongruent trials	T-test	Efficiency was calculated based on individual's accuracy and response time, so that could only use fixed effect model for comparison.	Based on the differences between congruent and incongruent trials, efficiency in congruent trials should be higher than that in incongruent trials (Ellefson et al., 2024).
Check the relationship between congruent and incongruent trails	Correlation	Double check the linear relationship between congruent and incongruent trials.	Congruent and incongruent trials should be correlated (Vosniadou et al., 2018).

Multicollinearity and Homoscedasticity.

Multicollinearity and homoscedasticity are critical assumptions in regression analyses (Field, 2018). Multicollinearity refers to the presence of co-linearity among independent variables (Vatcheva et al., 2016), which can result in an unstable regression model that is sensitive to minor variations in predictors (Vatcheva et al., 2016). To ensure valid parameter estimation and accurate model interpretation, I assessed multicollinearity in the preliminary analysis to avoid inflated standard errors and model overfitting (O'Brien, 2007). The variance inflation factor (VIF) and Tolerance are commonly used measures for assessing multicollinearity (Marcoulides & Raykov, 2019). A VIF exceeding 5 or 10, or a Tolerance below 0.1 or 0.2, may indicate high multicollinearity and a problematic model, respectively (Field, 2018; Marcoulides & Raykov, 2019; O'Brien, 2007).

Homoscedasticity, which refers to the consistency of error variance across independent variable values, is essential for ensuring the reliability of the regression line and the precision of standard error estimates (Breusch & Pagan, 1979). A violation of homoscedasticity (also known as heteroscedasticity) can lead to unreliable results due to inefficient parameter estimations and potentially incorrect standard errors (White, 1980). Therefore, I assessed homoscedasticity to ensure that no violations would impact confidence intervals and hypothesis tests (Rosopa et al., 2013). I used the Goldfeld-Quandt test to assess homoscedasticity, with non-significant results indicating homoscedasticity in the data (Goldfeld & Quandt, 1965). However, robust maximum likelihood can adjust for violations of homoscedasticity by providing standard errors that are robust to such violations (Wei et al., 2013).

Confirmatory Factor Analysis. Confirmatory factor analysis (CFA) is a methodological tool used to validate the underlying structure of latent constructs from a set of observed variables (Brown, 2015). Through CFA, I was able to examine the relationships between observed variables and their corresponding latent constructs (Schreiber et al., 2006). Additionally, CFA allowed me to predict the pattern of relationships and statistically evaluate this hypothesis, providing evidence for the construct validity of various measures (Kline, 2016).

I first assessed whether the observed variables in my research could form four latent variables: executive functions, explanatory coexistence, general cognitive ability (single indicator), and science education (single indicator). Based on the results of CFA, I then determined whether to proceed with structural equation modelling (SEM) or path analysis. In cases where CFA showed poor model fit or collinearity issues, I opted for path analysis instead of SEM for the main analysis.

Before conducting CFA, it was crucial to consider the initial assumptions underlying this statistical technique (Brown, 2015). Four key assumptions are critical for obtaining accurate and reliable results: (a) multivariate normality; (b) sufficient sample size; (c) random sampling; and (d) appropriate model specification. Table A14 provides the rationale and justification for these assumptions, as well as the decisions made regarding model construction to meet these requirements.

Table A14*Meeting Assumptions for Confirmatory Factor Analysis*

Assumptions	Rational	Justification	Decision
Multivariate normality	Multivariate normality influences the selection of the estimation method and the reliability of estimates (Byrne, 2016). Using an estimation method based on normal theory on a sample that does not conform to a multivariate normal distribution is risky (Knief & Forstmeier, 2021).	Substantial deviations from normality can inflate test statistics and skew standard errors, so that impacts the model's goodness-of-fit (Byrne, 2016).	I used robust maximum likelihood estimation to deal with nonnormal distribution of my dataset (Bishara & Hittner, 2012; Brown, 2015; Yuan & Bentler, 1998).
Sufficient sample size	Sufficient sample size is vital for the robustness and precision of parameter estimates (Field, 2018). Large sample sizes provide enhanced statistical power, crucial for identifying significant effects and relationships amongst variables, and also contribute to stable and accurate parameter estimates (Kline, 2016).	Rule-of-thumb based absolute minimum sample size: $N = 100$ or $N = 200$ (Bandalos, 2014). Model complexities based relative sample size: $n = 5-10$ or $n = 3-6$ per estimated parameters (Wang & Rhemtulla, 2021).	Given that my PhD research necessitates data collection from two countries, I took the feasibility of the study into account. In addition to the overall $N = 200$, I also considered the calculation method of collecting five participants for each parameter (Bentler & Chou, 1987). Counting the most complex confirmatory factor analysis model and path analysis model, a minimum of 180 participants were required for each country.

Table A14 continue

Random sampling	Random sampling provides an equal chance for individuals to be included in the research, making it suitable for studies involving homogenous populations where the target demographic is uniformly distributed (Elfil & Negida, 2017).	Select units individually, with each unit having an equal likelihood of being chosen at any given point (Singh, 2003).	My data collection was carried out online, this enhanced my ability to sample participants from various schools with diverse socioeconomic statuses from China and the United Kingdom.
Model specification	Model specification refers to the establishment of the linkages between variables, instead of deriving them from empirical or methodological perspectives (Kline, 2016),	Creation of models should base on the model, which is subsequently transformed into a path diagram depicting causal relationships between constructs/variables (Hox & Bechger, 1998).	Constructed model based on previous proposed the framework. I have four latent variables: with executive functions (behavioural inhibition, interference inhibition, switching, working memory) explanatory coexistence (animism, buoyancy, science statements), general cognitive ability(single indicator), and science education (single indicator).

After meeting the basic assumptions, there are four key steps in constructing confirmatory factor analysis: model identification, parameter estimation, model evaluation, and model modification (Hox & Bechger, 1998). In the following section, I explain these steps and the decisions I made during the construction of the confirmatory factor analysis.

Model Identification. Ensuring the model is identified is crucial (Bollen, 1989). An over-identified model rarely provides useful information about the relationships among variables (Bollen, 1989). Kline (2016) outlined three essential criteria for correctly specifying confirmatory factor analysis. In Table A15, I present these three criteria and explain how my model meets the requirements for model identification.

Table A15*Decision for Meeting Model Identification Criteria*

Criteria	Decision to Meet Criteria
Model should have degrees of freedom equal to or greater than zero, ensuring a non-negative degree of freedom	For all confirmatory factor analysis models, the degrees of freedom were greater than zero.
Scale setting for each latent variable, including error terms, requires fixing a factor loading and the path coefficient of the error term to one, or setting the latent variable's variance to one.	For latent variables with multiple observed variables, I fixed one factor loading to one and allow the error variances to be estimated. For a single indicator variable, I set the sole indicator's factor loading to one.
Every latent variable should be represented by a minimum of two indicators.	Executive functions had four observed variables; explanatory coexistence had three observed variables. Although general cognitive ability and science education had only one variable each, they fitted the case of a single indicator.

Parameter Estimation. Determining the numerical values for model parameters that most accurately reflect the relationships between observed and latent variables is crucial in constructing confirmatory factor analysis (Deng et al., 2018). This process involves computing model parameter values from the dataset, such as factor loadings, path coefficients, and variances and covariances (Brown, 2015). Various strategies, including maximum likelihood, generalised least squares, and robust maximum likelihood, can be used for this purpose (Cohen, 1988). Given the considerations about normality, I chose to use robust maximum likelihood to compute the model, as it ensures robust standard errors and provides a scaled test statistic.

I also distinguished between unstandardised and standardised estimates in the reporting of path coefficients. Unstandardised estimates indicate the expected change in the outcome variable for a one-unit change in the predictor variable, assuming all other predictors remain constant. In contrast, standardised estimates (also known as standardised coefficients or parameters) measure the relationships between variables after standardising all variables (Bring, 1994; Greenland et al., 1991). For ease of comparison across variables in the model, regardless of their scale or units of measurement, I primarily reported standardised coefficients in my analysis (Kwan & Chan, 2011).

Model Evaluation. Model evaluation uses fit indices to assess the performance and overall goodness of fit of the model (Goretzko et al., 2024). This evaluation depends on both specific path coefficients (such as p-values and standard errors) and the overall fit of the model (Brown, 2015). Fit indices provide valuable information on the goodness-of-fit for the designed models, making them essential in the construction of my confirmatory factor analysis model (Brown, 2015; Kline, 2016). Drawing on previous literature that used confirmatory factor analysis to study executive functions (Ellefson et al., 2020; Zonneveld et al., 2021), I selected six fit indices to measure the model fit. Details of these fit indices are presented in Table A16.

Table A16*Summary of Model Fit Indices*

Method	Rationale	Idea Results
Chi-square test (χ^2)	Detect discrepancies between the model-implied covariance matrix and the original covariance matrix (Kline, 2016).	$p > .05$ <i>Notes.</i> However, Chi-square test is sensitive to sample size, so that results from chi-square test should not be overly relied upon when sample size is large (Brown, 2015).
Root Mean Square Error of Approximation (<i>RMSEA</i>)	Identify model misspecification & <i>RMSEA</i> less sensitive to sample size compared to the χ^2 test (Xia & Yang, 2019).	$\leq .08$ for good fit (Kline, 2016).
Standardised Root Mean Square Residual (<i>SRMR</i>)	Quantify the standardised difference between the observed correlations and the ones predicted by the model (I.e., it compares the sample matrix to the model matrix) (Brown, 2015).	$\leq .08$ for good fit (Brown, 2015).
Comparative Fit Index (<i>CFI</i>)	Measure the variance accounted for in a covariance matrix and is less affected by sample size than the χ^2 test (Fan et al., 2016).	$\geq .90$ (Kline, 2016).

Table A16 continue		
Tucker-Lewis Index (<i>TLI</i>)	Build (mostly) out of ratios of chi-square / degrees of freedom and is independent of sample size (Bentler, 1990).	$\geq .90$ (Hu & Bentler, 1999).
Akaike Information Criterion (<i>AIC</i>)	Approximate the information lost when a given model is used to generate data (Akaike, 1998).	Model with the lowest <i>AIC</i> has better fit (Cavanaugh & Neath, 2019).
Bayesian Information Criterion (<i>BIC</i>)	Measure the efficiency of the parameterised model in terms of predicting the data (Akaike, 1998; Konishi & Kitagawa, 2008).	Model with the lowest <i>BIC</i> has better fit (Konishi & Kitagawa, 2008).

Model Modification. Based on model evaluation, model modification may be necessary to adjust the model and improve its fit (Brown, 2015). Model modification increases parameter constraints, particularly when shared relationships exist among the measured variables (Ellefson et al., 2020; Schreiber et al., 2006). A high modification index ($MI > 10$) may indicate a potential covariation, but such a constraint should only be incorporated if it can be theoretically justified (Hoyle & St, 2022). However, overfitting the model can compromise its generalisability (Schreiber et al., 2006). Therefore, model modification requires a careful balance between improving model fit and preserving the model's theoretical relevance. After constructing the confirmatory factor analysis, I determined that model modification was not necessary, as I ultimately selected path analysis.

A.9.2 Main Analysis

Through the preliminary analysis described earlier, I was able to proceed with the main data analysis, adopting appropriate analytical methods and adjustments. Due to issues with the latent variables formed by confirmatory factor analysis (see Chapters 5 & 6), I ultimately chose path analysis to explore the complex relationships between the variables. In the main analysis, correlation and path analysis were used in both Study 1 and Study 2, while multigroup analysis comparisons between the two countries, scatter plots, and t-tests were applied only in Study 2, which involved data from both countries.

Correlation. As my PhD project involved exploring complex relationships among multiple variables, I utilised correlation analysis to observe the degree and nature of covariation between variables (Field, 2018). Correlation analysis could assist me in understanding relationships, and validating measures and procedures (Schober et al., 2018). More importantly, correlation analysis assists in the interpretation of more complex analyses, including confirmatory factor and path analyses. Correlation analysis also helps to ensure robust correlations among variables intended to measure the same latent factor (Ellefson et al., 2020; Lawson & Farah, 2017; Xu et al., 2020).

For this thesis's central analysis, correlation was crucial for exploring adherence to the assumptions of confirmatory factor analysis, and for identifying potential multicollinearity issues. For instance, a r -value greater than .85 may indicate a possible multicollinearity problem (Brown, 2015; Dormann et al., 2013; Field, 2018). When conducting correlation analysis, I mainly used Pearson's correlation coefficient, as it is robust against violations of normality when using large sample sizes (Bishara & Hittner, 2012). Effect sizes for correlations are typically classified as large for r -values greater than .50, medium for r -values greater than .30, and small for r -values greater than .10 (Field, 2018). To further check the robustness of the thesis' results, a

non-parametric Spearman's rank correlation was also performed. These results were consistent with those obtained from Pearson's correlation, meaning that only the latter was reported on in the the results section.

Path Analysis. Path analysis is a multivariate statistical methodology which allows for both graphical and analytical approaches that depict and scrutinise intricate constellations of associations among all measured variables (Streiner, 2005). Path analysis was used in this thesis to provide a more nuanced examination into the complex relationships among all variables explored in this research, extending beyond what multiple regression could observe (Streiner, 2006). For example, path analysis could measure presumption of causal sequencing (Denis & Legerski McCormick, 2006), as well as offer structural equation modelling where only single indicators (i.e., observed variables) are utilised for each of the constructs within the model (Tarka, 2018). Path analysis was therefore used to analyse all key research questions for both Studies 1 and 2.

The evaluation of models and assumptions, as well as the steps for model construction in path analysis, are consistent with requirements and steps in confirmatory factor analysis (see Appendix A.9.1). However, one problem arose while constructing the path analysis: after constructing internal correlations among all dependent variables, the degrees of freedom for the full path model became zero, indicating a just-identified model. Therefore, after measuring all models, I removed one regression (the link between working memory and the Animism task) that was not significant across all models, which resulted in a degree of freedom of one; this allowed me to avoid issues of the model's being overidentified (Ellefson et al., 2024).

While observing the path analysis results, in addition to determining whether there was a significant relationship between variables based on the *p*-value, I also used standardised estimates to present the results. Standardised path coefficients, also known as beta coefficients, can reflect the strength and direction of the relationships

between variables, and are useful for assessing effect size (Schumacker et al., 2016). Standardised path coefficients provide a measure of how much the dependent variable is expected to change (in standard deviation units) for a one standard deviation change in the independent variable, holding other variables constant (Kline, 2015). In this study's result sections (Chapters 5 and 6), standardised estimates are represented by β . A β of .50 or greater indicates a large effect size, a β between .30 and .50 indicates a medium effect size, and a β between .10 and .30 indicates a small effect size. When β is less than .10, this suggests that, although the relationship between the two variables is significant, the predictor may have a negligible impact on the outcome variable (Schumacker et al., 2016).

Multigroup Analysis. Multigroup analysis was used to assess whether the relationships among executive functions, general cognitive ability, science education and explanatory coexistence are consistent across different countries. Multigroup analysis, also known as measurement invariance for confirmatory factor analysis or structure equation modelling, offers a robust method for evaluating link invariances among variables (Byrne, 2016). Multigroup analysis thus helps me check whether the relationships among variables are similar across diverse populations or conditions (Brown, 2015). Multigroup analysis in path analysis mainly focuses on assessing the equivalency of items' intercepts and path coefficients across groups (Chen et al., 2010; Evermann, 2010; Ryu, 2015). Therefore, when constructing multigroup analysis, I constrained both the intercepts and path coefficients in each group to be the same, and compared the constrained model with original model (Grace & Jutila, 1999). Non-significant comparison results indicated that the invariance was satisfied, and the relationships between executive functions, general cognitive ability, science education and explanatory coexistence do not vary by countries. However, significant multigroup analysis results indicated that the relationships

among variables between the two countries differed, necessitating separate model constructions (Evermann, 2010).

Scatterplot Illustration. Scatter plots graphically represent the relationship between two variables, providing insights into the pattern of their association (Sainani, 2016). By using scatter plots, I was able to observe the mutual influence between accuracy and response time in each task (Field, 2018). In Chapter 3, I discuss how direct mean comparisons between participants from the United Kingdom and China might introduce bias when interpreting cognitive abilities through a cross-cultural lens. Therefore, instead of using t-tests for direct comparisons of efficiency, I used scatter plots to illustrate the response patterns between accuracy and response time for executive functions, general cognitive ability, and explanatory coexistence in both countries. Although scatter plots do not directly provide comparison parameters, they enhance understanding by visualising the similarities and differences in response patterns across tasks. Thus, these plots provide valuable context for interpreting the results of the correlation and path analyses conducted in both countries.

Appendix B: Revision of Animism and Flanker Tasks

Revision of Flanker Task

In my pilot studies, I conducted a small study to decide which two complex incongruent trials to use for the Flanker task in order to balance the number of different trial types. This pre-research was conducted after my ethics application got approved. The data from 10 adults' participants were collected remotely from the total of 72 trials (see Table B1). Via analysing the mean accuracy, response time and efficiency score of these tasks (see Table B2), it was found that the complex incongruent trials number 7 and 12 had more variations on results. Thus, these two trials were selected as the complex incongruent trials.

Table B1*Materials of the Pilot Flanker Task*

Direction of Fish					Trial Type	Trial Number	Repeat times
Fish1	Fish2	Fish3	Fish4	Fish5			
right	right	right	right	right	Congruent	1	9
left	left	left	left	left	Congruent	2	9
right	right	left	right	right	Simple Incongruent	3	9
left	left	right	left	left	Simple Incongruent	4	9
left	right	left	right	left	Complex Incongruent	5	3
right	left	left	left	right	Complex Incongruent	6	3
left	right	left	right	right	Complex Incongruent	7	3
right	left	left	right	right	Complex Incongruent	8	3
right	right	left	left	right	Complex Incongruent	9	3
right	right	left	right	left	Complex Incongruent	10	3
right	left	right	left	left	Complex Incongruent	11	3
left	right	right	left	left	Complex Incongruent	12	3
left	left	right	right	left	Complex Incongruent	13	3
left	left	right	left	right	Complex Incongruent	14	3
right	left	right	left	right	Complex Incongruent	15	3
left	right	right	right	left	Complex Incongruent	16	3

Table B2*Results of the Pilot Flanker Task¹*

	Accuracy (N = 10)	Response Time (N = 10)	Efficiency (N = 10)
Complex Incongruent	0.95959596	0.429803333	2.23263964
Simple Incongruent	0.934343434	0.441083295	2.118292495
Congruent	0.96969697	0.42463798	2.283585114
1	0.974747475	0.431458333	2.259192602
2	0.964646465	0.417817626	2.308773982
3	0.929292929	0.442895556	2.098221392
4	0.939393939	0.439271035	2.138529208
5	0.96969697	0.435791667	2.225138854
6	0.924242424	0.416195606	2.220692412
7	0.924242424	0.442062045	2.090752721
8	0.939393939	0.419658485	2.238472409
9	0.954545455	0.417662348	2.285447702
10	0.924242424	0.434487803	2.12719993
11	1	0.446038106	2.241960914
12	0.939393939	0.452839924	2.07445035
13	0.984848485	0.415483788	2.370365616
14	0.984848485	0.419996136	2.344898916
15	0.96969697	0.419061364	2.313973689
16	1	0.438362727	2.281215846

Notes. Response Time here refers to response time for correct trials.

¹ As the time differences were small, I kept all decimal places recorded

Revision of Animism Task

For maintaining a balanced number of different categories of living/non-living and moving/non-moving trials within the Animism task, I have selected only two types in the 'Animal' sub-category. Thus, I analysed the data from my masters' research which shows that, whilst there are no significant differences between animal types in terms of response time and accuracy (see Table B1), there is a significant difference between the efficiency of each type. Through observing the violin plot of the efficiency score (see Figure B1), it was found that the difference is potentially caused by the outliers within the mammal group, which resulted in the outlier on its response time. Considering that choosing groups would generate more variance, I selected the groups 'Mammal' and 'Insect' for the 'Animal' subcategory in the PhD research.

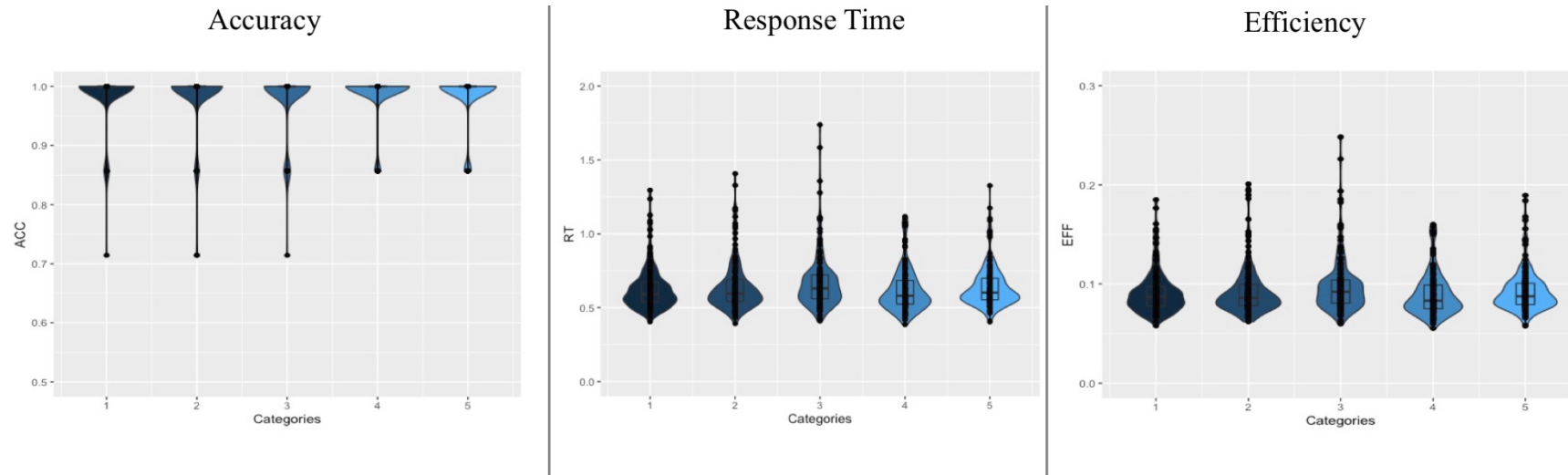
Table B3*ANOVA of Different Animal Sub-categories*

	Fish (N=155)		Insect (N=155)		Mammal (N=155)		Bird (N=155)		Reptile (N=155)		<i>F</i> (4,770)	<i>p</i>	η^2
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			
Accuracy	0.99	0.05	0.99	0.05	0.98	0.05	0.99	0.04	0.99	0.04	0.54	.708	0.01
Response Time	0.63	0.15	0.64	0.17	0.67	0.19	0.62	0.15	0.64	0.14	2.30	.057	0.04
Efficiency	0.099	0.02	0.09	0.02	0.10	0.03	0.09	0.02	0.09	0.02	2.66	.032	0.02

Notes. Response Time here refers to response time for correct trials.

Figure B1

Boxplot of the Different Animal Sub-categories in terms of Accuracy, Response Time and Efficiency



Notes. 1- Fish 2-Insect 3-Mammal 4-Bird 5-Reptile; Response Time here refers to response time for correct trials.

Appendix C: Influence of Fixed Button of Animism Task and Science Statement

Task

For the counterbalance setting of the reaction key, taking into consideration the fixed response option of the Science Statement and Animism tasks (see Table C1, Figure C2), we designed four versions in my masters' research:

- Version 1: Living -Right/Non-Living-Left; True-Left/False-Right;
- Version 2: Living-Left/Non-Living-Right; True-Right/False-Left;
- Version 3: Living-Left/Non-Living-Right; True-Left/False-Right;
- Version 4: Living-Right/Non-Living-Left; True-Left/False-Right.

No significant differences were present between four different versions of the Animism task. However, significant differences were present in the response time of the Science Statement task (see Table C1). Observations of the violin plot indicate that this difference has potentially been caused by the outliers in Version 3 (see Figure B1). An alternative analysis indicated that there was no difference regarding the placement of the true and false buttons on the left or right (see Table C2, Figure C3). Hence, I selected Version 3 in my pilot study as it would potentially generate more differences and it's also more commonly used in cognitive tasks. Response Time reported in later section refers to response time for correct trials.

Table C1*ANOVAs of Different Versions*

		Version 1 (n=39)		Version 2 (n=40)		Version 3 (n=37)		Version 4 (n=39)		<i>F</i>	<i>p</i>	η^2
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	(3, 151)		
Animism Task	Response Time	0.71	0.18	0.73	0.19	0.67	0.10	0.73	0.16	0.89	.45	0.06
	Accuracy	0.95	0.05	0.95	0.05	0.95	0.04	0.95	0.05	0.29	.84	0.00
	Efficiency	107.70	22.34	105.93	22.50	111.37	16.80	104.28	19.66	0.83	.48	0.58
Science Statement Task	Response Time	2.78	1.01	2.76	0.94	3.41	1.74	2.73	0.89	2.88	.04	0.26
	Accuracy	0.76	0.07	0.74	0.08	0.75	0.06	0.75	0.06	1.01	.39	0.03
	Efficiency	12.06	3.58	11.97	4.56	10.18	3.44	11.87	3.21	2.16	.10	0.40

Table C2*T-test of left/right Version of Science Statement Task*

		Versions 1&4 (n=78)		Versions 2&3 (n=77)		<i>t</i> (3, 151)	<i>p</i>	<i>d</i>
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			
Science Statement Task	Response Time	2.75	0.94	3.07	1.41	2.88	.103	0.27
	Correct Rate	0.75	0.07	0.74	0.07	1.01	.262	0.14
	Efficiency	11.96	3.38	11.11	4.14	2.16	.160	0.22

Figure C1

Violin plot of *Different Versions of Science Statement Task*

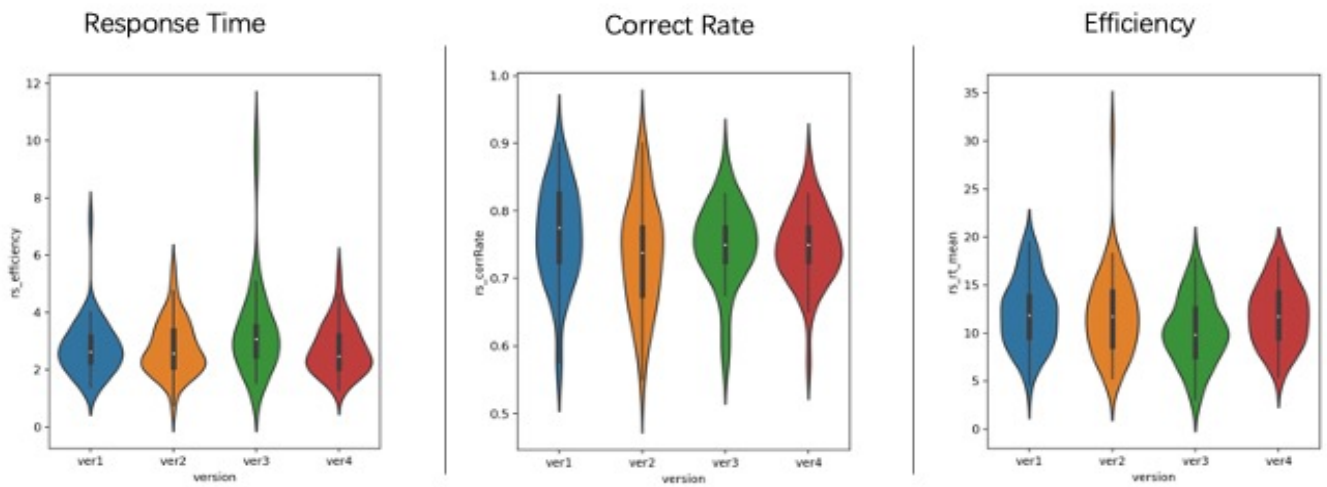


Figure C2

Violin plot *Different Versions of Animism Task*

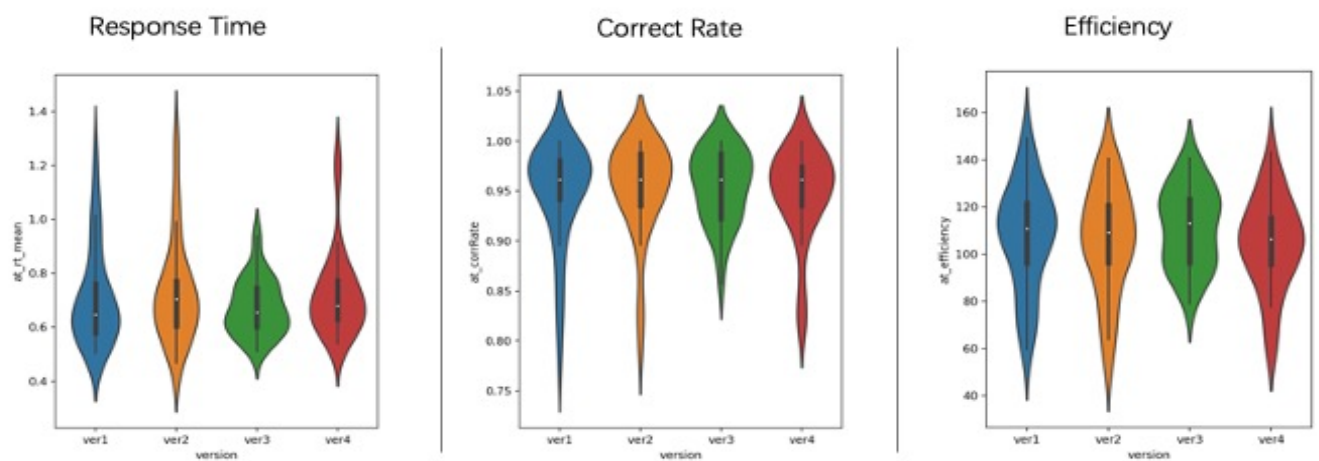
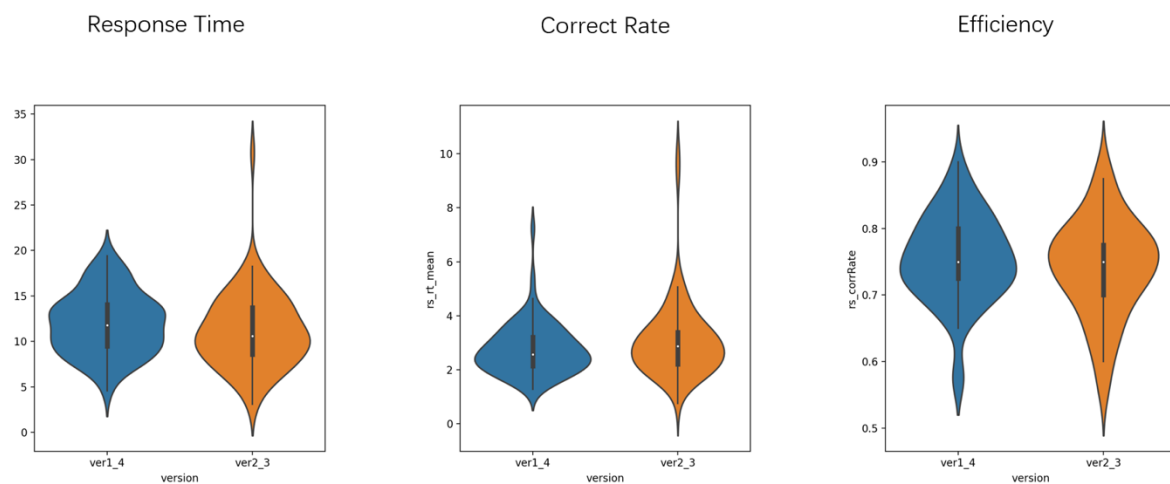


Figure C3

Violin plot of *left/right Version of Science Statement Task*



Appendix D: Participant Questionnaire

Participant Questionnaire

参与者调查

How old are you? (Please enter your age in years): _____

您的年龄是？(请输入您的年龄) _____ 岁

What is Your Gender (choose one): Female; Male; Non-binary; Prefer not to say

您的性别是（选择一项）：女；男；其他；不想说

Are you currently a student? (choose one): Yes ; No

您现在是一名学生吗？（选择一项）是；不是

What level is your current level of study? (choose one): Undergraduate; PGCE; Masters; Doctorate; Other(Please describe)

您目前的学历是？（选择一项）本科生；教师培训学位；硕士；博士；其他（请说明）

（注：PGCE，即教育学研究生，是一个英国才有的学历制度，相当于国内的教师资格证。如果调查的对象以国内学生为主，建议删去“教育学研究生”）

What subject(s) are you studying? If appropriate, please include your major and minor.

您在学习什么学科？如果可以的话，请给出您的主修和辅修专业。

In What country is your current university located?

Please write the name of the country _____

您现在的大学在哪个国家（地区）？

请给出国家（地区）的名称 _____

（注：部分填写问卷的人士就读的大学可能位于尚未被国际认可为国家的地区（如台湾、库拉索等）。为避免由此带来的麻烦，故作此增译。如认为不需要，可以删除“地区”）。

What is the subject of your undergraduate degree(s)? If appropriate, please include your major and minor.

您获得了哪些学科的学士学位？如果可以的话，请给出您的主修和辅修专业所在的学科。

In which country is the university of your undergraduate degree(s) located?

Please write the name of the country_____

您在哪个国家（地区）获得了学士学位？

请给出国家（地区）的名称_____

Have you completed any other masters degrees besides the one you are studying now? Yes No

除了正在攻读的硕士学位以外，您还获得了其他硕士学位吗？是 否

What is the subject of your previous masters degree(s)?

您之前获得了哪些学科的硕士学位？

In what country is the university of your previous masters degree(s) located?

Please write the name of the country_____

您以前攻读硕士学位的大学在哪个国家（地区）？

请给出国家（地区）的名称_____

Please write all of the countries where you studied before attending university. Please list each country with the level(s) of education in brackets beside it.

Please write the name of the country_____

请给出您在上大学之前所有就读过的学校所在的国家（地区）。每写一个国家（地区），请在旁边用括号注明在该国家（地区）取得的学历。

请给出国家（地区）的名称_____

What is your highest level of education in the following subjects? (choose one)

在以下学科中，您的最高学历是什么？（每个学科选择一项）

Biology (None; Secondary/Middle School; GCSE/O-Levels/Junior Cert; some A-Levels/Leaving Cert; completed A-Levels; some undergraduate; completed bachelor's degree; some masters; completed master; some doctoral; completed doctoral; other_____)

生物学（无；中学；普通中等教育证书/普通教育证书/初中毕业证书；部分高中课程/高中毕业证书；完成高中课程；部分专科生课程；取得专科学位；部分本科课程；取得学士学位；部分硕士课程；取得硕士学位；部分博士课程；取得博士学位；其他_____）

（注：原文中并未出现“专科课程”（即绿字部分）对应的内容。从翻译契合的角度来说，建议删除，但考虑到中国存在专科，请进一步考虑是否删除）

Chemistry (None; Secondary/Middle School; GCSE/O-Levels/Junior Cert; some A-Levels/Leaving Cert; completed A-Levels; some undergraduate; completed bachelor's degree; some masters; completed master; some doctoral; completed doctoral; other_____)

化学（无；中学；普通中等教育证书/普通教育证书/初中毕业证书；部分高中课程/高中毕业证书；完成高中课程；部分专科生课程；取得专科学位；部分本科课程；取得学士学位；部分硕士课程；取得硕士学位；部分博士课程；取得博士学位；其他_____）

Physics (None; Secondary/Middle School; GCSE/O-Levels/Junior Cert; some A-Levels/Leaving Cert; completed A-Levels; some undergraduate; completed bachelor's degree; some masters; completed master; some doctoral; completed doctoral; other_____)

物理学（无；中学；普通中等教育证书/普通教育证书/初中毕业证书；部分高中课程/高中毕业证书；完成高中课程；部分专科生课程；取得专科学位；部分本科课程；取得学士学位；部分硕士课程；取得硕士学位；部分博士课程；取得博士学位；其他_____）

Engineering (None; Secondary/Middle School; GCSE/O-Levels/Junior Cert; some A-Levels/Leaving Cert; completed A-Levels; some undergraduate; completed bachelor's degree; some masters; completed master; some doctoral; completed doctoral; other_____)

工程学（无；中学；普通中等教育证书/普通教育证书/初中毕业证书；部分高中课程/高中毕业证书；完成高中课程；部分专科生课程；取得专科学位；部分本科课程；取得学士学位；部分硕士课程；取得硕士学位；部分博士课程；取得博士学位；其他_____）

Mathematics (None; Secondary/Middle School; GCSE/O-Levels/Junior Cert; some A-Levels/Leaving Cert; completed A-Levels; some undergraduate; completed bachelor's degree; some masters; completed master; some doctoral; completed doctoral; other_____)

数学（无；中学；普通中等教育证书/普通教育证书/初中毕业证书；部分高中课程/高中毕业证书；完成高中课程；部分专科生课程；取得专科学位；部分本科课程；取得学士学位；部分硕士课程；取得硕士学位；部分博士课程；取得博士学位；其他_____）

Psychology (None; Secondary/Middle School; GCSE/O-Levels/Junior Cert; some A-Levels/Leaving Cert; completed A-Levels; some undergraduate; completed bachelor's degree; some

masters; completed master; some doctoral; completed doctoral; other_____)

心理学(无; 中学; 普通中等教育证书/普通教育证书/初中毕业证书; 部分高中课程/高中毕业证书; 完成高中课程; 部分专科生课程; 取得专科学位; 部分本科课程; 取得学士学位; 部分硕士课程; 取得硕士学位; 部分博士课程; 取得博士学位; 其他_____)

Other science (Please describe) (None; Secondary/Middle School; GCSE/O-Levels/Junior Cert; some A-Levels/Leaving Cert; completed A-Levels; some undergraduate; completed bachelor's degree; some masters; completed master; some doctoral; completed doctoral; other_____)

其他理科类学科(请指明)(无; 中学; 普通中等教育证书/普通教育证书/初中毕业证书; 部分高中课程/高中毕业证书; 完成高中课程; 部分专科生课程; 取得专科学位; 部分本科课程; 取得学士学位; 部分硕士课程; 取得硕士学位; 部分博士课程; 取得博士学位; 其他_____)

(注: 此处私认为有必要强调“理科类”(蓝字部分), 因为本问卷中并未涉及任何人文学科的内容, 只说“学科”容易影响答案分布。当然, 具体强调与否由您依据实际情况来决定。)

Have you ever taught a science subject? (choose one) Yes; No

您有没有教过科学课程? (选择一项) 有; 没有

If yes-

若上一问您的答案是“有”,

What science subject have you ever taught? Please select:

您教的是什么科目? 请选择:

Biology; Physics; Chemistry; Engineering; Mathematics; Psychology; Other sciences_____

生物学; 物理学; 化学; 工程学; 数学; 心理学; 其他学科_____

What level of that science subject have you ever taught? Please select:

您教授的课程是哪个阶段的? 请选择:

Primary School; Secondary/Middle School; High School; Undergraduate; Master; Doctoral; Others_____

小学; 中学; 高中; 本科; 硕士; 博士; 其他_____

Appendix E: Consent Form

CONSENT FORM

Thinking about Science

If you agree to participate in this study, then please complete this consent form.

	Please circle one:	
I agree to participate in this research study.	Yes	No
I have read the information letter, had an opportunity to consider the information, ask questions, and received satisfactory answers.	Yes	No
I understand that:		
My participation is voluntary and that I am free to withdraw at any time without giving any reason.	Yes	No
Scientific data collected as part of this study will be stored anonymously on computer or in paper form. This anonymised data may contribute to scientific papers and presentations.	Yes	No
Scientific data might be shared according to Open Science standards, but no identifying information will allow others to trace responses back to me.	Yes	No

Lead Researcher for this Study: Rui Wang (email: rw623@cam.ac.uk)

This study has been reviewed by University of Cambridge Faculty of Education ethics panel

知情同意书

科学学习的思考研究

如果您同意参与这项研究，请您完成这份知情同意书。

	请选择一项	
我同意参与这项研究	是	否
我已经阅读过这项研究的信息说明，并已思考过有关这项实验的信息，了解到我有机会研究人员提出问题，并且得到满意的解答。	是	否
我理解：		
我的研究是完全自愿的，我可以在任何情况下选择退出实验，且并不需要给予任何理由。	是	否
科学数据作为这项研究的一部分将以电子或纸质形式匿名储存。这种匿名的数据将有助于科学论文和报告。	是	否
根据开放科学的标准，我的数据可能会被共享，但任何信息都无法被追踪和识别到我本人。	是	否

研究主试：王瑞 (邮件: rw623@cam.ac.uk)

本研究已经通过剑桥大学教育学校系伦理委员会批准

Appendix F: Information Sheet

English Version:

INFORMATION SHEET *Thinking About Science*

What is being investigated? We are studying people's thinking skills involved in science understanding.

How long does it take? The research takes approximately 50-60 minutes

What will I do? You will complete a short demographical questionnaire and complete nine computerised tasks on general thinking skills and science using PsychoPy (each taking about 5-10 minutes each).

How will my data be used? The data collected for this study are likely to be analysed and reported at professional conferences and/or in academic books or articles. These reports will be written based on group data. To help advance our field to promote Open Science, sometimes individual responses are shared with other researchers. For the individual responses, no identifying information will allow others to trace responses back to the person who gave them.

Is this research voluntary? This study is completely voluntary, that means that it is your decision whether or not you want to complete the tasks. If you decide to participate, then you are free to withdraw your consent or to discontinue your participation at any time without penalty.

Will by responses be kept anonymous and confidential? All information gathered will be kept strictly confidential and only used for research purposes. We store the data using random ID numbers so that the data cannot be linked back to you or your family. The data will be stored on secured electronic files or in a locked cabinet in the Faculty of Education, and it will be accessible only by the PI, the lead researcher for this study (Wang Rui).

What are the risks? There are likely no known risks associated with participating in this study.

Ethics. Approval for this study have been granted by the ethics committees in the Faculty of Education at the University of Cambridge.

Contact Information. Please do not hesitate to contact us if you have any questions or concerns you would like to discuss.

Rui Wang: rw623@cam.ac.uk

Prof. Michelle Ellefson: mre33@cam.ac.uk

Note - you must be 18 years of age or older to participate in this study

Please do not hesitate to contact us if you have any questions or concerns you would like to discuss.

Chinese Version

信息说明 科学学习的思考研究

这项研究调查什么？我们正在研究人们在理解科学知识方面的思维能力。

完成这项研究需要多长时间？研究大约需要 50-60 分钟

我需要做什么？你将完成一份简短的人口学统计调查问卷，并完成 9 项关于一般思维技能和科学学习的电脑程序任务（每个任务大约需要 5-10 分钟）。

我的实验数据将如何被使用？本研究收集的数据可能会被报告在专业的学术会议，学术书籍或学术文章中。这些报告将被以数据集的形式进行分析。此外，为了推进心理学领域建设，促进开放科学，属于将在开放科学的网站上与其他的研究人员分享。但是对于个人的回答，并不会有任何可识别信息可以追溯到您。

参与这项研究是自愿的吗？是的，参与这项研究是完全自愿的。您可以自己决定是否完成实验。在研究中您可以随时拒绝参与或中止您的回答，且不受任何处罚。

我的数据是否会被匿名保密？所有收集到的信息将被严格保密，并仅用于学术研究目的。我们使用随机 ID 号存储数据，这样数据就无法追溯回您或您家人的身份。这些数据将存储在安全的电子文件中，或者存储在剑桥大学教育学院，只有实验室负责人和这项研究的首席研究员（王瑞）才能访问这些数据。

这项研究有什么风险？目前参与这项研究可能没有已知风险。

伦理审查：剑桥大学教育学院的伦理委员会已经批准了这项研究。

联系方式：如果您有任何问题或顾虑，请随时与我们联系。

研究员王瑞：rw623@cam.ac.uk

Michelle Ellefson 教授： mre33@cam.ac.uk

注意-您必须年满 18 岁才能参加本研究

如果您有任何问题或顾虑，请随时与我们联系。

Appendix G: Science Statements Task

Table F1

Science Statements

Sentence English	Sentence Chinese	Domain	Concept	Naïve	Science
Physics Statements					
Rocks are composed of matter.	岩石由物质构成。	Matter	Mass	T	T
Numbers are composed of matter.	数字由物质构成。	Matter	Mass	F	F
Fire is composed of matter.	火由物质构成。	Matter	Mass	T	F
Air is composed of matter.	空气由物质构成。	Matter	Mass	F	T
A liter of water weighs more than a liter of air.	一公升水比一公升空气重。	Matter	Weight	T	T
A pound of steal weighs more than a ton of steel.	一磅的钢铁比一吨的钢铁重。	Matter	Weight	F	F
A pound of steal weighs more than a pound of feathers.	一磅的钢铁比一磅的羽毛重。	Matter	Weight	T	F
A liter of water weighs more than a liter of ice.	一升水比一升冰重。	Matter	Weight	F	T
Steal is denser than foam.	钢铁比泡沫密度大。	Matter	Density	T	T
Foam is denser than brick.	泡沫比砖密度大。	Matter	Density	F	F
Ice is denser than water.	冰比水密度大。	Matter	Density	T	F
A cold penny is denser than a hot penny.	冷的一分钱比热的一分钱密度大。	Matter	Density	F	T
Forces can be strong or weak.	力可以是强大的, 也可以是弱小的。	Mechanics	Force	T	T
Forces can be male or female.	力可以是男的, 也可以是女的。	Mechanics	Force	F	F
Forces can be straight or curved.	力可以是直的, 也可以是弯曲的。	Mechanics	Force	T	F

Forces can be balanced or unbalanced.	力可以是平衡的，也可以是不平衡的。	Mechanics	Force	F	T
A moving bullet loses speed.	移动中的子弹会失去速度。	Mechanics	Velocity	T	T
A moving bullet loses weight.	移动中的子弹会失去重量。	Mechanics	Velocity	F	F
A moving bullet loses force.	移动中的子弹会失去力。	Mechanics	Velocity	T	F
A moving bullet loses height.	移动中的子弹会失去高度。	Mechanics	Velocity	F	T
Anvils fall through air faster than feathers.	铁砧从空中掉落的速度比羽毛快。	Mechanics	Gravity	T	T
Bright objects fall through air faster than dark objects.	亮的物体从空中掉落的速度比暗的物体快。	Mechanics	Gravity	F	F
Heavy balls fall through air faster than light balls.	重的球从空中掉落的速度比轻的球快。	Mechanics	Gravity	T	F
Pointy objects fall through air faster than flat objects.	尖锐的物体从空中掉落的速度比扁平的物体快。	Mechanics	Gravity	F	T
Constant acceleration requires constant force.	恒定的加速度需要恒定的力。	Mechanics	Acceleration	T	T
Constant speed requires constant acceleration.	恒定的速度需要恒定的加速度。	Mechanics	Acceleration	F	F
Constant motion requires constant force.	恒定的运动需要恒定的力。	Mechanics	Acceleration	T	F
Constant force can yield constant rest.	恒定的力可以产生恒定的静止。	Mechanics	Acceleration	F	T
The sun produces light.	太阳产生光。	Astronomy	Star	T	T
The sun produces sound.	太阳产生声音。	Astronomy	Star	F	F
The moon produces light.	月亮产生光。	Astronomy	Star	T	F
The sun produces gravity.	太阳产生引力。	Astronomy	Star	F	T
Turtles can die.	乌龟会死。	Physiology	Death	T	T
Screwdrivers can die.	螺丝刀会死。	Physiology	Death	F	F
Clouds can die.	云会死。	Physiology	Death	T	F

Mushrooms can die.	蘑菇会死。	Physiology	Death	F	T
Tigers reproduce.	老虎可以繁殖。	Physiology	Reproduction	T	T
Chairs can reproduce.	椅子可以繁殖。	Physiology	Reproduction	F	F
Caterpillars can reproduce.	毛毛虫可以繁殖。	Physiology	Reproduction	T	F
Ferns can reproduce	女性可以繁殖。	Physiology	Reproduction	F	T
People turn food into energy.	人把食物变成能量。	Physiology	Metabolism	T	T
Rocks turn food into energy.	岩石把食物变成能量。	Physiology	Metabolism	F	F
Plants turn food into energy.	植物把食物变成能量。	Physiology	Metabolism	T	F
Virus turn food into energy.	病毒把食物变成能量。	Physiology	Metabolism	T	F
Bacteria turn food into energy.	细菌把食物变成能量。	Physiology	Metabolism	F	T
Biological species evolve.	生物物种会进化。	Evolution	Adaptation	T	T
Inanimate objects evolve.	没有生命的物体会进化。	Evolution	Adaptation	F	F
Individual organisms evolve.	生物个体会进化。	Evolution	Adaptation	T	F
Computer viruses evolve.	计算机病毒会进化。	Evolution	Adaptation	F	T
Hair colour is heritable.	发色是可遗传的。	Genetics	Heritability	T	T
Pierced ears are heritable.	打了耳洞的耳朵是可遗传的。	Genetics	Heritability	F	F
Immunity to chickenpox is heritable.	对水痘的免疫能力是可遗传的。	Genetics	Heritability	T	F
Intelligence is heritable.	智力是可遗传的。	Genetics	Heritability	F	T
Rotting meat contains germs.	腐肉中含有细菌。	Germ	Contamination	T	T
Sunshine contains germs.	阳光中含有细菌。	Germ	Contamination	F	F
Urine contains germs.	尿液中含有细菌。	Germ	Contamination	T	F
Dish sponges contain germs.	海绵中含有细菌。	Germ	Contamination	F	T
Germ enter the body through cuts.	细菌通过伤口进入人体。	Germ	Infection	T	T
Germ enter the body through the hair.	细菌通过头发进入人体。	Germ	Infection	F	F

Germs enter the body through the skin.	细菌通过皮肤进入人体。	Germs	Infection	T	F
Germs enter the body through the eyes.	细菌通过眼睛进入人体。	Germs	Infection	F	T
Fish are alive.	鱼是活的。	Physiology	Life	T	T
Rocks are alive.	岩石是活的。	Physiology	Life	F	F
The sun is alive.	太阳是活的。	Physiology	Life	T	F
Coral is alive	珊瑚是活的。	Physiology	Life	F	T

Note: Regarding statement that "plants turn food into energy," there was different results between the accuracy obtained in my master's research and the results from (Shtulman and Valcarcel, (2012)' s original paper. After consulting with several biology teachers, I added the statement "Viruses turn food into energy." Experiments show that in the United Kingdom, the accuracy for "Viruses turn food into energy" is 78.4%, while for "Plants turn food into energy" it is 19.5%. In China, the accuracy for "Viruses turn food into energy" is 66.4%, and for "Plants turn food into energy," it is 50.2%. Therefore, in our final analysis, I chose to only analyse the statement "Viruses turn food into energy."

Appendix H: Accuracy Analysis (the United Kingdom)

Table H1

Descriptive statistics for Accuracy (the United Kingdom)

Variables	<i>M</i>	<i>SD</i>	<i>Skewness</i>	<i>Kurtosis</i>
General Cognitive Ability	.55	.19	.26	-.89
Executive Functions				
Stop Signal Task	.90	.06	-3.12	17.97
Figure Matching Task	.96	.05	-4.82	3.66
Fish Flanker Task	.96	.07	-5.81	36.67
Corsi Block Task	.55	.12	-.48	.56
Explanatory Coexistence				
Buoyancy Task				
Overall	.94	.11	-4.49	23.14
Congruent	.96	.10	-5.60	34.87
Incongruent	.91	.15	-3.94	16.92
Animism Task				
Overall	.94	.06	-2.10	6.01
Congruent	.96	.05	-4.00	22.78
Incongruent	.91	.09	-1.61	2.87
Science Statement Task				
Overall	.73	.08	-.32	.65
Congruent	.88	.09	-1.88	5.74
Incongruent	.57	.11	.18	-.22

Table H2*Age adjustment test for Accuracy (the United Kingdom)*

	Linear	Quadratic	Cubic
General Cognitive Ability	.01	.01	.00
Executive Functions			
Stop Signal Task	.00	.00	.01
Figure Matching Task	.00	.01	.00
Fish Flanker Task	.00	-.01	-.01
Corsi Block Task	.01	.01	.00
Explanatory Coexistence			
Buoyancy Task			
Overall	.00	.00	.00
Congruent	.00	.00	.00
Incongruent	.00	.00	.00
Animism Task			
Overall	.00	.00	.00
Congruent	-.01	.00	.00
Incongruent	-.01	.00	-.01
Science Statement Task			
Overall	.00	.00	.00
Congruent	-.01	.00	.00
Incongruent	-.01	.00	-.01

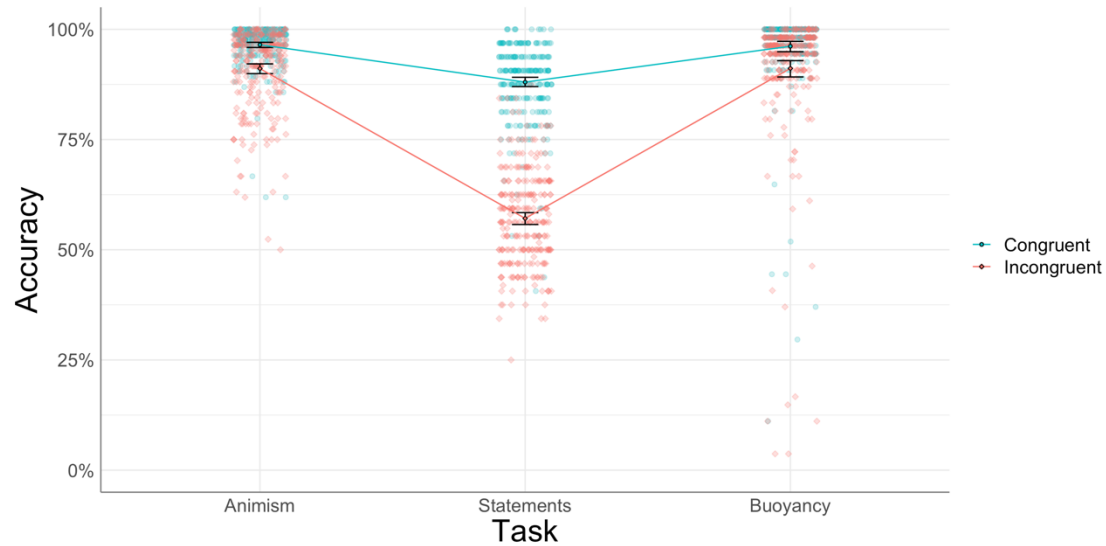
*Notes. * $p < .05$; ** $p < .01$; *** $p < .001$*

Table H3*Shapiro-Wilk test for Accuracy (the United Kingdom)*

Variables	<i>Shapiro-Wilk</i>	<i>p</i>
Science Education	.76	< .001
General Cognitive Ability	.97	< .001
Executive Function skills		
Stop Signal Task	.77	< .001
Figure Matching Task	.57	< .001
Fish Flanker Task	.37	< .001
Corsi Block Task	.96	< .001
Explanatory Coexistence		
Buoyancy Task	.44	< .001
Overall	.44	< .001
Congruent	.36	< .001
Incongruent	.51	< .001
Animism Task	.79	< .001
Overall	.79	< .001
Congruent	.62	< .001
Incongruent	.83	< .001
Science Statement Task	.98	< .001
Overall	.98	< .001
Congruent	.85	< .001
Incongruent	.99	.03

Figure H1

Mean Comparison of Accuracy between Congruent and Incongruent Trials for Naïve Thinking Tasks (the United Kingdom)



Notes. In accuracy analysis, significant differences between congruent and incongruent trials were observed in the Animism task ($t = 188.47, p < .001$), the Buoyancy task ($t = -15.99, p < .001$) and the Science Statement task ($t = 46.46, p < .001$); Thanks to Clemens Brunner for sharing the R script used to create this figure in Stricker et al. (2021).

Table H4*Multicollinearity test for Accuracy (the United Kingdom)*

	Overall		Congruent		Incongruent	
	<i>Tolerance</i>	<i>VIF</i>	<i>Tolerance</i>	<i>VIF</i>	<i>Tolerance</i>	<i>VIF</i>
Regression on Animism						
Science Education	.90	1.11	.90	1.11	.90	1.11
General Cognitive Ability	.72	1.39	.72	1.39	.72	1.39
Stop Signal Task	.79	1.26	.79	1.26	.79	1.26
Figure Matching Task	.78	1.28	.78	1.28	.78	1.28
Fish Flanker Task	.97	1.03	.97	1.03	.97	1.03
Corsi Block Task	.78	1.28	.78	1.28	.78	1.28
Regression on Buoyancy						
Science Education	.90	1.11	.90	1.11	.90	1.11
General Cognitive Ability	.72	1.39	.72	1.39	.72	1.39
Stop Signal Task	.79	1.26	.79	1.26	.79	1.26
Figure Matching Task	.78	1.28	.78	1.28	.78	1.28
Fish Flanker Task	.97	1.03	.97	1.03	.97	1.03
Corsi Block Task	.78	1.28	.78	1.28	.78	1.28
Regression on Science Statement						
Science Education	.90	1.11	.90	1.11	.90	1.11
General Cognitive Ability	.72	1.39	.72	1.39	.72	1.39
Stop Signal Task	.79	1.26	.79	1.26	.79	1.26
Figure Matching Task	.78	1.28	.78	1.28	.78	1.28
Fish Flanker Task	.97	1.03	.97	1.03	.97	1.03
Corsi Block Task	.78	1.28	.78	1.28	.78	1.28

Table H5*Homoscedasticity test for Accuracy (the United Kingdom)*

	<i>GQ</i>	<i>p</i>
Regression on Animism for Overall trials	1.25	.11
Regression on Animism Congruent trials	1.82	.00
Regression on Animism Incongruent trials	.97	.57
Regression on Buoyancy Overall trials	.54	1.00
Regression on Buoyancy Congruent trials	.67	.99
Regression on Buoyancy Incongruent trials	.89	.75
Regression on Science Statement Overall trials	1.04	.41
Regression on Science Statement Congruent trials	1.54	.01
Regression on Science Statement Incongruent trials	.76	.94

Notes. Significant GQ value indicated non homoscedasticity

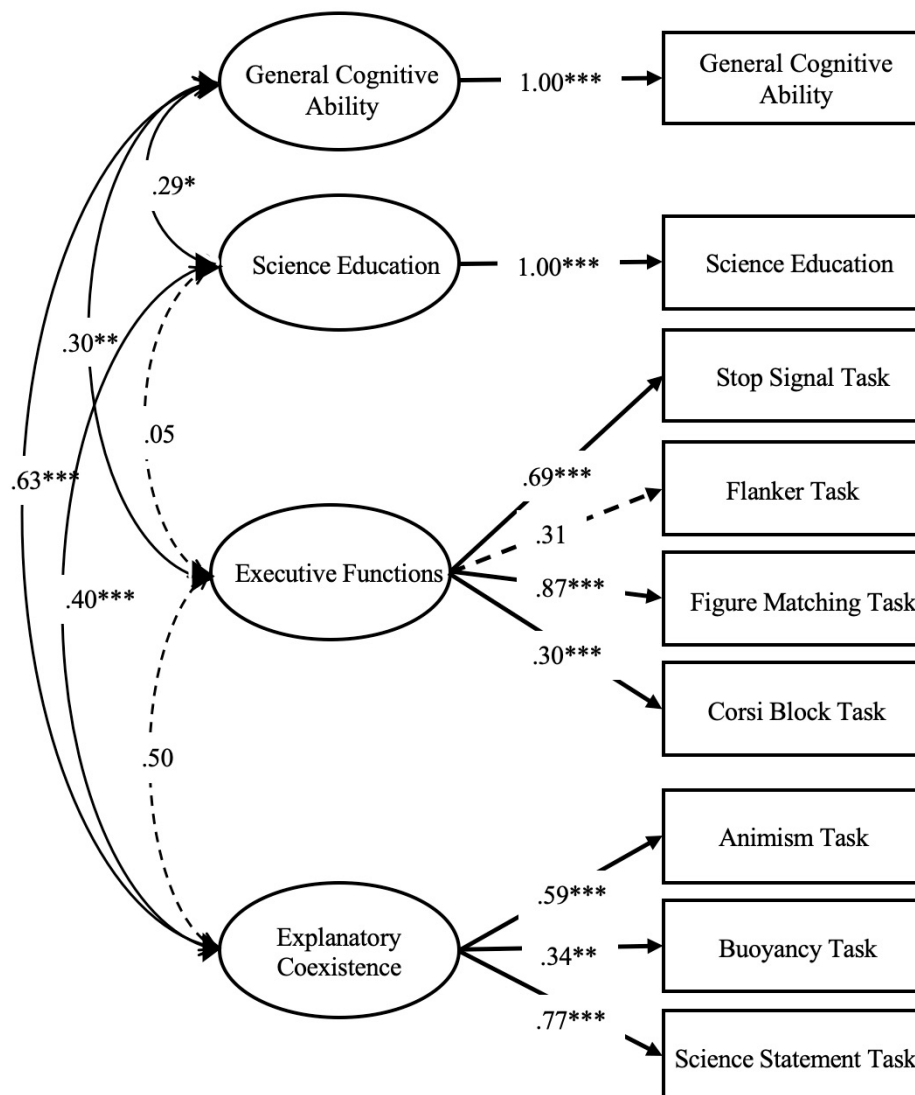
Table H6*Accuracy Confirmatory Factor Analysis Model Fit Indices (the United Kingdom)*

	$\chi^2(df)$	<i>CFI</i>	<i>TLI</i>	<i>AIC</i>	<i>BIC</i>	<i>RMSEA</i>	<i>SRMR</i>
Overall	64.45***	.87	.80	5455.37	5565.99	.10	.08
Congruent	69.93***	.86	.78	5428.65	5539.27	.10	.08
Incongruent	64.73***	.81	.71	6417.29	6528.02	.09	.07

Notes. Significant χ^2 was presented with *; * $p < .05$; ** $p < .01$; *** $p < .001$; χ^2 = Chi-square; df = degree of freedom; *CFI* = Comparative Fit Indices; *TLI* = Tucker and Lewis' Index; *AIC* = Akaike Information Criterion; *BIC* = Bayesian Information Criterion; *RMSEA* = Root Mean Square of Approximation; *SRMR* = Standardized Root Mean Square; Robust *CFI*; *TLI*; *RMSEA* and *SRMR* were reported in the Table.

Figure H2

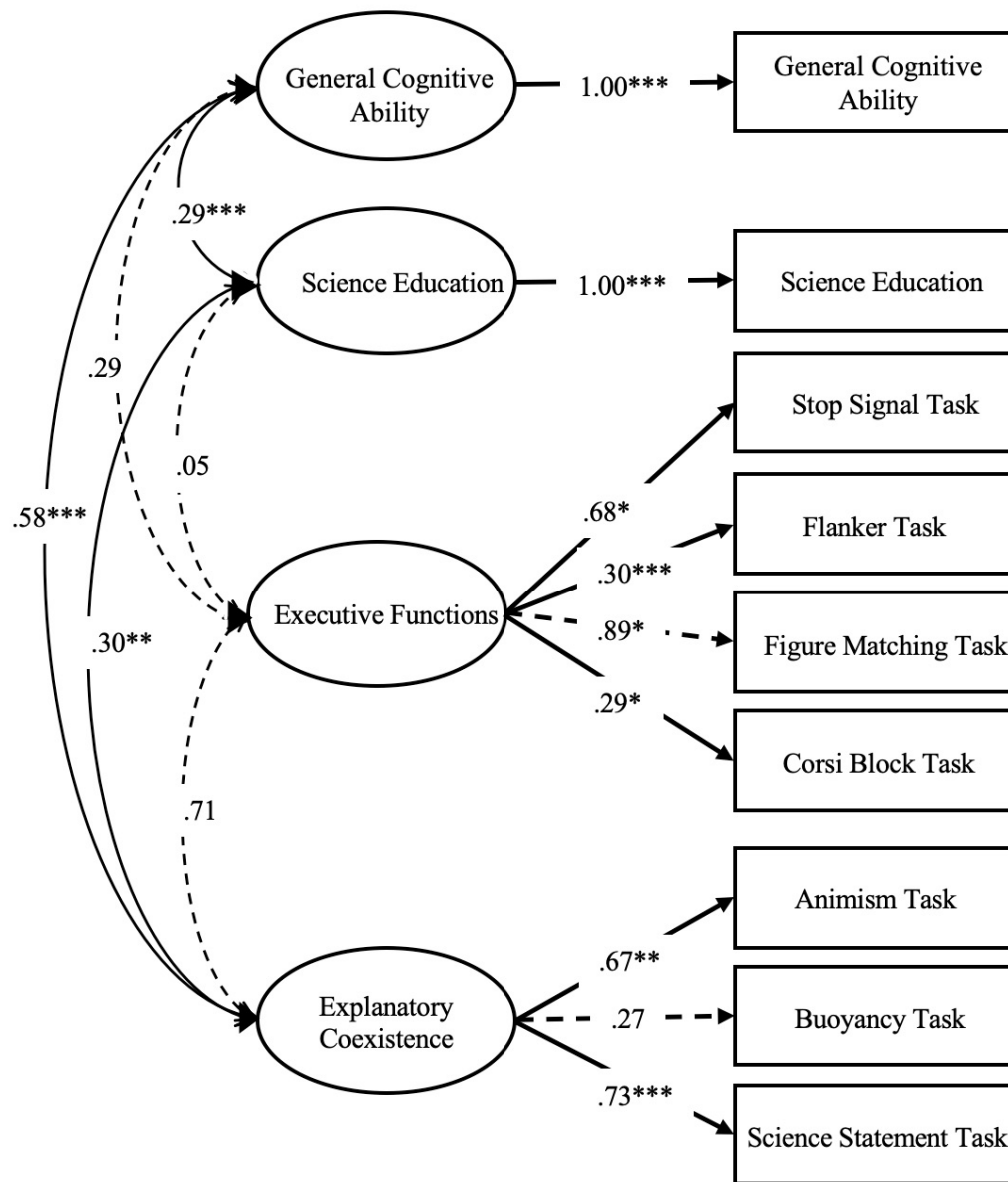
Accuracy for Overall Trials (the United Kingdom)



Notes. Standardised Path Coefficients were reported. * $p < .05$; ** $p < .01$; *** $p < .001$.

Figure H3

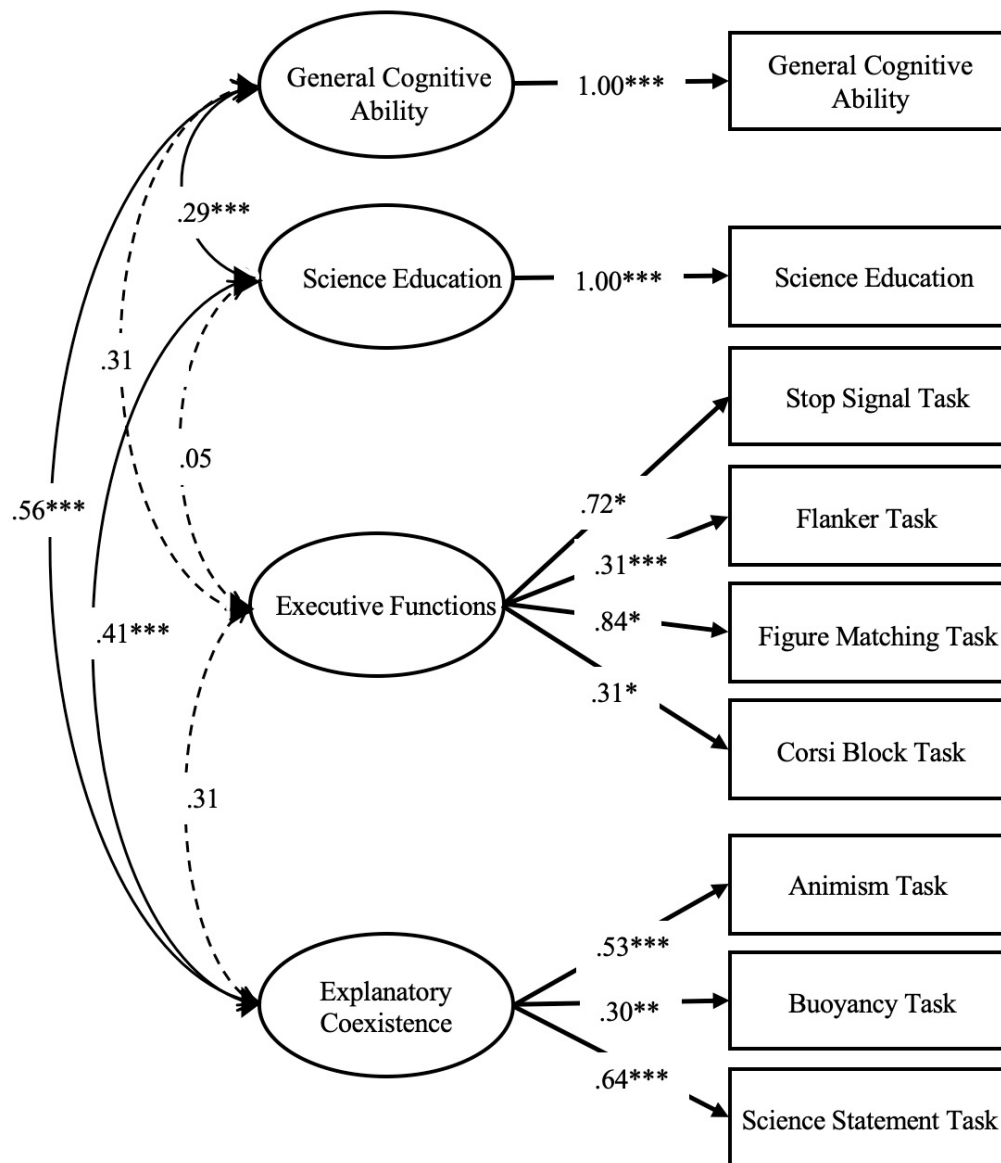
Accuracy for Congruent Trials (the United Kingdom)



Notes. Standardised Path Coefficients were reported. * $p < .05$; ** $p < .01$; *** $p < .001$.

Figure H4

Accuracy for Incongruent Trials (the United Kingdom)



Notes. Standardised Path Coefficients were reported. * $p < .05$; ** $p < .01$; *** $p < .001$.

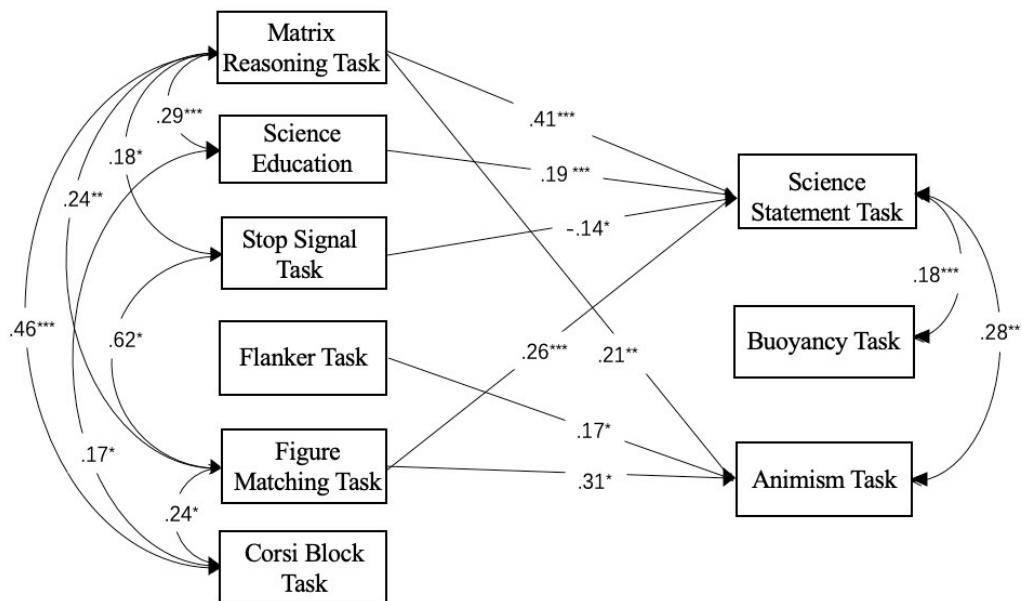
Table H7*Correlation Analysis on Accuracy (the United Kingdom)*

Variables	1	2	3	4	5	6	7	8
<i>Overall Trials</i>								
1. Science Education	-							
2. Matrix Reasoning	.29^{***}							
3. Stop Signal	-.03	.18^{**}	-					
4. Figure Matching	.03	.24^{***}	.62^{***}	-				
5. Flanker	.08	.10	.24^{***}	.24^{***}	-			
6. Corsi Block	.17^{**}	.46^{***}	.16^{**}	.24^{***}	.19^{**}	-		
7. Science Statement	.33^{***}	.51^{***}	.11	.30^{***}	.16[*]	.26^{***}		
8. Animism	.20^{**}	.32^{***}	.20^{***}	.36^{***}	.26^{***}	.22^{***}	.45^{***}	-
9. Buoyancy	.12	.19^{**}	.21^{***}	.14[*]	.13[*]	.14[*]	.26^{***}	.20^{**}
<i>Congruent Trials</i>								
1. Science Education	-							
2. Matrix Reasoning	.29^{***}	-						
3. Stop Signal	-.03	.18^{***}	-					
4. Figure Matching	.03	.24^{***}	.62^{***}	-				
5. Flanker	.08	.10	.24^{***}	.24^{***}	-			
6. Corsi Block	.17^{**}	.46^{***}	.16^{**}	.24^{***}	.19^{**}	-		
7. Science Statement	.27^{***}	.49^{***}	.24^{***}	.39^{***}	.14[*]	.29^{***}		
8. Animism	.13[*]	.28^{***}	.31^{***}	.54^{***}	.27^{***}	.24^{***}	.49^{***}	-
9. Buoyancy	.12	.20^{**}	.20^{***}	.12	.15[*]	.17^{**}	.22^{***}	.13[*]
<i>Incongruent Trials</i>								
1. Science Education	-							
2. Matrix Reasoning	.29^{***}	-						
3. Stop Signal	-.03	.18^{**}	-					
4. Figure Matching	.03	.24^{***}	.62^{***}	-				
5. Flanker	.08	.10	.24^{***}	.24^{***}	-			
6. Corsi Block	.17^{**}	.46^{***}	.16^{**}	.24^{***}	.19^{**}	-		
7. Science Statement	.28^{***}	.37^{***}	-.03	.13[*]	.13[*]	.17^{**}		
8. Animism	.20^{**}	.29^{**}	.11	.21^{***}	.21^{***}	.17^{**}	.34^{***}	-
9. Buoyancy	.10	.15[*]	.16^{**}	.11	.08	.09	.20^{**}	.17^{**}

Notes. ^{*} $p < .05$; ^{**} $p < .01$; ^{***} $p < .001$; Correlations $< .30$ are small effect size, between $.30$ and $.50$ are medium effect size, and $> .50$ are large effect size.

Figure H5

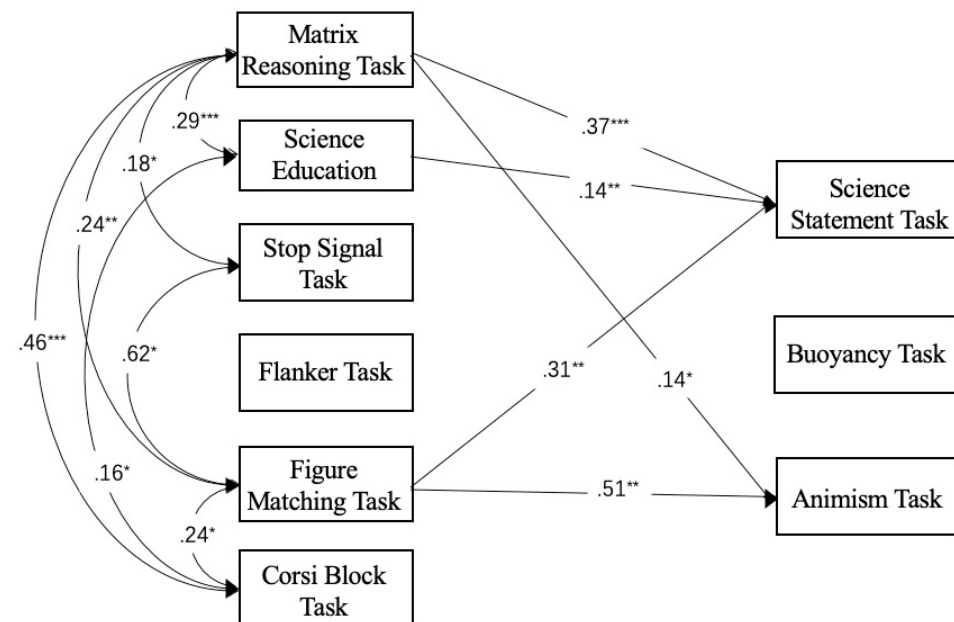
Path Analysis for Overall trials on Accuracy (the United Kingdom)



Notes. Standardised Path Coefficients were reported. * $p < .05$; ** $p < .01$; *** $p < .001$.

Figure H6

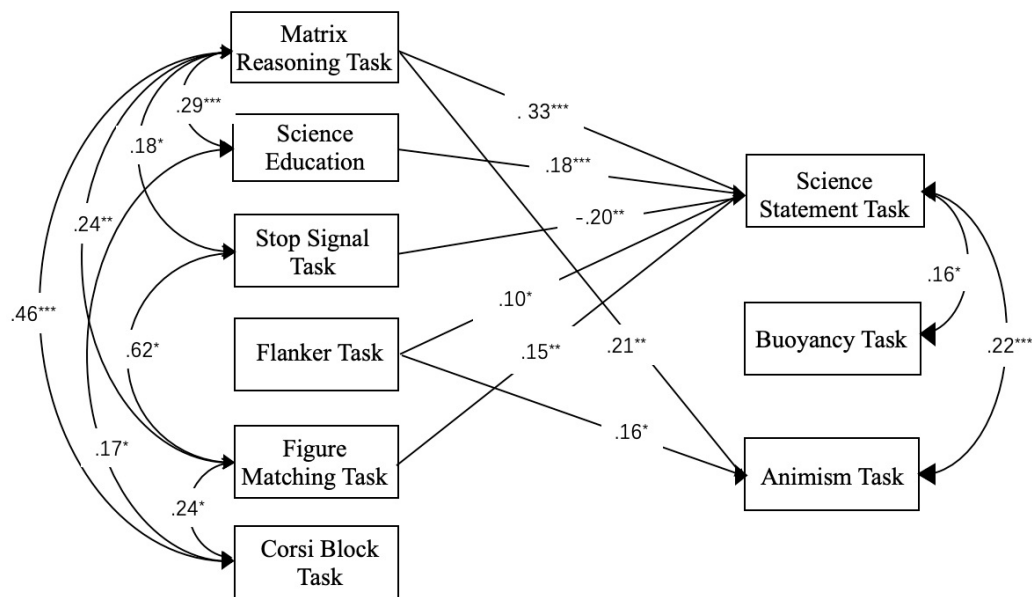
Path Analysis for Congruent trials on Accuracy (the United Kingdom)



Notes. Standardised Path Coefficients were reported. * $p < .05$; ** $p < .01$; *** $p < .001$.

Figure H7

Path Analysis for Incongruent trials on Accuracy (the United Kingdom)



Notes. Standardised Path Coefficients were reported. * $p < .05$; ** $p < .01$; *** $p < .001$.

Table H8

Path Analysis Model Fit Indices Summary on Accuracy (the United Kingdom)

Trials	$\chi^2(df)$	CFI	TLI	AIC	BIC	RMSEA	SRMR
Overall	.02(1)	1.00	1.09	5407.17	5596.29	.00	.00
Congruent	.57(1)	1.00	1.04	5368.09	5557.21	.00	.01
Incongruent	.03(1)	1.00	1.12	5504.84	5693.96	.00	.00

Notes. Significant χ^2 was presented with *; * $p < .05$; ** $p < .01$; *** $p < .001$; χ^2 = Chi-square; df = degree of freedom; CFI = Comparative Fit Indices; TLI = Tucker and Lewis' Index; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; RMSEA = Root Mean Square of Approximation; SRMR = Standardized Root Mean Square; Robust CFI; TLI; RMSEA and SRMR were reported in the Table.

Appendix I: Response Time Analysis (the United Kingdom)

Response Time in Appendix I refers to response time for correct trial

Table I1

Descriptive statistics for Response Time (the United Kingdom)

Variables	<i>M</i>	<i>SD</i>	<i>Skewness</i>	<i>Kurtosis</i>
General Cognitive Ability	7.88	3.70	.11	-.55
Executive Function skills				
Stop Signal Task	.69	.28	1.04	1.55
Figure Matching Task	.89	.24	1.44	3.40
Fish Flanker Task	.47	.12	2.63	1.08
Corsi Block Task	2.67	1.29	1.02	12.27
Explanatory Coexistence				
Buoyancy Task				
Overall	.70	.24	1.97	5.39
Congruent	.67	.21	1.78	5.51
Incongruent	.73	.31	2.91	12.75
Animism Task				
Overall	.65	.19	3.29	17.57
Congruent	.63	.18	2.85	11.91
Incongruent	.66	.21	4.00	25.42
Science Statement Task				
Overall	2.58	1.00	1.59	4.24
Congruent	2.44	.97	1.73	4.90
Incongruent	2.83	1.58	7.16	81.93

Table I2*Age adjustment test for Response Time (the United Kingdom)*

Variables	Linear	Quadratic	Cubic
General Cognitive Ability	.00	.00	-.01
Executive Function skills			
Stop Signal Task	.00	-.01	-.01
Figure Matching Task	.04***	.05***	.07***
Fish Flanker Task	.03**	.02*	.04**
Corsi Block Task	.04***	.06***	.06***
Explanatory Coexistence			
Buoyancy Task			
Overall	.00	.00	< .001
Congruent	.00	.00	< .001
Incongruent	.00	.00	< .001
Animism Task			
Overall	.00	.01	< .001
Congruent	.00	.00	< .001
Incongruent	.00	.00	< .001
Science Statement Task			
Overall	.00	.00	< .001
Congruent	-.01	-.01	< .001
Incongruent	-.01	-.01	< .001

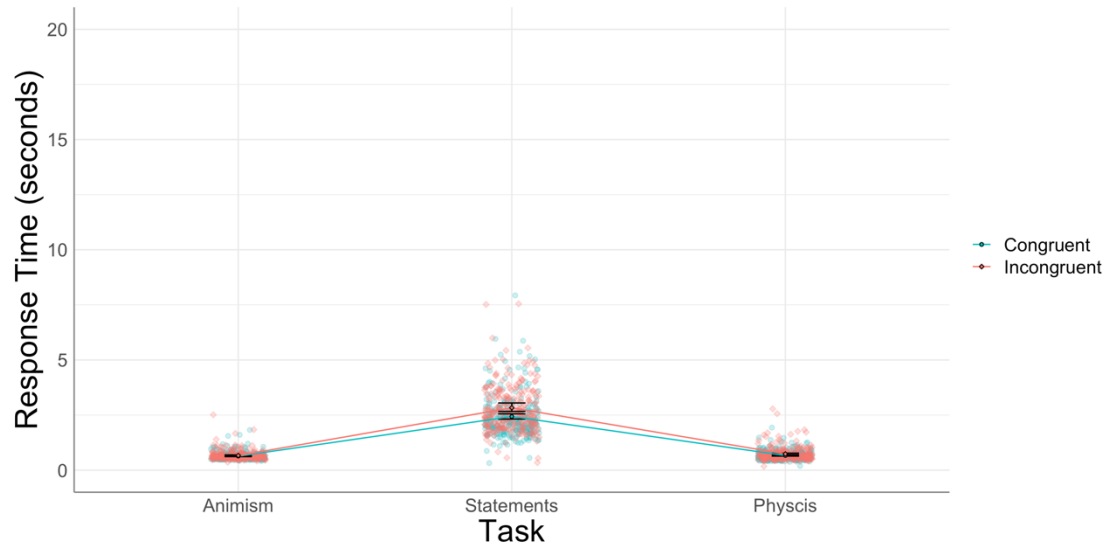
Notes. * $p < .05$; ** $p < .01$; *** $p < .001$.

Table I3*Shapiro-Wilk test for Response Time (the United Kingdom)*

Variables	<i>Shapiro-Wilk</i>	<i>p</i>
Science Education	.99	.04
General Cognitive Ability	.93	< .001
Executive Function skills		
Stop Signal Task	.93	< .001
Figure Matching Task	.93	< .001
Fish Flanker Task	.77	< .001
Corsi Block Task	.68	< .001
Explanatory Coexistence		
Buoyancy Task		
Overall	.84	< .001
Congruent	.87	< .001
Incongruent	.75	< .001
Animism Task		
Overall	.72	< .001
Congruent	.75	< .001
Incongruent	.67	< .001
Science Statement Task		
Overall	.89	< .001
Congruent	.87	< .001
Incongruent	.58	< .001

Figure I1

Mean Comparison of Response Time between Congruent and Incongruent Trials for Naïve Thinking Tasks (the United Kingdom)



Notes. In response time analysis, significant differences between congruent and incongruent trials were observed in the Animism task ($t = 8.94$, $p < .001$) and the Buoyancy task ($t = 6.19$, $p < .001$), but not in the Science Statement task ($t = 1.71$, $p = .089$); Response Time here refers to response time for correct trial; Thanks to Clemens Brunner for sharing the R script used to create this figure in Stricker et al. (2021).

Table I4*Multicollinearity test for Response Time (the United Kingdom)*

	Overall		Congruent		Incongruent	
	Tolerance	VIF	Tolerance	VIF	Tolerance	VIF
Regression on Animism						
Science Education	.97	1.03	.97	1.03	.97	1.03
General Cognitive Ability	.89	1.12	.89	1.12	.89	1.12
Stop Signal Task	.88	1.14	.88	1.14	.88	1.14
Figure Matching Task	.67	1.50	.67	1.50	.67	1.50
Fish Flanker Task	.60	1.68	.60	1.68	.60	1.68
Corsi Block Task	.88	1.14	.88	1.14	.88	1.14
Regression on Buoyancy						
Science Education	.97	1.03	.97	1.03	.97	1.03
General Cognitive Ability	.89	1.12	.89	1.12	.89	1.12
Stop Signal Task	.88	1.14	.88	1.14	.88	1.14
Figure Matching Task	.67	1.50	.67	1.50	.60	1.68
Fish Flanker Task	.60	.60	.60	1.68	.67	1.50
Corsi Block Task	.88	1.14	.88	1.14	.88	1.14
Regression on Science Statement						
Science Education	.97	1.03	.97	1.03	.97	1.03
General Cognitive Ability	.89	1.12	.89	1.12	.89	1.12
Stop Signal Task	.88	1.14	.88	1.14	.88	1.14
Figure Matching Task	.67	1.50	.67	1.50	.67	1.50
Fish Flanker Task	.60	1.68	.60	1.68	.60	1.68
Corsi Block Task	.88	1.14	.88	1.14	.88	1.14

Table I5*Homoscedasticity test for Response Time (the United Kingdom)*

	<i>GQ</i>	<i>p</i>
Regression on Animism for Overall trials	1.17	.19
Regression on Animism Congruent trials	1.17	.20
Regression on Animism Incongruent trials	1.09	.32
Regression on Buoyancy Overall trials	1.18	.18
Regression on Buoyancy Congruent trials	.81	.88
Regression on Buoyancy Incongruent trials	1.00	.49
Regression on Science Statement Overall trials	.83	.84
Regression on Science Statement Congruent trials	.98	.55
Regression on Science Statement Incongruent trials	.25	1.00

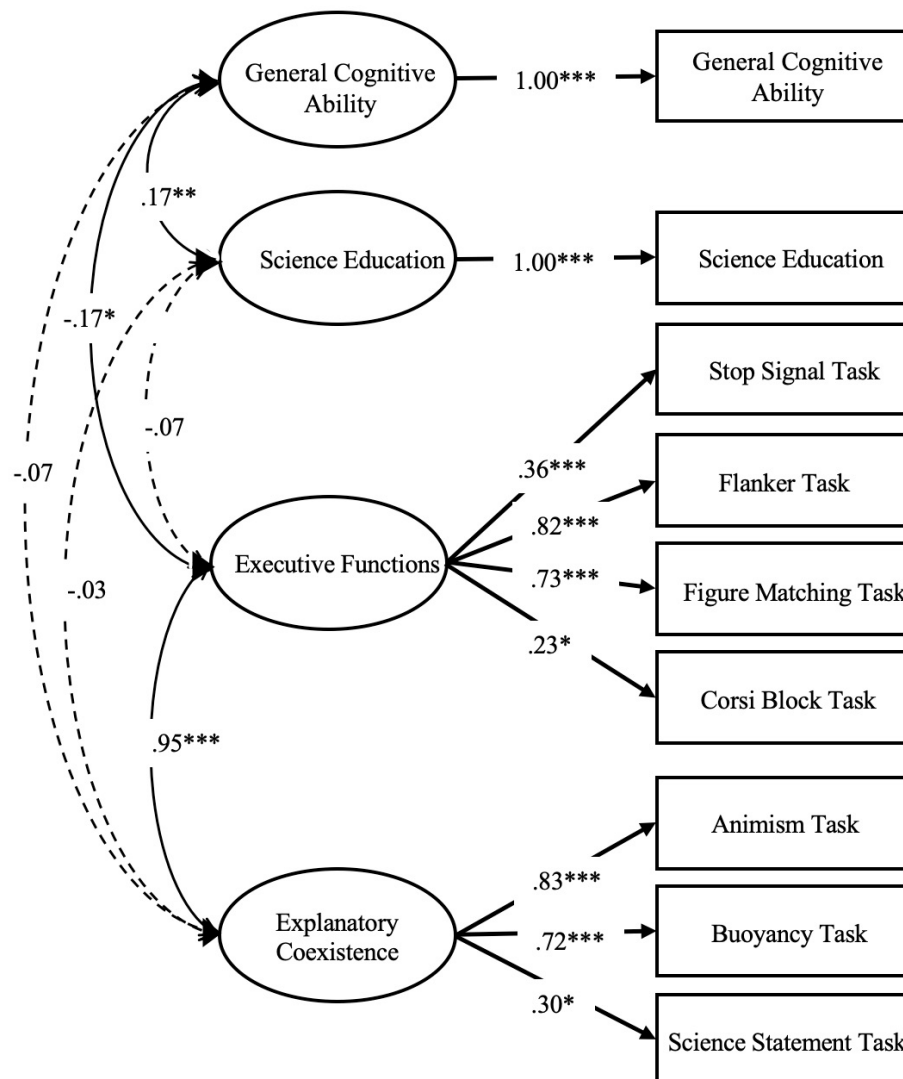
Table I6*Response Time Confirmatory Factor Analysis Model Fit Indices (the United Kingdom)*

	$\chi^2(df)$	<i>CFI</i>	<i>TLI</i>	<i>AIC</i>	<i>BIC</i>	<i>RMSEA</i>	<i>SRMR</i>
Overall	89.81***	.83	.73	6050.24	6128.23	.12	.08
Congruent	86.50***	.85	.76	6070.02	6148.02	.12	.08
Incongruent	49.88***	.90	.84	6139.28	6217.28	.09	.07

Notes. Significant χ^2 was presented with *; * $p < .05$; ** $p < .01$; *** $p < .001$; χ^2 = Chi-square; *df* = degree of freedom; *CFI* = Comparative Fit Indices; *TLI* = Tucker and Lewis' Index; *AIC* = Akaike Information Criterion; *BIC* = Bayesian Information Criterion; *RMSEA* = Root Mean Square of Approximation; *SRMR* = Standardized Root Mean Square; Robust *CFI*; *TLI*; *RMSEA* and *SRMR* were reported in the Table.

Figure I2

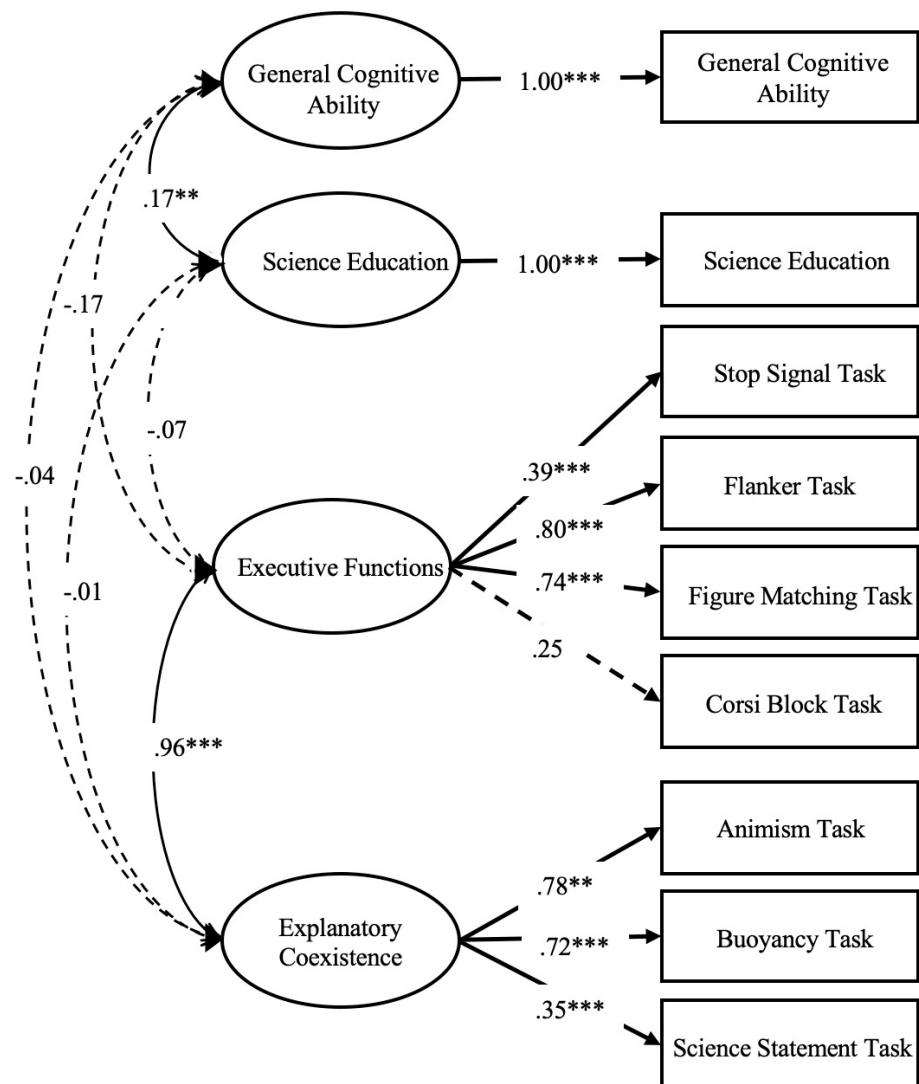
Response Time for Overall Trials (the United Kingdom)



Notes. Standardised Path Coefficients were reported. * $p < .05$; ** $p < .01$; *** $p < .001$.

Figure I3

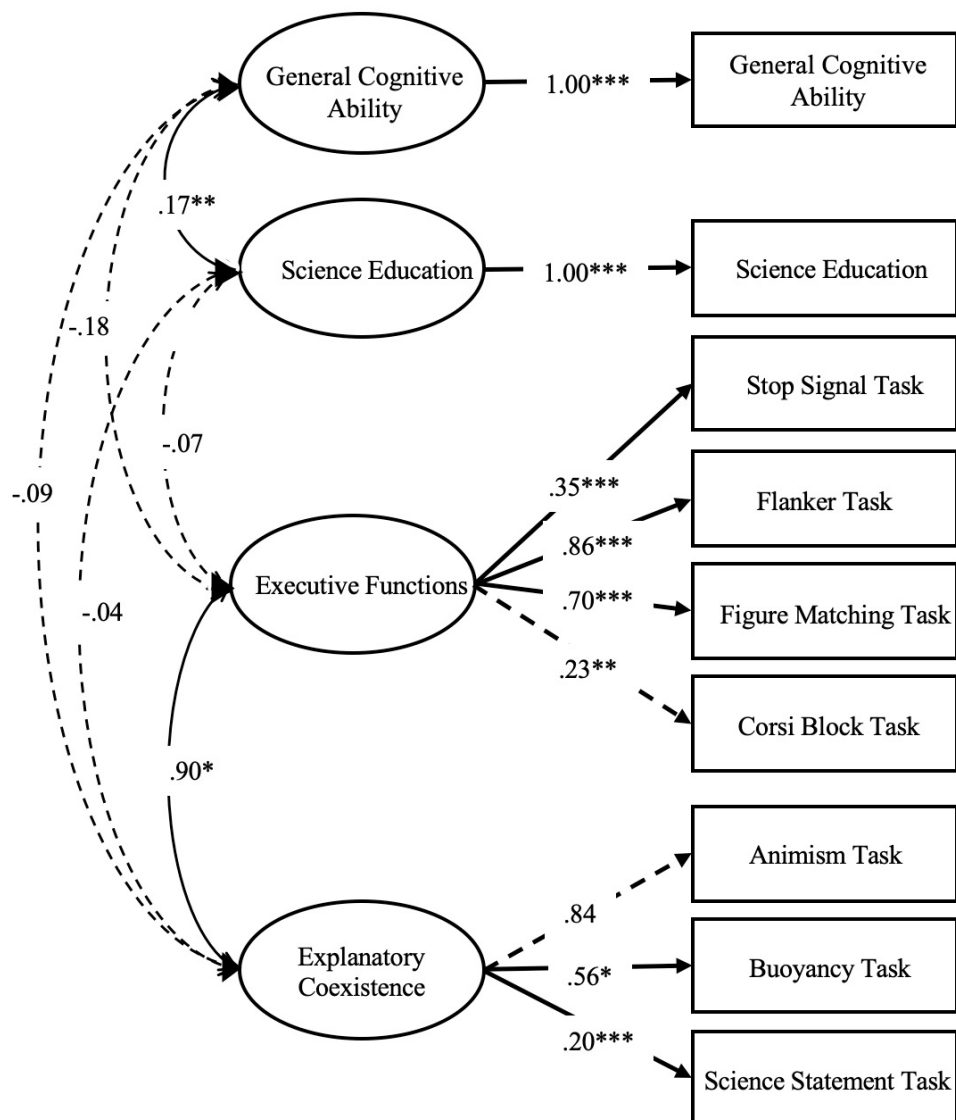
Response Time for Congruent Trials (the United Kingdom)



Notes. Standardised Path Coefficients were reported. * $p < .05$; ** $p < .01$; *** $p < .001$.

Figure I4

Response Time for Incongruent Trials (the United Kingdom)



Notes. Standardised Path Coefficients were reported. * $p < .05$; ** $p < .01$; *** $p < .001$.

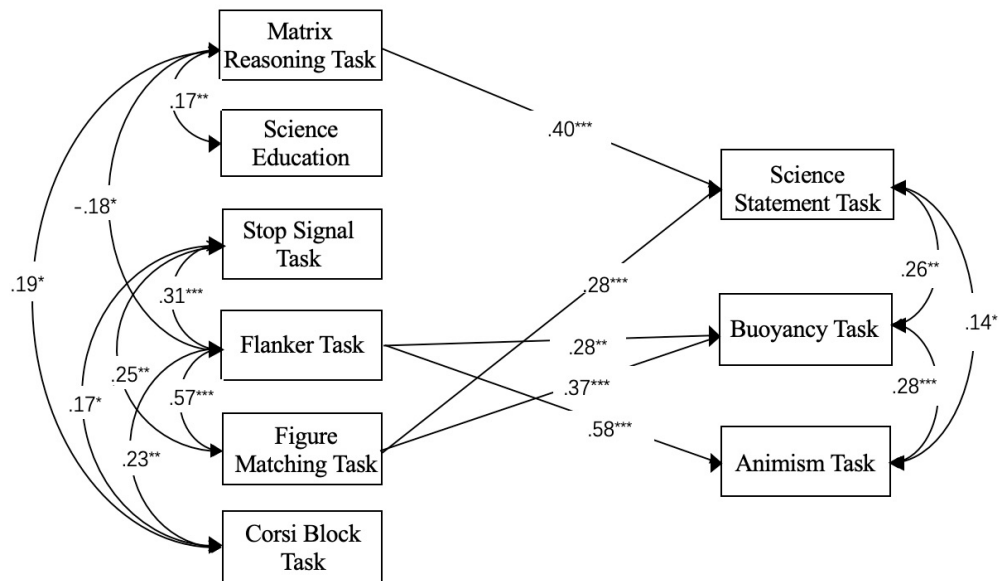
Table I7*Correlation Analysis on Response Time (the United Kingdom)*

Variables	1	2	3	4	5	6	7	8
<i>Overall Trials</i>								
1. Science Education	-							
2. Matrix Reasoning	.17**							
3. Stop Signal	-.13*	.00	-					
4. Figure Matching	-.02	-.10	.25***	-				
5. Flanker	-.06	-.18**	.31***	.57***	-			
6. Corsi Block	.00	.19**	.17**	.10	.23***	-		
7. Science Statement	.07	.36***	.13*	.26***	.12	.10		
8. Animism	.00	-.11	.24***	.50***	.68***	.18**	.20***	-
9. Buoyancy	-.09	-.06	.25***	.55***	.52***	.18**	.33***	.55***
<i>Congruent Trials</i>								
1. Science Education	-							
2. Matrix Reasoning	.17**	-						
3. Stop Signal	-.13*	.00	-					
4. Figure Matching	-.02	-.10	.25***	-				
5. Flanker	-.06	-.18**	.31***	.57***	-			
6. Corsi Block	.00	.19**	.17**	.10	.23***	-		
7. Science Statement	.07	.31***	.15*	.27***	.14*	.11		
8. Animism	.01	-.11	.26***	.47***	.64***	.19**	.20***	-
9. Buoyancy	-.05	-.03	.32***	.55***	.50***	.24***	.40***	.52***
<i>Incongruent Trials</i>								
1. Science Education	-							
2. Matrix Reasoning	.17***	-						
3. Stop Signal	-.13*	.00	-					
4. Figure Matching	-.02	-.10	.25***	-				
5. Flanker	-.06	-.18**	.31***	.57***	-			
6. Corsi Block	.00	.19**	.17**	.10	.23***	-		
7. Science Statement	.04	.28***	.09	.18**	.07	.04		
8. Animism	-.02	-.11	.21***	.48***	.67***	.15*	.18**	-
9. Buoyancy	-.10	-.03	.16**	.44***	.41***	.14*	.13*	.45***

Notes. * $p < .05$; ** $p < .01$; *** $p < .001$; Correlations $< .30$ are small effect size, between $.30$ and $.50$ are medium effect size, and $> .50$ are large effect size.

Figure I5

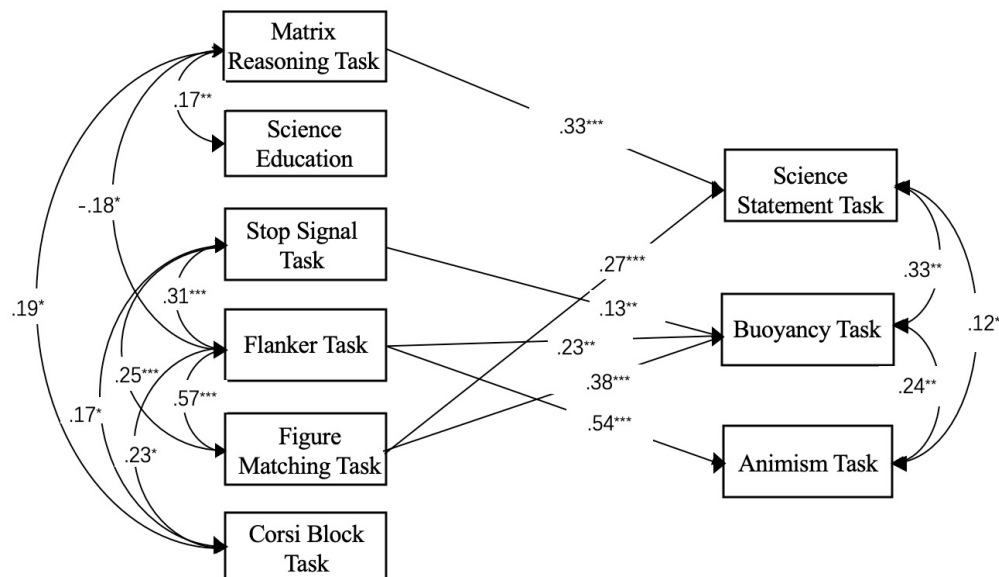
Path Analysis for Overall trials on Response Time (the United Kingdom)



Notes. Standardised Path Coefficients were reported. * $p < .05$; ** $p < .01$; *** $p < .001$.

Figure I6

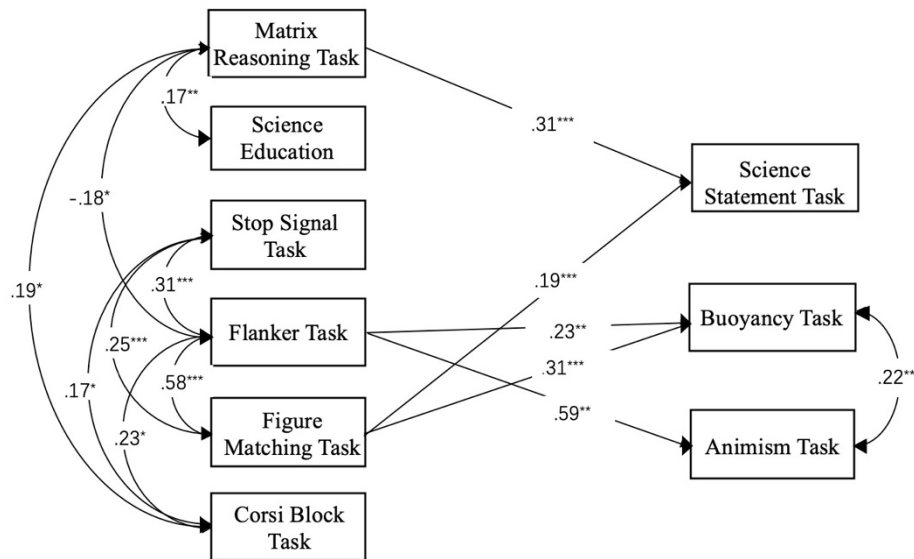
Path Analysis for Congruent trials on Response Time (the United Kingdom)



Notes. Standardised Path Coefficients were reported. * $p < .05$; ** $p < .01$; *** $p < .001$.

Figure I7

Path Analysis for Incongruent trials on Response Time (the United Kingdom)



Notes. Standardised Path Coefficients were reported. * $p < .05$; ** $p < .01$; *** $p < .001$.

Table I8

Path Analysis Model Fit Indices Summary on Response Time (the United Kingdom)

Trials	$\chi^2(df)$	CFI	TLI	AIC	BIC	RMSEA	SRMR
Overall	.12(1)	1.00	1.13	6186.77	6375.89	.00	.00
Congruent	.46(1)	1.00	1.09	6209.83	6398.96	.00	.01
Incongruent	.00(1)	1.00	1.18	6300.89	6490.02	.00	.00

Notes. Significant χ^2 was presented with *; * $p < .05$; ** $p < .01$; *** $p < .001$; χ^2 = Chi-square; df = degree of freedom; CFI = Comparative Fit Indices; TLI = Tucker and Lewis' Index; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; RMSEA = Root Mean Square of Approximation; SRMR = Standardized Root Mean Square; Robust CFI; TLI; RMSEA and SRMR were reported in the Table.

Appendix J: Efficiency Analysis (the United Kingdom)

Table J1

Descriptive statistics for Efficiency (the United Kingdom)

Variables	<i>M</i>	<i>SD</i>	<i>Skewness</i>	<i>Kurtosis</i>
General Cognitive Ability	.09	.08	6.61	56.25
Executive Function				
Stop Signal Task	1.49	.51	.38	-.70
Figure Matching Task	1.14	.26	-.09	-.22
Fish Flanker Task	2.14	.41	-1.22	2.30
Corsi Block Task	.24	.09	.67	1.41
Explanatory Coexistence				
Buoyancy Task				
Overall	1.47	.42	-.26	-.13
Congruent	1.57	.44	-.32	.04
Incongruent	1.39	.45	-.51	.24
Animism Task				
Overall	1.52	.31	-.72	.58
Congruent	.33	1.65	-.68	.40
Incongruent	1.45	.30	-.68	.66
Science Statement Task				
Overall	.32	.12	2.15	12.25
Congruent	.41	.14	1.41	6.63
Incongruent	.24	.13	4.42	3.32

Table J2*Age adjustment test for Efficiency (the United Kingdom)*

Variables	Linear	Quadratic	Cubic
General Cognitive Ability	.01	.00	.00
Executive Function skills			
Stop Signal Task	.00	-.01	-.01
Figure Matching Task	.02*	.02*	.03*
Fish Flanker Task	.02**	.02*	.03*
Corsi Block Task	.04***	.04**	.04**
Explanatory Coexistence			
Buoyancy Task			
Overall	.00	.00	-.01
Congruent	.00	.00	.00
Incongruent	.00	.00	-.01
Animism Task			
Overall	.01*	.00	.01
Congruent	.00	.00	.00
Incongruent	.01	.01	.01
Science Statement Task			
Overall	.00	.00	.00
Congruent	-.01	-.01	-.01
Incongruent	-.01	-.01	-.01

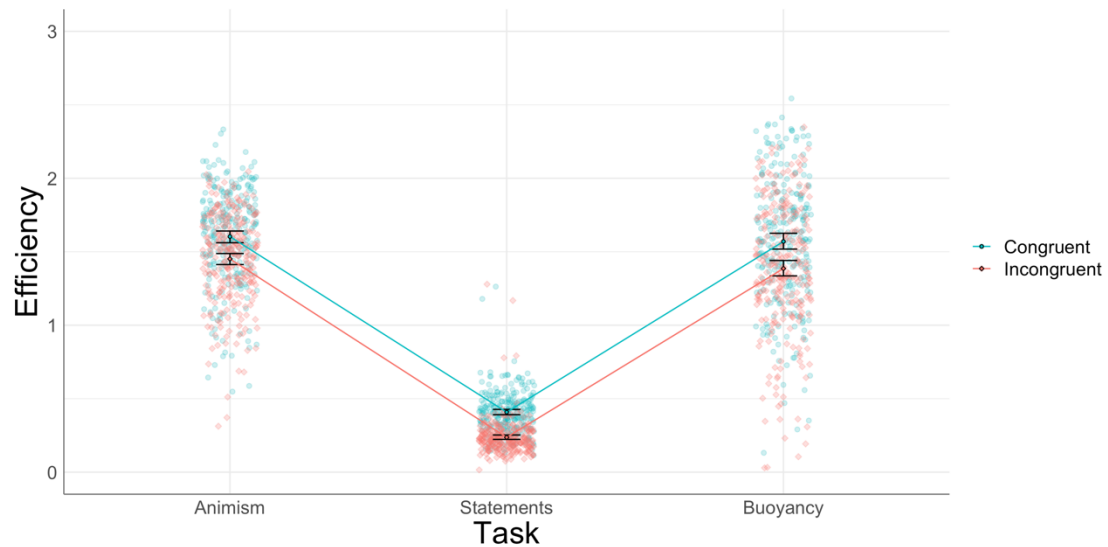
Notes. * $p < .05$; ** $p < .01$; *** $p < .001$.

Table J3*Shapiro-Wilk test for Efficiency (the United Kingdom)*

Variables	<i>Shapiro-Wilk</i>	<i>p</i>
General Cognitive Ability	.45	< .001
Executive Function		
Stop Signal Task	.97	< .001
Figure Matching Task	1.00	.61
Fish Flanker Task	.93	< .001
Corsi Block Task	.97	< .001
Explanatory Coexistence		
Buoyancy Task		
Overall	.99	.05
Congruent	.99	.03
Incongruent	.98	< .001
Animism Task		
Overall	.97	< .001
Congruent	.97	< .001
Incongruent	.97	< .001
Science Statement Task		
Overall	.87	< .001
Congruent	.92	< .001
Incongruent	.66	< .001

Figure J1

Mean Comparison of Efficiency between Congruent and Incongruent Trials for Naïve Thinking Tasks (the United Kingdom)



Notes. In efficiency analysis, significant differences between congruent and incongruent trials were observed in the Animism task ($t = 12.98, p < .001$), the Buoyancy task ($t = -9.52, p < .001$), and the Science Statement task ($t = 25.77, p < .001$); Thanks to Clemens Brunner for sharing the R script used to create this figure in Stricker et al. (2021).

Table J4*Multicollinearity test for Efficiency (the United Kingdom)*

	Overall		Congruent		Incongruent	
	<i>Tolerance</i>	<i>VIF</i>	<i>Tolerance</i>	<i>VIF</i>	<i>Tolerance</i>	<i>VIF</i>
Regression on Animism						
Science Education	.99	1.03	.97	1.03	.97	1.03
General Cognitive Ability	.99	1.01	.99	1.01	.99	1.01
Stop Signal Task	.83	1.20	.83	1.20	.83	1.20
Figure Matching Task	.68	1.48	.68	1.48	.68	1.48
Fish Flanker Task	.60	1.66	.60	1.66	.60	1.66
Corsi Block Task	.86	1.16	.86	1.16	.86	1.16
Regression on Buoyancy						
Science Education	.97	1.03	.97	1.03	.97	1.03
General Cognitive Ability	.99	1.01	.99	1.01	.99	1.01
Stop Signal Task	.83	1.20	.83	1.20	.83	1.20
Figure Matching Task	.68	1.48	.68	1.48	.68	1.48
Fish Flanker Task	.60	1.66	.60	1.66	.60	1.66
Corsi Block Task	.86	1.16	.86	1.16	.86	1.16
Regression on Science Statement						
Science Education	.97	1.03	.97	1.03	.97	1.03
General Cognitive Ability	.99	1.01	.99	1.01	.99	1.01
Stop Signal Task	.83	1.20	.83	1.20	.83	1.20
Figure Matching Task	.68	1.48	.68	1.48	.68	1.48
Fish Flanker Task	.60	1.66	.60	1.66	.60	1.66
Corsi Block Task	.86	1.16	.86	1.16	.86	1.16

Table J5*Homoscedasticity test for Efficiency (the United Kingdom)*

	<i>GQ</i>	<i>p</i>
Regression on Animism for Overall trials	1.14	.23
Regression on Animism Congruent trials	1.21	.15
Regression on Animism Incongruent trials	1.07	.35
Regression on Buoyancy Overall trials	.95	.61
Regression on Buoyancy Congruent trials	.87	.77
Regression on Buoyancy Incongruent trials	1.12	.27
Regression on Science Statement Overall trials	.74	.95
Regression on Science Statement Congruent trials	1.13	.25
Regression on Science Statement Incongruent trials	.86	.80

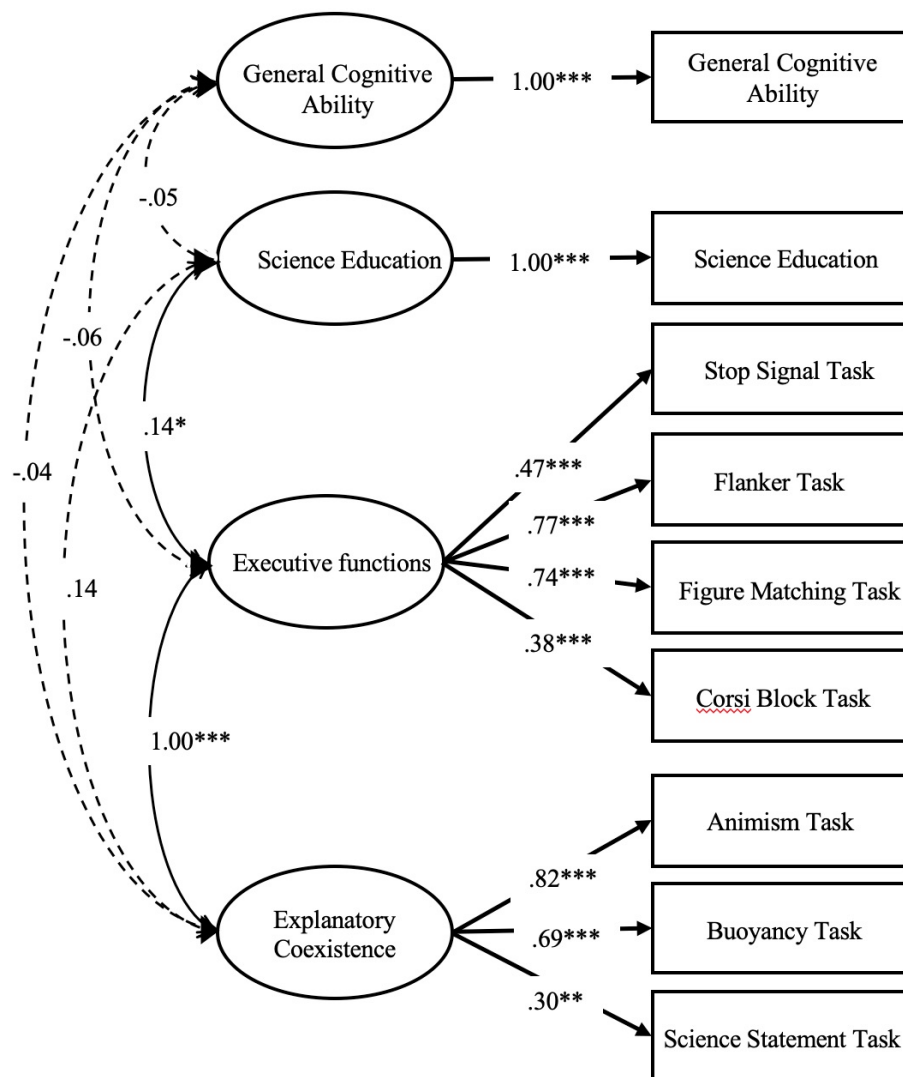
Table J6*Efficiency for Confirmatory Factor Analysis Model Fit Indices (the United Kingdom)*

	$\chi^2(df)$	<i>CFI</i>	<i>TLI</i>	<i>AIC</i>	<i>BIC</i>	<i>RMSEA</i>	<i>SRMR</i>
Overall	45.14***	.95	.91	6183.54	6294.16	.07	.05
Congruent	36.01*	.97	.95	6016.49	6094.49	.05	.05
Incongruent	31.82	.97	.96	6007.48	6085.47	.04	.04

Notes. Significant χ^2 was presented with *; * $p < .05$; ** $p < .01$; *** $p < .001$; χ^2 = Chi-square; *df* = degree of freedom; *CFI* = Comparative Fit Indices; *TLI* = Tucker and Lewis' Index; *AIC* = Akaike Information Criterion; *BIC* = Bayesian Information Criterion; *RMSEA* = Root Mean Square of Approximation; *SRMR* = Standardized Root Mean Square; Robust *CFI*; *TLI*; *RMSEA* and *SRMR* were reported in the Table.

Figure J2

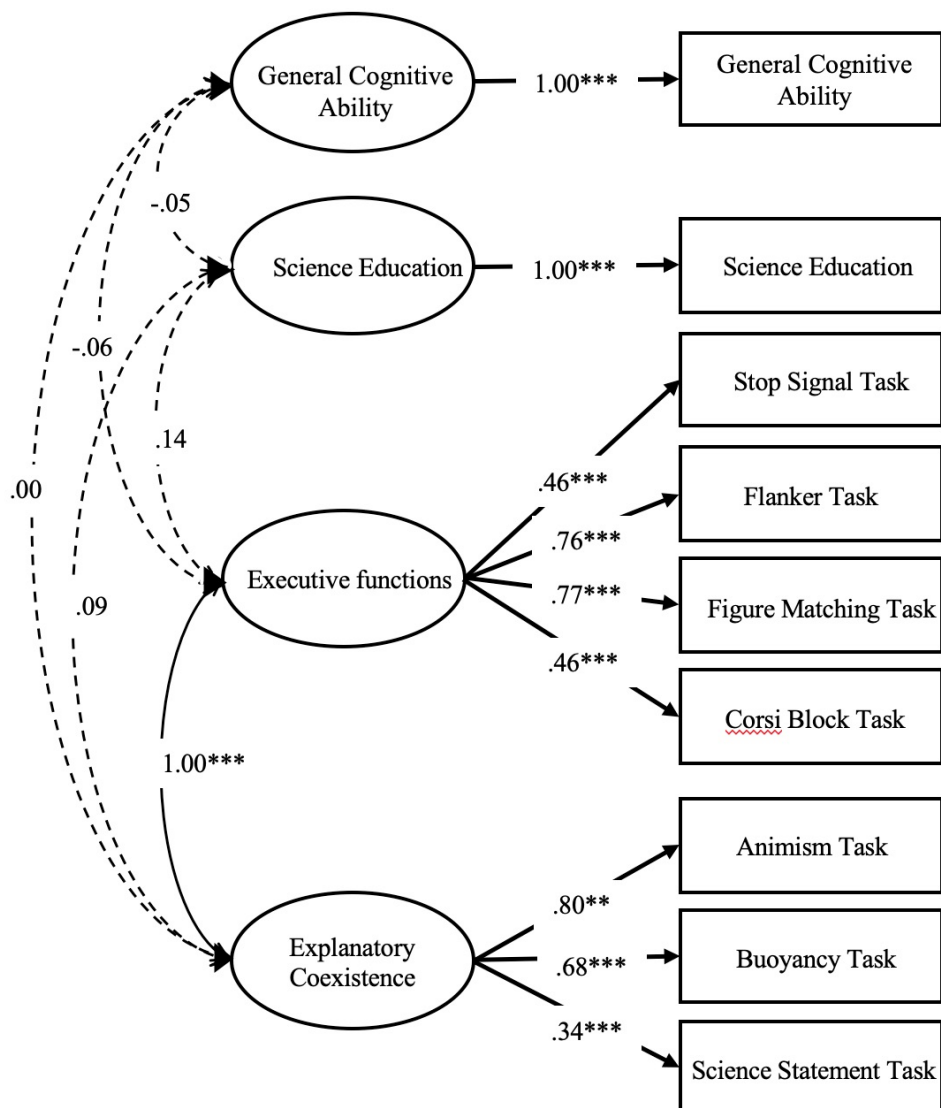
Efficiency for Overall Trials (the United Kingdom)



Notes. Standardised Path Coefficients were reported. * $p < .05$; ** $p < .01$; *** $p < .001$.

Figure J3

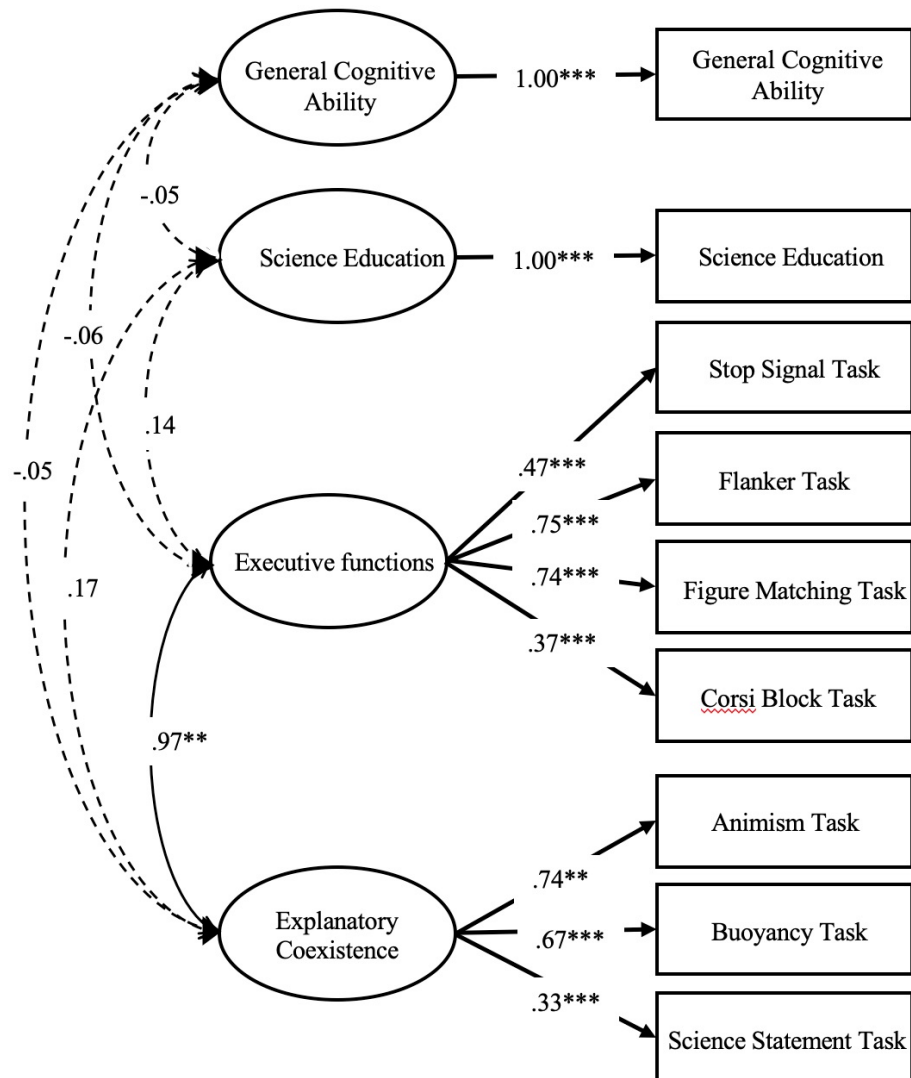
Efficiency for Congruent Trials (the United Kingdom)



Notes. Standardised Path Coefficients were reported. * $p < .05$; ** $p < .01$; *** $p < .001$.

Figure J4

Efficiency for Incongruent Trials (the United Kingdom)



Notes. Standardised Path Coefficients were reported. * $p < .05$; ** $p < .01$; *** $p < .001$.

Appendix K: Accuracy Analysis (China)

Table K1

Descriptive statistics for Accuracy (China)

Variables	<i>M</i>	<i>SD</i>	<i>Skewness</i>	<i>Kurtosis</i>
General Cognitive Ability	0.71	0.19	-0.88	-0.26
Executive Function				
Stop Signal Task	0.90	0.06	-1.63	6.66
Figure Matching Task	0.97	0.04	-5.47	42.86
Fish Flanker Task	0.98	0.02	-2.31	6.94
Corsi Block Task	0.61	0.12	-0.65	3.66
Explanatory Coexistence				
Buoyancy Task				
Overall	0.96	0.09	-4.83	29.54
Congruent	0.96	0.08	-4.46	27.02
Incongruent	0.94	0.12	-4.48	25.08
Animism Task				
Overall	0.96	0.04	-1.08	4.07
Congruent	0.97	0.03	-2.84	13.44
Incongruent	0.94	0.06	-2.32	9.54
Science Statement Task				
Overall	0.77	0.07	-0.34	0.55
Congruent	0.87	0.07	-1.18	3.26
Incongruent	0.67	0.10	0.03	-0.84

Table K2*Age adjustment test for Accuracy (China)*

Variables	Linear	Quadratic	Cubic
General Cognitive Ability	.01*	.01	.01
Executive Function skills			
Stop Signal Task	-.00	-.01	-.01
Figure Matching Task	-.00	-.00	-.01
Fish Flanker Task	.03**	.02*	.03*
Corsi Block Task	-.00	-.00	.02
Explanatory Coexistence			
Buoyancy Task			
Overall	-.00	-.00	-.01
Congruent	-.00	-.00	.01
Incongruent	-.00	.01	-.01
Animism Task			
Overall	-.00	.00	-.01
Congruent	.02*	.02	.01
Incongruent	-.00	-.00	-.01
Science Statement Task			
Overall	.01	.01	.01
Congruent	.04**	.03**	.03*
Incongruent	-.00	-.00	.01

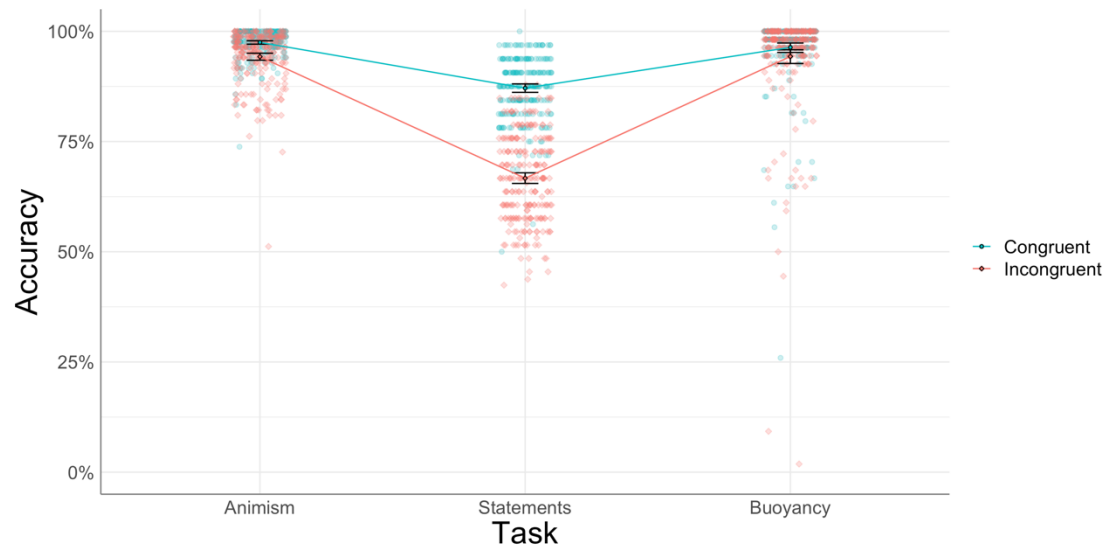
Notes. * $p < .05$; ** $p < .01$; *** $p < .001$.

Table K3*Shapiro-Wilk test for Accuracy (China)*

Variables	<i>Shapiro-Wilk</i>	<i>p</i>
General Cognitive Ability	.90	< .001
Executive Function		
Stop Signal Task	.90	< .001
Figure Matching Task	.56	< .001
Fish Flanker Task	.80	< .001
Corsi Block Task	.94	< .001
Explanatory Coexistence		
Buoyancy Task		
Overall	.44	< .001
Congruent	.47	< .001
Incongruent	.48	< .001
Animism Task		
Overall	.82	< .001
Congruent	.76	< .001
Incongruent	.78	< .001
Science Statement Task		
Overall	.98	.004
Congruent	.94	< .001
Incongruent	.68	.002

Figure K1

Mean Comparison of Accuracy between Congruent and Incongruent Trials for Naïve Thinking Tasks (China)



Notes. In accuracy analysis, significant differences between congruent and incongruent trials were observed in the Animism task ($t = -5.80, p < .001$) and the Buoyancy task ($t = -8.59, p < .001$), but not in the Science Statement task ($t = -1.23, p = .106$); Thanks to Clemens Brunner for sharing the R script used to create this figure in Stricker et al. (2021).

Table K4*Multicollinearity test for Accuracy (China)*

	Overall		Congruent		Incongruent	
	Tolerance	VIF	Tolerance	VIF	Tolerance	VIF
Regression on Animism						
Science Education	.99	1.07	.99	1.07	.99	1.07
General Cognitive Ability	.94	1.01	.94	1.01	.94	1.01
Stop Signal Task	.95	1.06	.95	1.06	.95	1.06
Figure Matching Task	.85	1.18	.85	1.18	.85	1.18
Fish Flanker Task	.90	1.11	.90	1.11	.90	1.11
Corsi Block Task	.99	1.01	.99	1.01	.99	1.01
Regression on Buoyancy						
Science Education	.99	1.07	.99	1.07	.99	1.07
General Cognitive Ability	.94	1.01	.94	1.01	.94	1.01
Stop Signal Task	.95	1.06	.95	1.06	.95	1.06
Figure Matching Task	.85	1.18	.85	1.18	.85	1.18
Fish Flanker Task	.90	1.11	.90	1.11	.90	1.11
Corsi Block Task	.99	1.01	.99	1.01	.99	1.01
Regression on Science Statement						
Science Education	.99	1.07	.99	1.07	.99	1.07
General Cognitive Ability	.94	1.01	.94	1.01	.94	1.01
Stop Signal Task	.95	1.06	.95	1.06	.95	1.06
Figure Matching Task	.85	1.18	.85	1.18	.85	1.18
Fish Flanker Task	.90	1.11	.90	1.11	.90	1.11
Corsi Block Task	.99	1.01	.99	1.01	.99	1.01

Table K5*Homoscedasticity test for Accuracy (China)*

	<i>GQ</i>	<i>p</i>
Regression on Animism for Overall trials	.54	.99
Regression on Animism Congruent trials	.65	.99
Regression on Animism Incongruent trials	.62	.99
Regression on Buoyancy Overall trials	3.52	< .001
Regression on Buoyancy Congruent trials	3.35	< .001
Regression on Buoyancy Incongruent trials	3.67	< .001
Regression on Science Statement Overall trials	1.38	.05
Regression on Science Statement Congruent trials	.98	.53
Regression on Science Statement Incongruent trials	1.43	.03

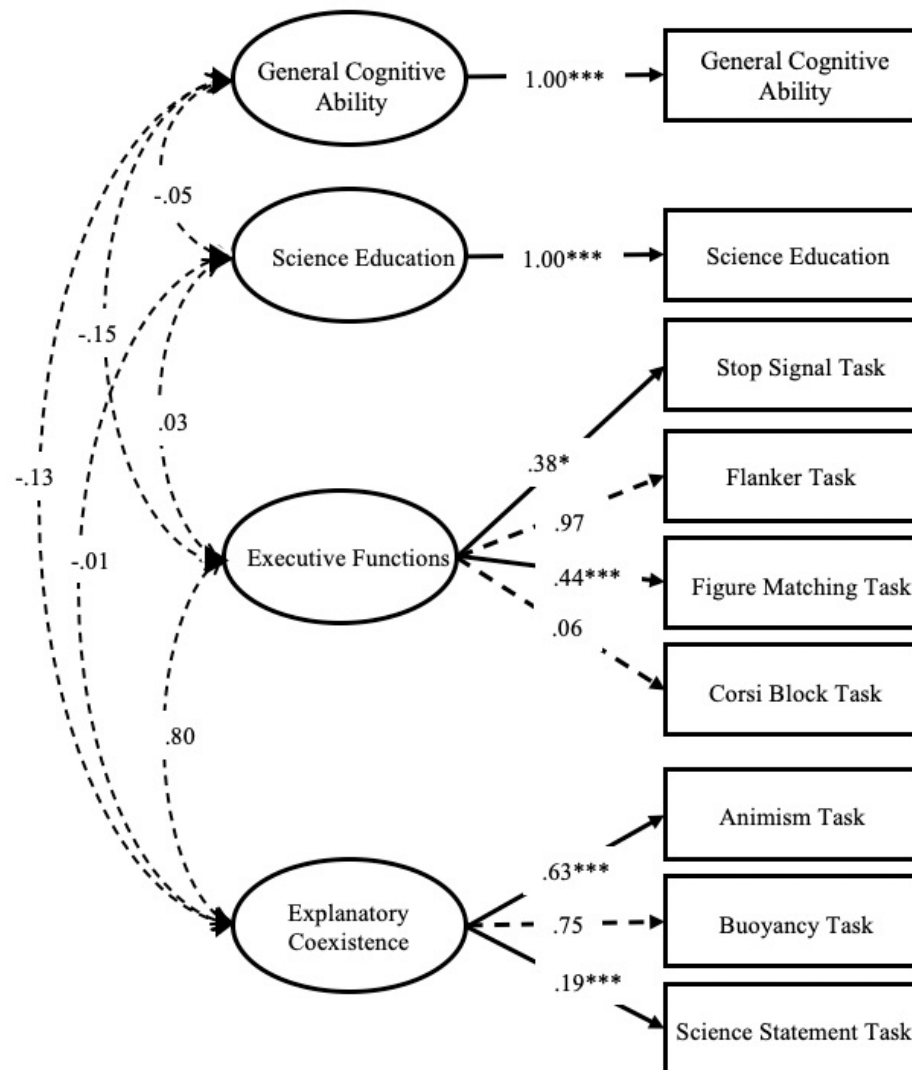
Table K6*Accuracy Confirmatory Factor Analysis Model Fit Indices (China)*

	$\chi^2(df)$	<i>CFI</i>	<i>TLI</i>	<i>AIC</i>	<i>BIC</i>	<i>RMSEA</i>	<i>SRMR</i>
Overall	31.65	.93	.89	5849.49	5956.61	.05	.05
Congruent	21.97	1.00	1.00	5879.99	5879.11	.00	.00
Incongruent	64.73***	.81	.71	6417.29	6528.02	.09	.07

Notes. Significant χ^2 was presented with *; * $p < .05$; ** $p < .01$; *** $p < .001$; χ^2 = Chi-square; *df* = degree of freedom; *CFI* = Comparative Fit Indices; *TLI* = Tucker and Lewis' Index; *AIC* = Akaike Information Criterion; *BIC* = Bayesian Information Criterion; *RMSEA* = Root Mean Square of Approximation; *SRMR* = Standardized Root Mean Square; Robust *CFI*; *TLI*; *RMSEA* and *SRMR* were reported in the Table.

Figure K2

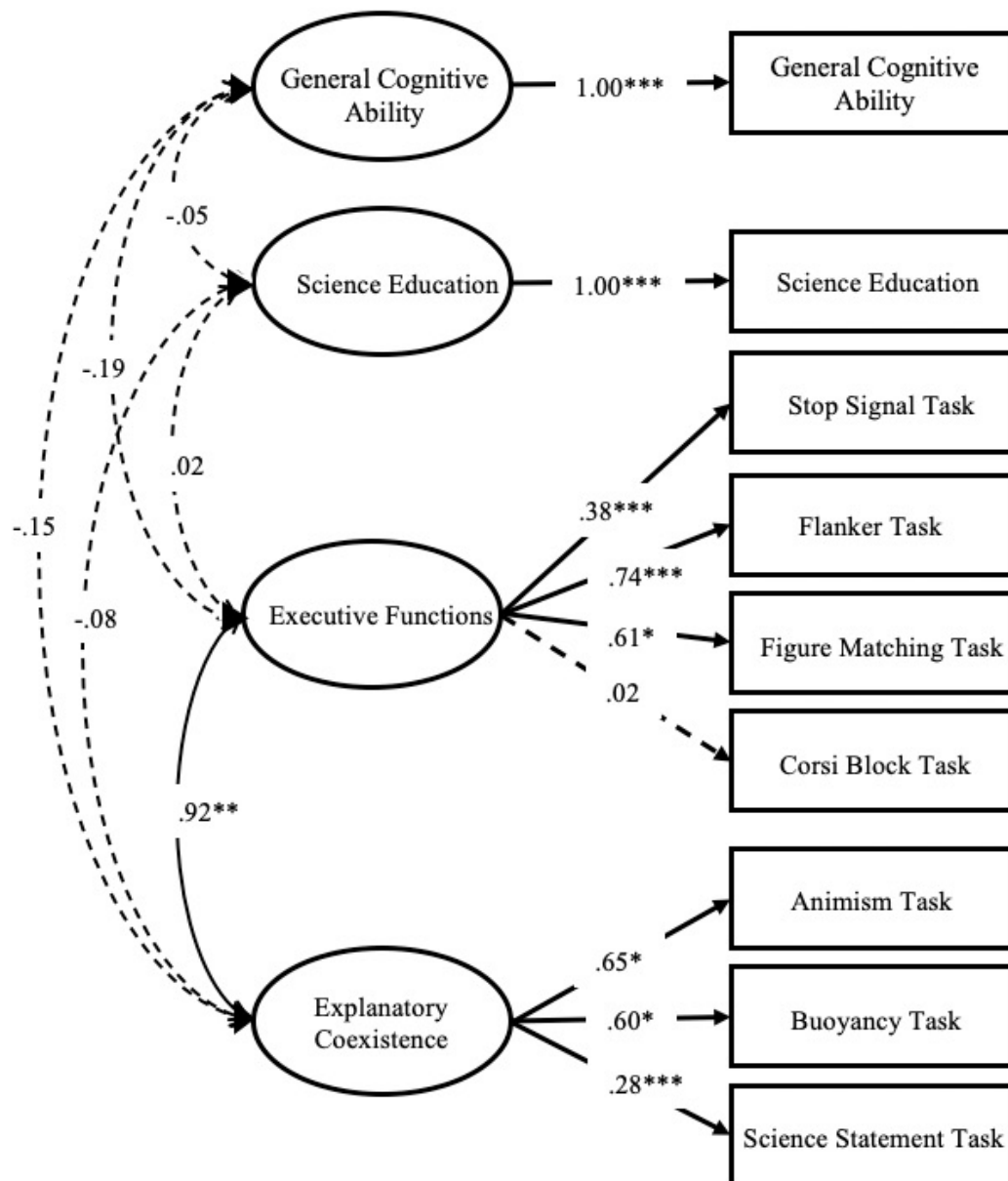
Accuracy for Overall Trials (China)



Notes. Standardised Path Coefficients were reported. * $p < .05$; ** $p < .01$; *** $p < .001$.

Figure K3

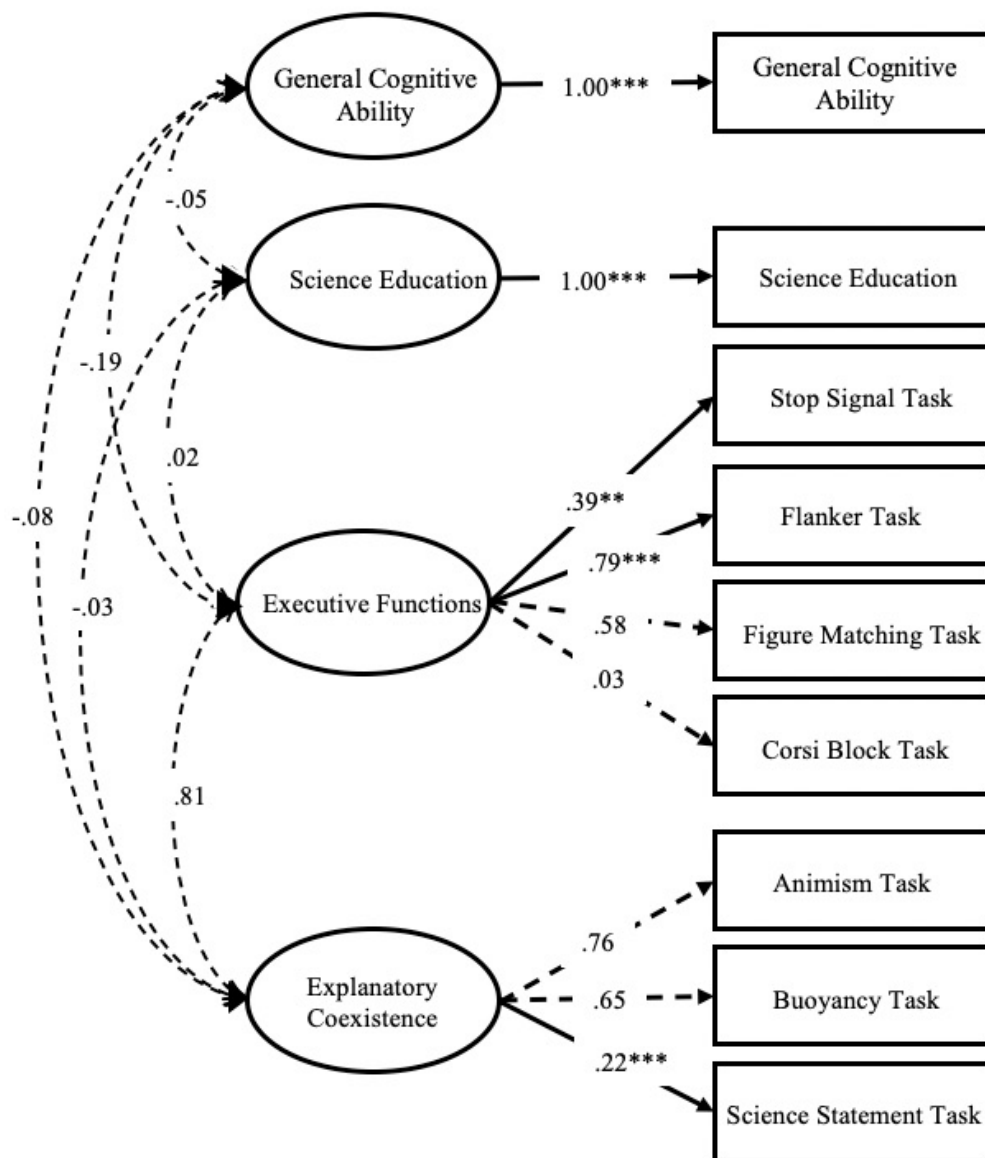
Accuracy for Congruent Trials (China)



Notes. Standardised Path Coefficients were reported. * $p < .05$; ** $p < .01$; *** $p < .001$.

Figure K4

Accuracy for Incongruent Trials (China)



Notes. Standardised Path Coefficients were reported. * $p < .05$; ** $p < .01$; *** $p < .001$.

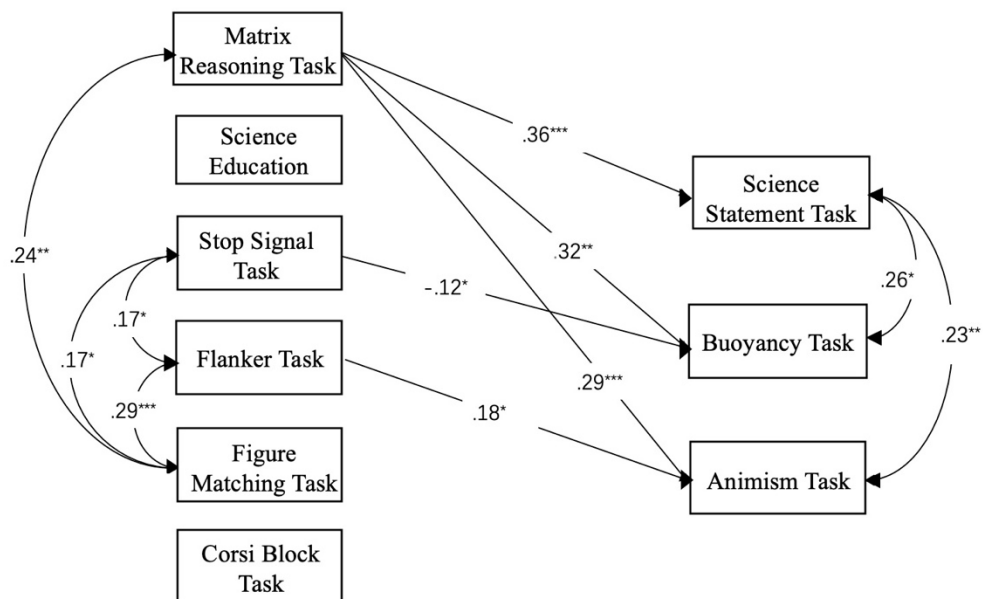
Table K7*Correlation Analysis on Accuracy (China)*

Variables	1	2	3	4	5	6	7	8
<i>Overall Trials</i>								
1. Science Education	-							
2. Matrix Reasoning	.00							
3. Stop Signal	.05	-.01	-					
4. Figure Matching	-.02	.24***	.17*	-				
5. Flanker	.00	.11	.17*	.29***	-			
6. Corsi Block	.06	.01	.04	-.05	.00	-		
7. Science Statement	.08	.39***	.06	.21**	.07	.04		
8. Animism	.10	.33***	.06	.20**	.24***	.00	.35***	-
9. Buoyancy	.08	.33***	-.11	.09	.06	-.03	.35***	.08
<i>Congruent Trials</i>								
1. Science Education	-							
2. Matrix Reasoning	.00	-						
3. Stop Signal	.05	-.01	-					
4. Figure Matching	-.02	.24***	.17*	-				
5. Flanker	.00	.11	.17*	.29***	-			
6. Corsi Block	.06	.01	.04	-.05	.00	-		
7. Science Statement	.04	.34***	.09	.23***	.07	.05		
8. Animism	.16*	.33***	-.03	.13*	.23***	.03	.22***	-
9. Buoyancy	.00	.31***	-.06	.13*	.12	-.01	.19**	.12
<i>Incongruent Trials</i>								
1. Science Education	-							
2. Matrix Reasoning	.00	-						
3. Stop Signal	.05	-.01	-					
4. Figure Matching	-.02	.24***	.17*	-				
5. Flanker	.00	.11	.17*	.29***	-			
6. Corsi Block	.06	.01	.04	-.05	.00	-		
7. Science Statement	.09	.33***	.02	.14*	.05	.02		
8. Animism	.05	.27***	.09	.20**	.20**	-.02	.29***	-
9. Buoyancy	.11	.29***	-.11	.06	.00	-.06	.34***	.04

Notes. * $p < .05$; ** $p < .01$; *** $p < .001$; Correlations $< .30$ are small effect size, between $.30$ and $.50$ are medium effect size, and $> .50$ are large effect size.

Figure K5

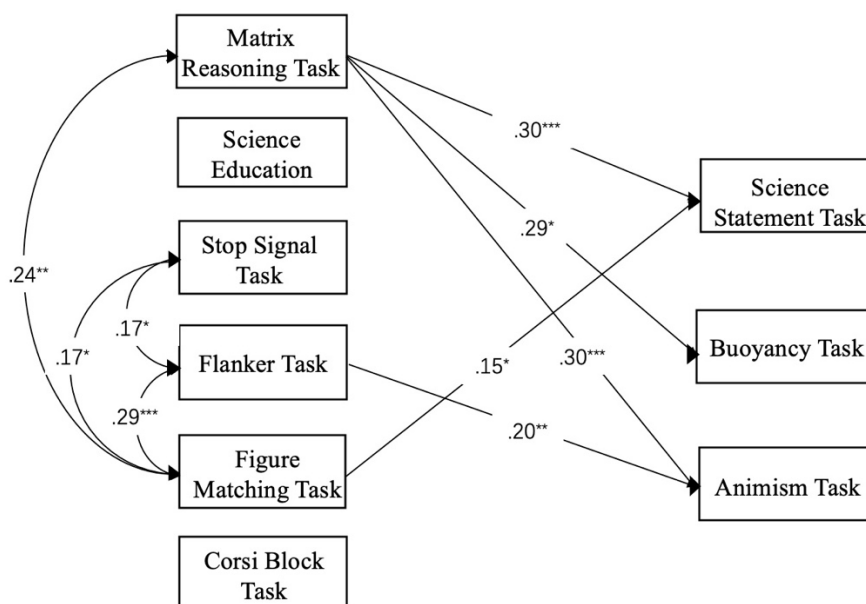
Path Analysis for Overall trials on Accuracy (China)



Notes. Standardised Path Coefficients were reported. * $p < .05$; ** $p < .01$; *** $p < .001$.

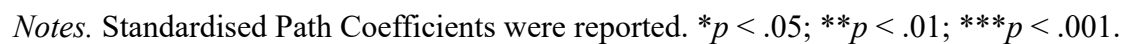
Figure K6

Path Analysis for Congruent trials on Accuracy (China)



Notes. Standardised Path Coefficients were reported. * $p < .05$; ** $p < .01$; *** $p < .001$.

Path Analysis for Incongruent trials on Accuracy (China)



Path Analysis Model Fit Indices Summary on Accuracy (China)

Notes. Significant χ^2 was presented with *; * $p < .05$; ** $p < .01$; *** $p < .001$; χ^2 = Chi-square; df = degree of freedom; CFI = Comparative Fit Indices; TLI = Tucker and Lewis' Index; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; $RMSEA$ = Root Mean Square of Approximation; $SRMR$ = Standardized Root Mean Square; Robust CFI ; TLI ; $RMSEA$ and $SRMR$ were reported in the Table.

Appendix L: Response Time Analysis (China)

Response Time in Appendix L refers to response time for correct trial

Table L1

Descriptive statistics for Response Time (China)

Variables	<i>M</i>	<i>SD</i>	<i>Skewness</i>	<i>Kurtosis</i>
General Cognitive Ability	11.07	3.26	-0.78	0.47
Executive Function				
Stop Signal Task	0.74	0.40	1.59	2.72
Figure Matching Task	0.96	0.23	1.13	2.90
Fish Flanker Task	0.48	0.14	5.53	39.72
Corsi Block Task	3.37	2.26	3.58	16.69
Explanatory Coexistence				
Buoyancy Task				
Overall	0.78	0.59	12.59	176.65
Congruent	0.74	0.22	1.53	59.59
Incongruent	0.79	0.40	5.97	51.71
Animism Task				
Overall	0.72	0.19	2.13	9.23
Congruent	0.69	0.19	2.59	5.03
Incongruent	0.74	0.20	1.93	15.40
Science Statement Task				
Overall	3.14	1.57	2.60	9.70
Congruent	2.75	1.28	2.60	11.45
Incongruent	3.62	2.14	2.84	11.09

Table L2*Age adjustment test for Response Time (China)*

Variables	Linear	Quadratic	Cubic
General Cognitive Ability	.02*	.02*	.01
Executive Function skills			
Stop Signal Task	.02*	.02*	.02
Figure Matching Task	-.00	-.01	-.01
Fish Flanker Task	-.00	-.01	-.01
Corsi Block Task	-.00	.02	.03*
Explanatory Coexistence			
Buoyancy Task			
Overall	-.00	-.01	-.01
Congruent	-.00	-.01	-.01
Incongruent	-.00	-.01	-.01
Animism Task			
Overall	-.00	-.01	-.01
Congruent	-.00	-.01	-.01
Incongruent	-.00	-.01	-.01
Science Statement Task			
Overall	.01	.01	-.00
Congruent	-.00	-.01	-.01
Incongruent	.02*	.02	.02

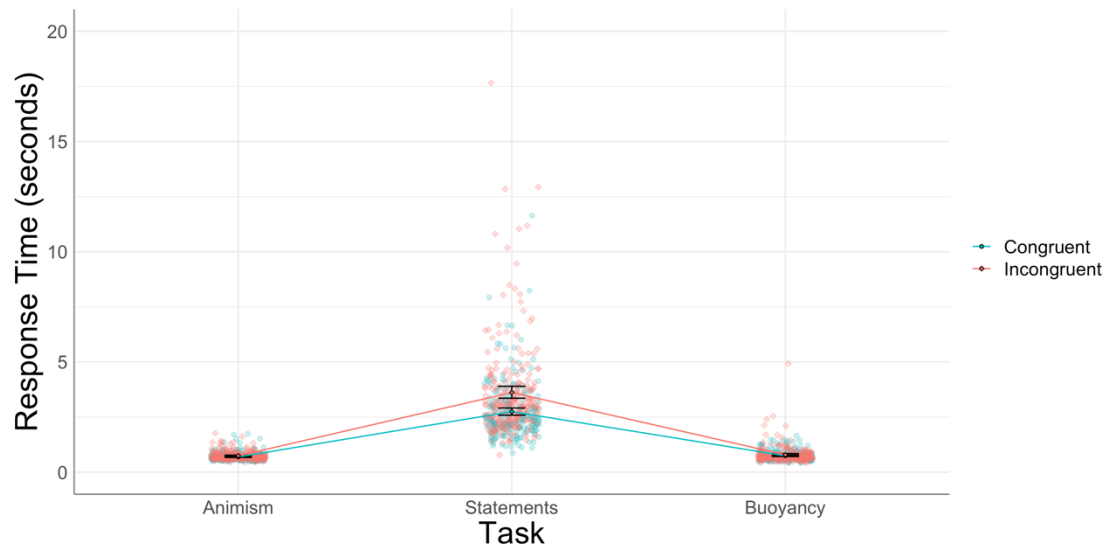
Notes. * $p < .05$; ** $p < .01$; *** $p < .001$.

Table L3*Shapiro-Wilk test for Response Time (China)*

Variables	<i>Shapiro-Wilk</i>	<i>p</i>
General Cognitive Ability	.96	< .001
Executive Function		
Stop Signal Task	.86	< .001
Figure Matching Task	.91	< .001
Fish Flanker Task	.52	< .001
Corsi Block Task	.63	< .001
Explanatory Coexistence		
Buoyancy Task		
Overall	.25	< .001
Congruent	.86	< .001
Incongruent	.53	< .001
Animism Task		
Overall	.81	< .001
Congruent	.76	< .001
Incongruent	.83	< .001
Science Statement Task		
Overall	.76	< .001
Congruent	.79	< .001
Incongruent	.73	< .001

Figure L1

Mean Comparison of Response Time between Congruent and Incongruent Trials for Naïve Thinking Tasks (China)



Notes. In response time analysis, significant differences between congruent and incongruent trials were observed in the Animism task ($t = 8.94, p < .001$) and the Buoyancy task ($t = 2.65, p = .008$), but not in the Science Statement task ($t = 0.80, p = .427$); Response time here refers to response time in correct trials; Thanks to Clemens Brunner for sharing the R script used to create this figure in Stricker et al. (2021).

Table L4*Multicollinearity test for Response Time (China)*

	Overall		Congruent		Incongruent	
	Tolerance	VIF	Tolerance	VIF	Tolerance	VIF
Regression on Animism						
Science Education	.98	1.02	.98	1.02	.98	1.02
General Cognitive Ability	.95	1.05	.95	1.05	.95	1.05
Stop Signal Task	.85	1.18	.85	1.18	.85	1.18
Figure Matching Task	.77	1.30	.77	1.30	.77	1.30
Fish Flanker Task	.69	1.46	.69	1.46	.69	1.46
Corsi Block Task	.96	1.04	.96	1.04	.96	1.04
Regression on Buoyancy						
Science Education	.98	1.02	.98	1.02	.98	1.02
General Cognitive Ability	.95	1.05	.95	1.05	.95	1.05
Stop Signal Task	.85	1.18	.85	1.18	.85	1.18
Figure Matching Task	.77	1.30	.77	1.30	.77	1.30
Fish Flanker Task	.69	1.46	.69	1.46	.69	1.46
Corsi Block Task	.96	1.04	.96	1.04	.96	1.04
Regression on Science Statement						
Science Education	.98	1.02	.98	1.02	.98	1.02
General Cognitive Ability	.95	1.05	.95	1.05	.95	1.05
Stop Signal Task	.85	1.18	.85	1.18	.85	1.18
Figure Matching Task	.77	1.30	.77	1.30	.77	1.30
Fish Flanker Task	.69	1.46	.69	1.46	.69	1.46
Corsi Block Task	.96	1.04	.96	1.04	.96	1.04

Table L5*Homoscedasticity test for Response Time (China)*

	<i>GQ</i>	<i>p</i>
Regression on Animism for Overall trials	.66	.98
Regression on Animism Congruent trials	.69	.97
Regression on Animism Incongruent trials	.65	.99
Regression on Buoyancy Overall trials	1.23	.14
Regression on Buoyancy Congruent trials	1.65	.005
Regression on Buoyancy Incongruent trials	1.79	.001
Regression on Science Statement Overall trials	.67	.98
Regression on Science Statement Congruent trials	.61	.99
Regression on Science Statement Incongruent trials	.82	.85

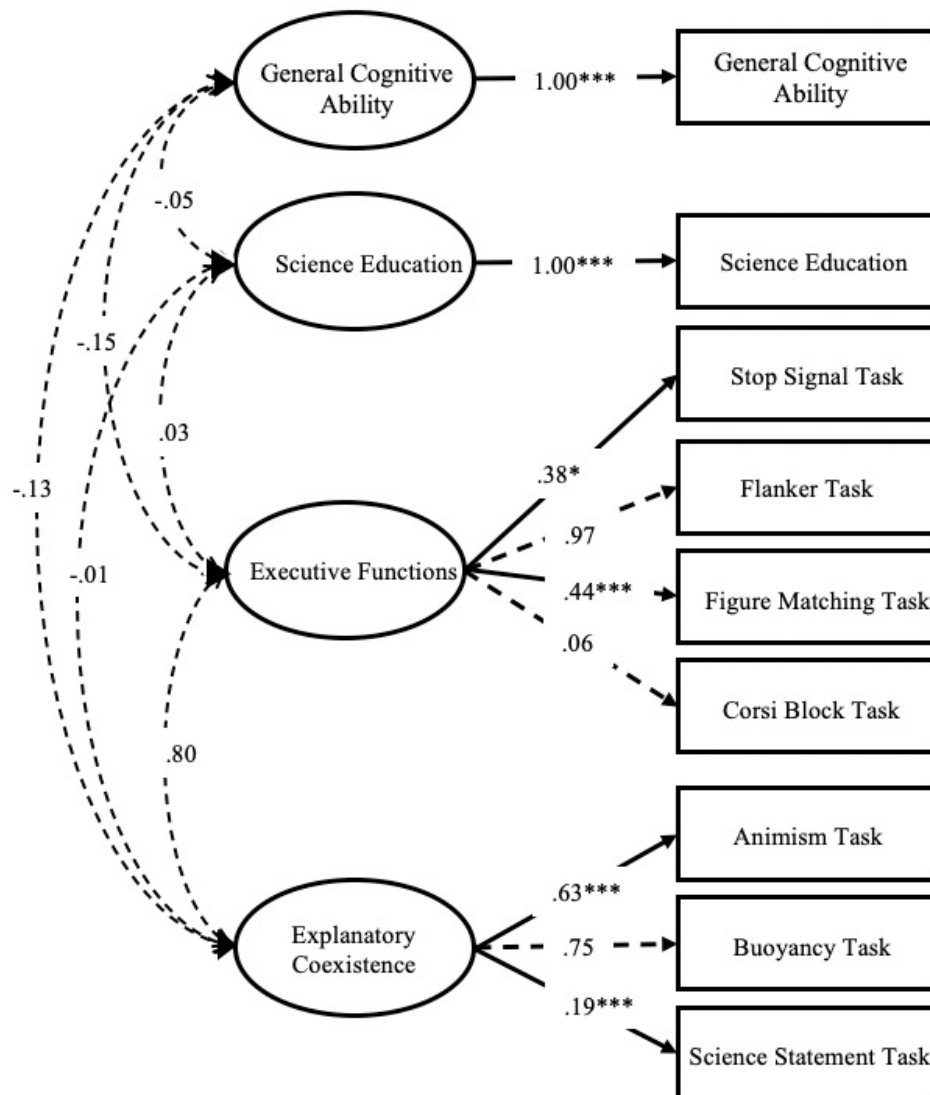
Table L6*Response Time Confirmatory Factor Analysis Model Fit Indices (China)*

	$\chi^2(df)$	<i>CFI</i>	<i>TLI</i>	<i>AIC</i>	<i>BIC</i>	<i>RMSEA</i>	<i>SRMR</i>
Overall	78.36***	.75	.61	5741.62	5848.74	.11	.07
Congruent	61.07***	.78	.65	5787.18	5894.30	.09	.06
Incongruent	48.43**	.82	.71	5769.42	5876.53	.09	.07

Notes. Significant χ^2 was presented with *; * $p < .05$; ** $p < .01$; *** $p < .001$; χ^2 = Chi-square; *df* = degree of freedom; *CFI* = Comparative Fit Indices; *TLI* = Tucker and Lewis' Index; *AIC* = Akaike Information Criterion; *BIC* = Bayesian Information Criterion; *RMSEA* = Root Mean Square of Approximation; *SRMR* = Standardized Root Mean Square; Robust *CFI*; *TLI*; *RMSEA* and *SRMR* were reported in the Table.

Figure L2

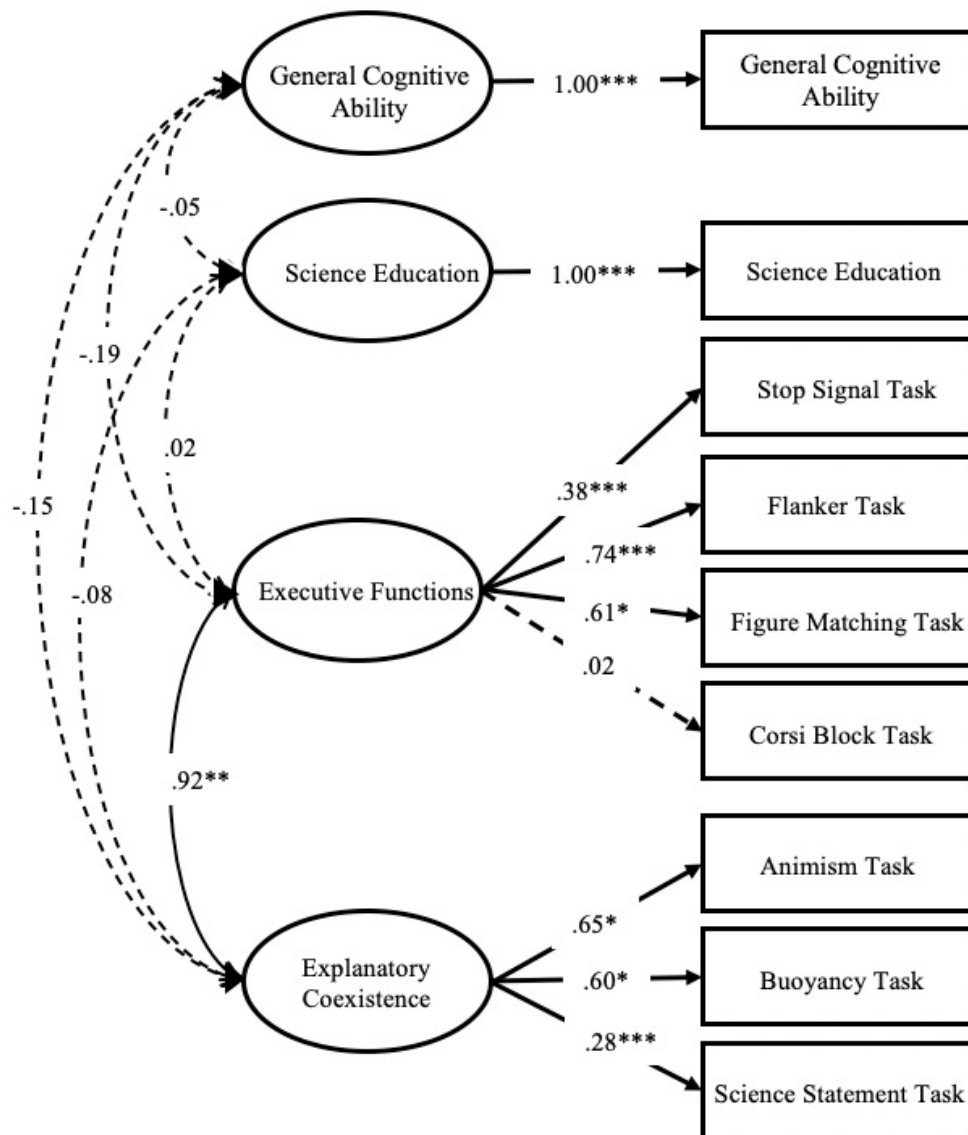
Response Time for Overall Trials (China)



Notes. Standardised Path Coefficients were reported. * $p < .05$; ** $p < .01$; *** $p < .001$.

Figure L3

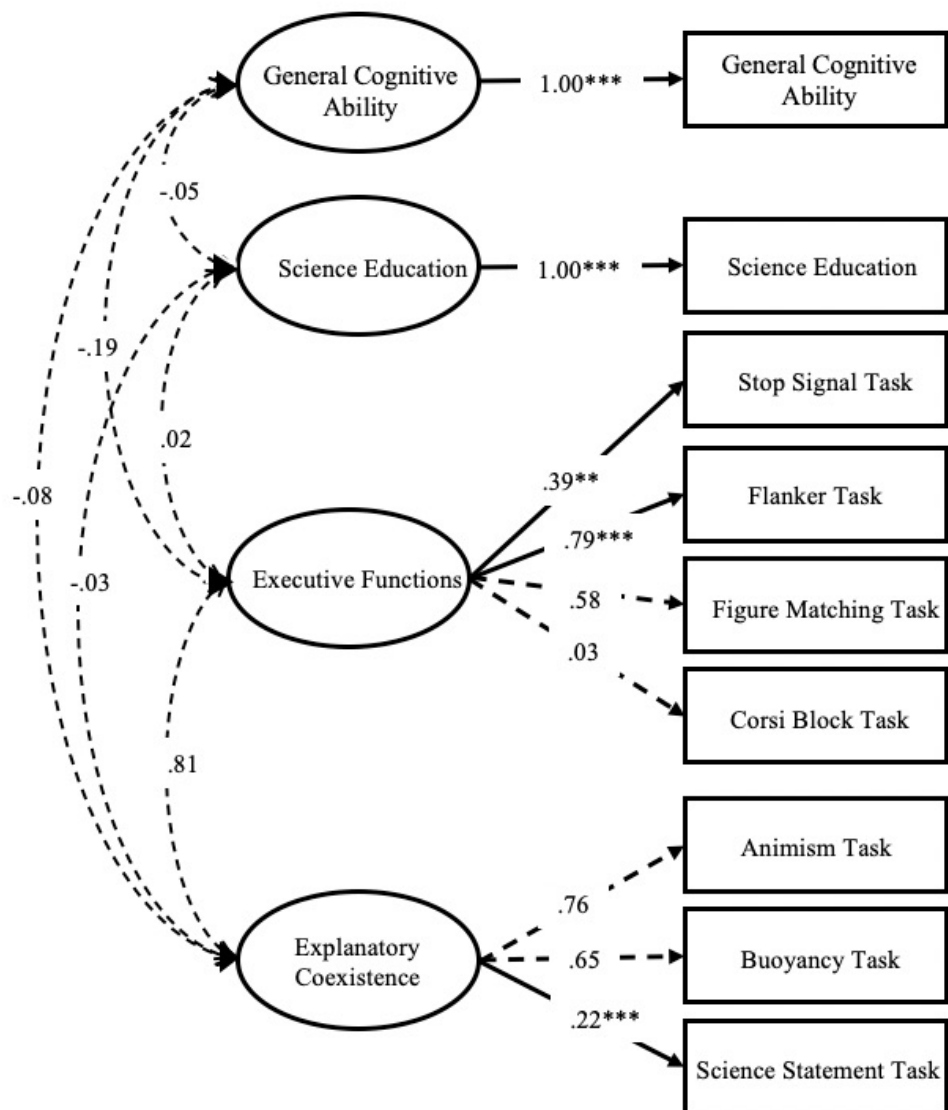
Response Time for Congruent Trials (China)



Notes. Standardised Path Coefficients were reported. * $p < .05$; ** $p < .01$; *** $p < .001$.

Figure L4

Response Time for Incongruent Trials (China)



Notes. Standardised Path Coefficients were reported. * $p < .05$; ** $p < .01$; *** $p < .001$.

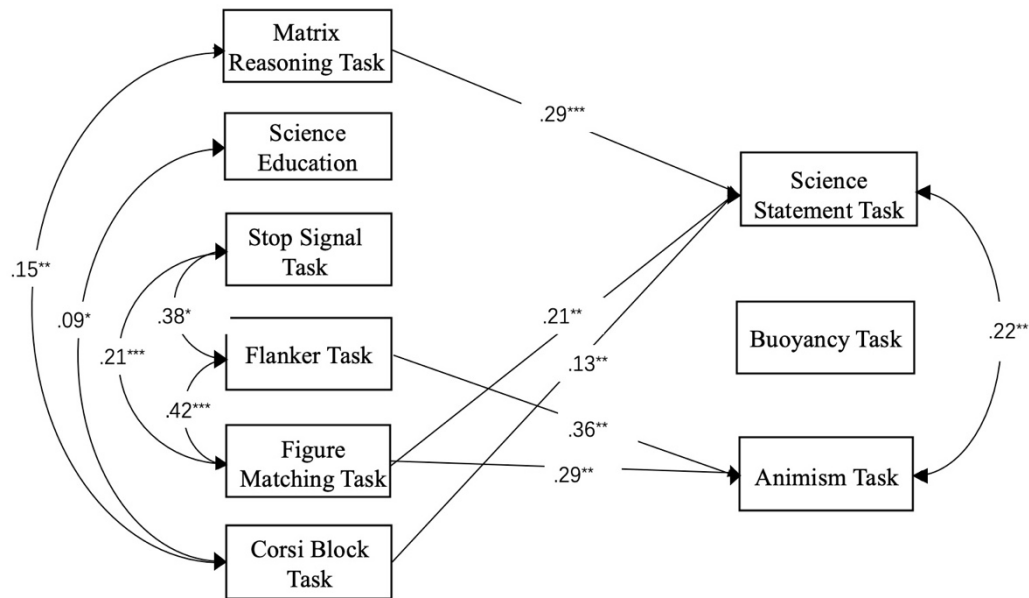
Table L7*Correlation Analysis on Response Time (China)*

Variables	1	2	3	4	5	6	7	8
<i>Overall Trials</i>								
1. Science Education	-							
2. Matrix Reasoning	-.05							
3. Stop Signal	-.07	-.14*	-					
4. Figure Matching	.01	-.09	.21**	-				
5. Flanker	.03	-.14*	.38***	.42***	-			
6. Corsi Block	.09	.15*	-.01	-.01	.05	-		
7. Science Statement	.01	.29***	-.02	.21**	.13*	.18**		
8. Animism	-.05	-.14*	.15*	.43***	.47***	.00	.26***	-
9. Buoyancy	.02	-.09	.14*	.19**	.60***	.01	.07	.47***
<i>Congruent Trials</i>								
1. Science Education	-							
2. Matrix Reasoning	-.05	-						
3. Stop Signal	-.07	-.14*	-					
4. Figure Matching	.01	-.09	.21**	-				
5. Flanker	.03	-.14*	.38***	.42***	-			
6. Corsi Block	.09	.15*	-.01	-.01	.05	-		
7. Science Statement	-.04	.27***	-.01	.24***	.14*	.10		
8. Animism	-.05	-.18**	.16*	.37***	.46***	.01	.23***	-
9. Buoyancy	-.04	-.09	.15*	.41***	.40***	-.02	.17*	.37***
<i>Incongruent Trials</i>								
1. Science Education	-							
2. Matrix Reasoning	-.05	-						
3. Stop Signal	-.07	-.14*	-					
4. Figure Matching	.01	-.09	.21**	-				
5. Flanker	.03	-.14*	.38***	.42***	-			
6. Corsi Block	.09	.15*	-.01	-.01	.05	-		
7. Science Statement	.04	.27***	-.03	-.17**	.10	.21**		
8. Animism	-.04	-.10	.14*	.46***	.45***	-.01	.23***	-
9. Buoyancy	.00	-.04	.12	.28***	.47***	.03	.05	.49***

Notes. * $p < .05$; ** $p < .01$; *** $p < .001$; Correlations $< .30$ are small effect size, between $.30$ and $.50$ are medium effect size, and $> .50$ are large effect size.

Figure L5

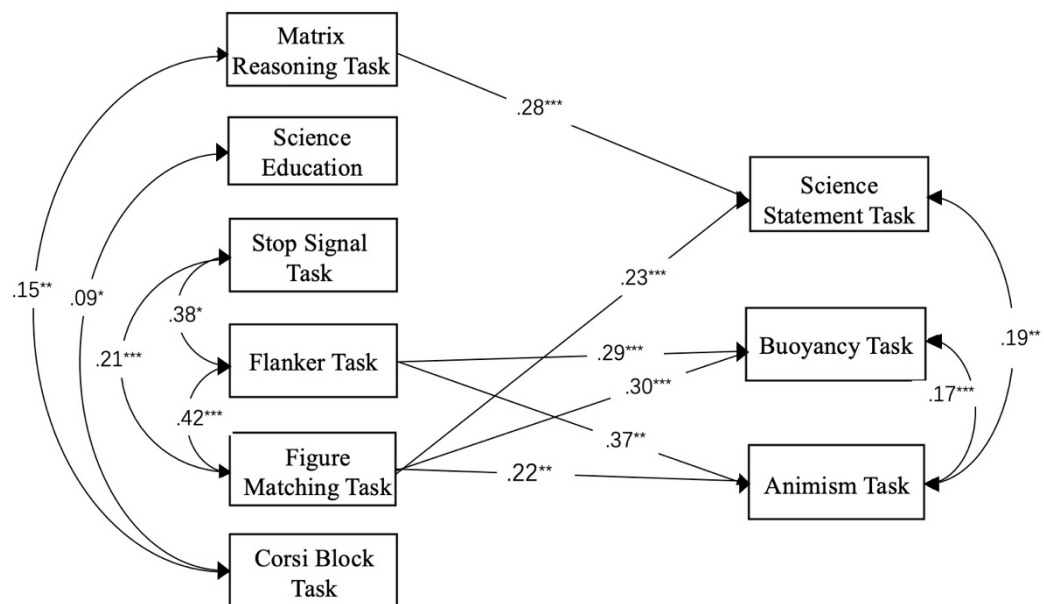
Path Analysis for Overall trials on Response Time (China)



Notes. Standardised Path Coefficients were reported. * $p < .05$; ** $p < .01$; *** $p < .001$.

Figure L6

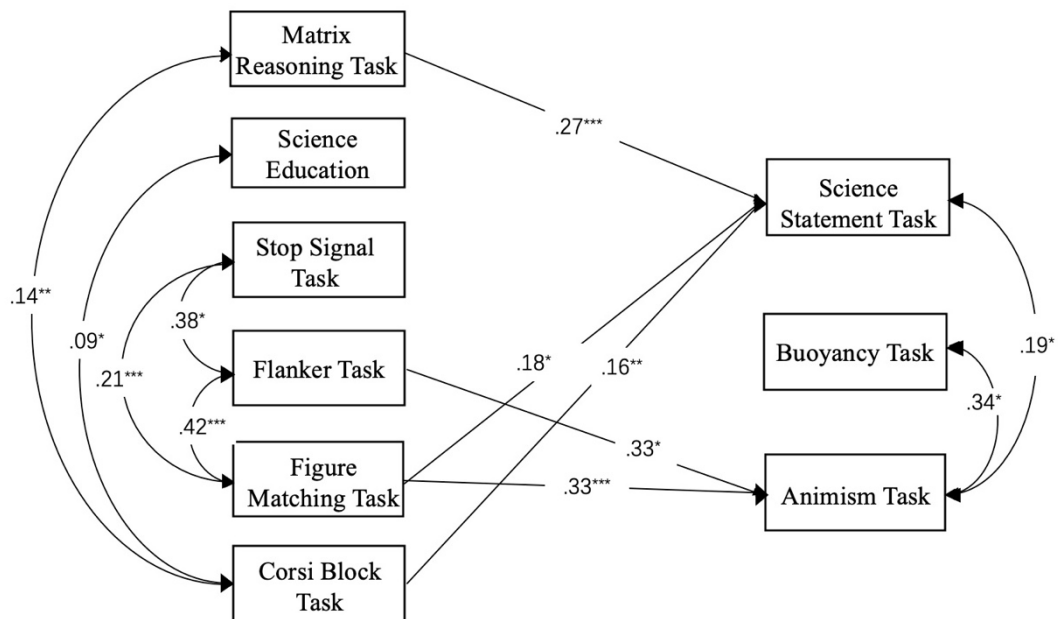
Path Analysis for Congruent trials on Response Time (China)



Notes. Standardised Path Coefficients were reported. * $p < .05$; ** $p < .01$; *** $p < .001$.

Figure L7

Path Analysis for Incongruent trials on Response Time (China)



Notes. Standardised Path Coefficients were reported. * $p < .05$; ** $p < .01$; *** $p < .001$.

Table L8

Path Analysis Model Fit Indices Summary on Response Time (China)

Trial	$\chi^2(df)$	CFI	TLI	AIC	BIC	RMSEA	SRMR
Overall	.20(1)	1.00	1.15	5688.68	5871.81	.00	.01
Congruent	.73(1)	1.00	1.02	5767.04	5950.18	.00	.01
Incongruent	.01(1)	1.00	1.09	5736.14	5919.27	.00	.00

Notes. Significant χ^2 was presented with *; * $p < .05$; ** $p < .01$; *** $p < .001$; χ^2 = Chi-square; df = degree of freedom; CFI = Comparative Fit Indices; TLI = Tucker and Lewis' Index; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; $RMSEA$ = Root Mean Square of Approximation; $SRMR$ = Standardized Root Mean Square; Robust CFI ; TLI ; $RMSEA$ and $SRMR$ were reported in the Table.

Appendix M: Efficiency Analysis (China)

Table M1

Descriptive statistics for Efficiency (China)

Variables	<i>M</i>	<i>SD</i>	<i>Skewness</i>	<i>Kurtosis</i>
General Cognitive Ability	0.07	0.02	3.93	26.69
Executive Function				
Stop Signal Task	1.47	0.55	0.00	-1.02
Figure Matching Task	1.05	0.22	0.08	0.40
Fish Flanker Task	2.13	0.33	-1.42	4.53
Corsi Block Task	0.21	0.08	0.39	0.32
Explanatory Coexistence				
Buoyancy Task				
Overall	1.37	0.36	-0.45	0.71
Congruent	1.40	0.37	-0.34	5.89
Incongruent	1.34	0.39	-0.61	0.94
Animism Task				
Overall	1.40	0.29	-0.50	-0.04
Congruent	1.48	0.30	-.68	0.30
Incongruent	1.34	0.29	-.68	-0.35
Science Statement Task				
Overall	0.29	0.10	0.43	0.44
Congruent	0.37	0.13	0.43	0.20
Incongruent	0.22	0.09	0.45	0.64

Table M2*Age adjustment test for Efficiency (China)*

Variables	Linear	Quadratic	Cubic
General Cognitive Ability	.01	.00	.00
Executive Function skills			
Stop Signal Task	.01*	.01	.02
Figure Matching Task	-.00	-.01	-.01
Fish Flanker Task	-.00	-.00	-.00
Corsi Block Task	-.00	.00	.01
Explanatory Coexistence			
Buoyancy Task			
Overall	-.00	-.01	-.01
Congruent	-.00	-.01	-.01
Incongruent	-.00	-.01	-.01
Animism Task			
Overall	-.00	-.01	-.01
Congruent	-.00	-.01	-.01
Incongruent	-.00	-.01	-.01
Science Statement Task			
Overall	.02*	.02	.01
Congruent	.00	.00	-.01
Incongruent	.03*	.03**	.03*

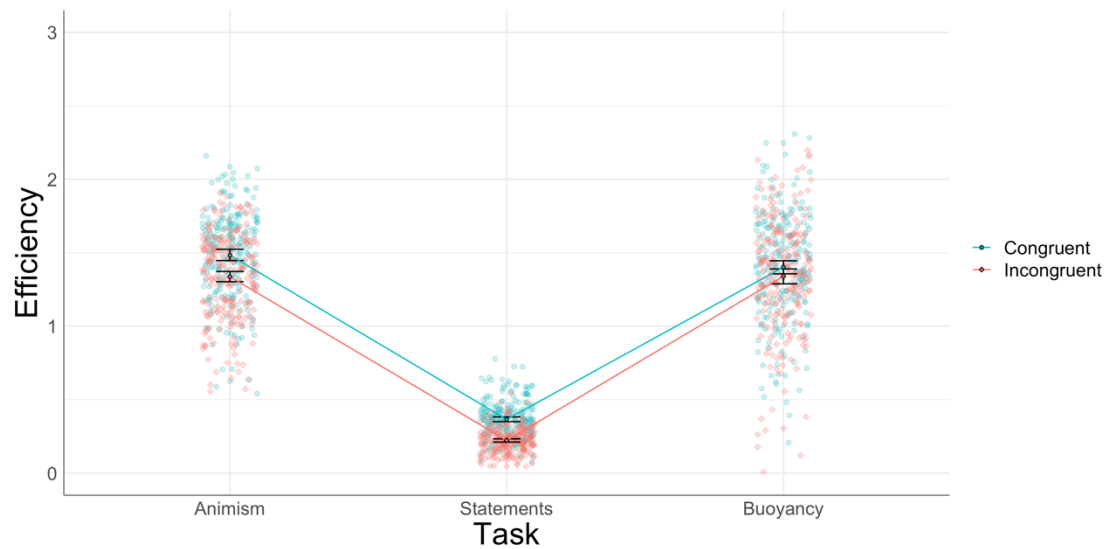
Notes. * $p < .05$; ** $p < .01$; *** $p < .001$.

Table M3*Shapiro-Wilk test for Efficiency (China)*

Variables	<i>Shapiro-Wilk</i>	<i>p</i>
General Cognitive Ability	.73	< .001
Executive Function		
Stop Signal Task	.98	< .001
Figure Matching Task	.99	.19
Fish Flanker Task	.90	< .001
Corsi Block Task	.98	< .001
Explanatory Coexistence		
Buoyancy Task		
Overall	.98	.001
Congruent	.99	.051
Incongruent	.97	< .001
Animism Task		
Overall	.98	< .001
Congruent	.98	.001
Incongruent	.98	< .001
Science Statement Task		
Overall	.99	.049
Congruent	.99	.038
Incongruent	.99	.051

Figure M1

Mean Comparison of Efficiency between Congruent and Incongruent Trials for Naïve Thinking Tasks (China)



Notes. In efficiency analysis, significant differences between congruent and incongruent trials were observed in the Animism task ($t = 13.78, p < .001$), the Buoyancy task ($t = 3.89, p < .001$), and the Science Statement task ($t = 27.15, p < .001$); Thanks to Clemens Brunner for sharing the R script used to create this figure in Stricker et al. (2021).

Table M4*Multicollinearity test for Efficiency (China)*

	Overall		Congruent		Incongruent	
	Tolerance	VIF	Tolerance	VIF	Tolerance	VIF
Regression on Animism						
Science Education	.99	1.01	.99	1.01	.99	1.01
General Cognitive Ability	.97	1.03	.97	1.03	.97	1.03
Stop Signal Task	.87	1.15	.87	1.15	.87	1.15
Figure Matching Task	.64	1.57	.64	1.57	.64	1.57
Fish Flanker Task	.59	1.70	.59	1.70	.59	1.70
Corsi Block Task	.87	1.17	.87	1.17	.87	1.17
Regression on Buoyancy						
Science Education	.99	1.01	.99	1.01	.99	1.01
General Cognitive Ability	.97	1.03	.97	1.03	.97	1.03
Stop Signal Task	.87	1.15	.87	1.15	.87	1.15
Figure Matching Task	.64	1.57	.64	1.57	.64	1.57
Fish Flanker Task	.59	1.70	.59	1.70	.59	1.70
Corsi Block Task	.87	1.17	.87	1.17	.87	1.17
Regression on Science Statement						
Science Education	.99	1.01	.99	1.01	.99	1.01
General Cognitive Ability	.97	1.03	.97	1.03	.97	1.03
Stop Signal Task	.87	1.15	.87	1.15	.87	1.15
Figure Matching Task	.64	1.57	.64	1.57	.64	1.57
Fish Flanker Task	.59	1.70	.59	1.70	.59	1.70
Corsi Block Task	.87	1.17	.87	1.17	.87	1.17

Table M5*Homoscedasticity test for Efficiency (China)*

	<i>GQ</i>	<i>p</i>
Regression on Animism for Overall trials	.71	.96
Regression on Animism Congruent trials	.77	.91
Regression on Animism Incongruent trials	.67	.98
Regression on Buoyancy Overall trials	1.16	.22
Regression on Buoyancy Congruent trials	1.32	.08
Regression on Buoyancy Incongruent trials	1.09	.33
Regression on Science Statement Overall trials	.91	.69
Regression on Science Statement Congruent trials	.96	.58
Regression on Science Statement Incongruent trials	.85	.80

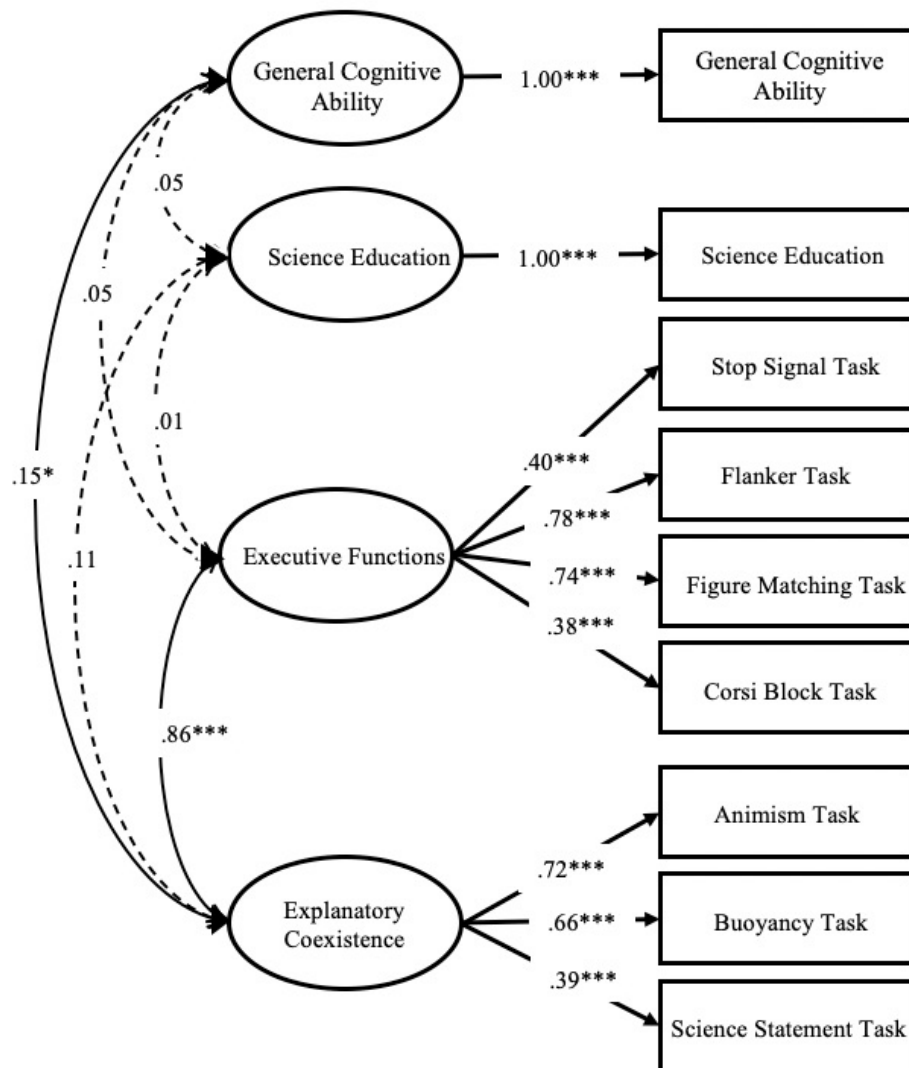
Table M6*Efficiency Confirmatory Factor Analysis Model Fit Indices (China)*

	$\chi^2(df)$	<i>CFI</i>	<i>TLI</i>	<i>AIC</i>	<i>BIC</i>	<i>RMSEA</i>	<i>SRMR</i>
Overall	35.07	.96	.93	5660.46	5767.57	.05	.05
Congruent	33.93	.96	.94	5679.37	5786.48	.05	.04
Incongruent	33.95	.96	.93	5683.03	5790.14	.05	.05

Notes. Significant χ^2 was presented with *; * $p < .05$; ** $p < .01$; *** $p < .001$; χ^2 = Chi-square; *df* = degree of freedom; *CFI* = Comparative Fit Indices; *TLI* = Tucker and Lewis' Index; *AIC* = Akaike Information Criterion; *BIC* = Bayesian Information Criterion; *RMSEA* = Root Mean Square of Approximation; *SRMR* = Standardized Root Mean Square; Robust *CFI*; *TLI*; *RMSEA* and *SRMR* were reported in the Table.

Figure M2

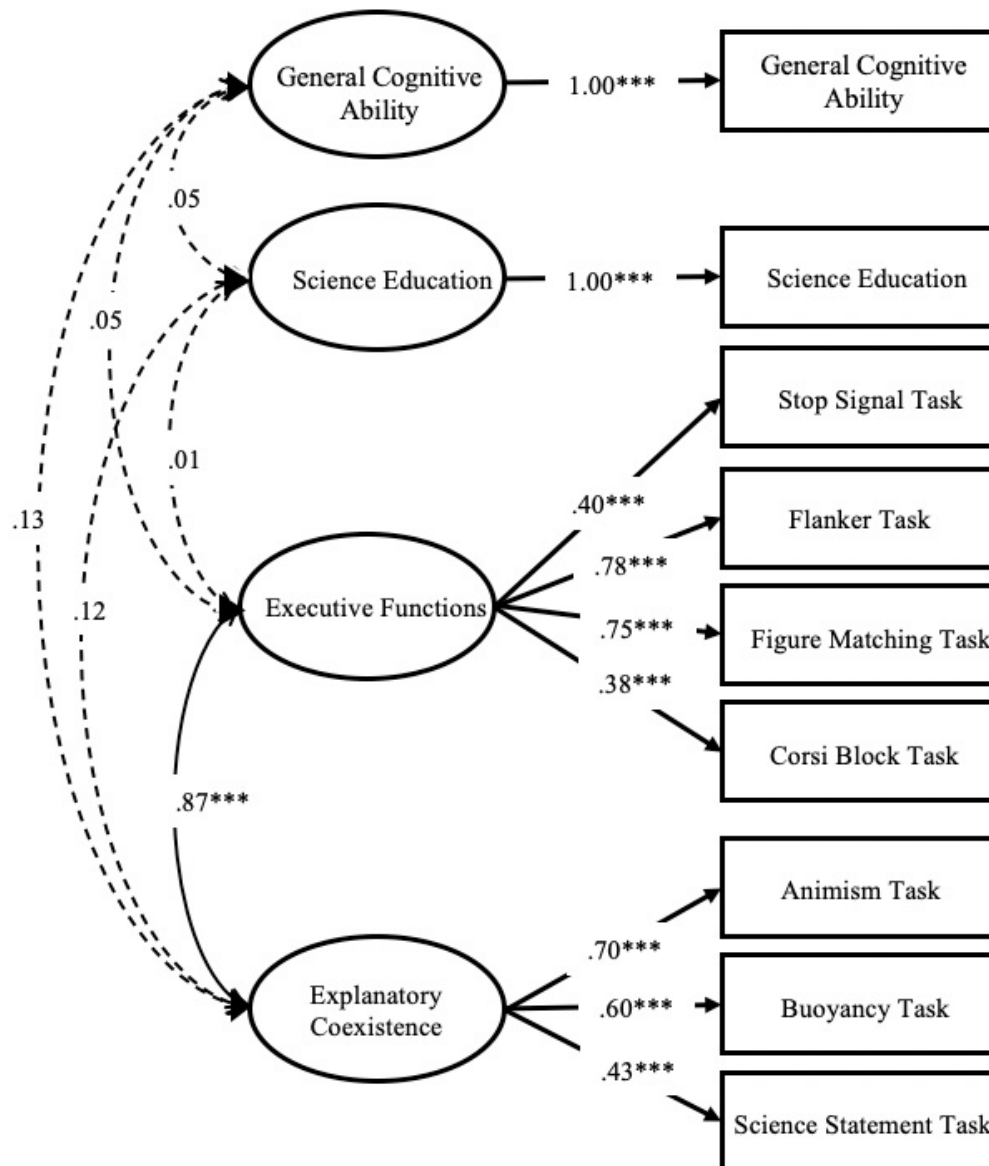
Efficiency for Overall Trials (China)



Notes. Standardised Path Coefficients were reported. * $p < .05$; ** $p < .01$; *** $p < .001$.

Figure M3

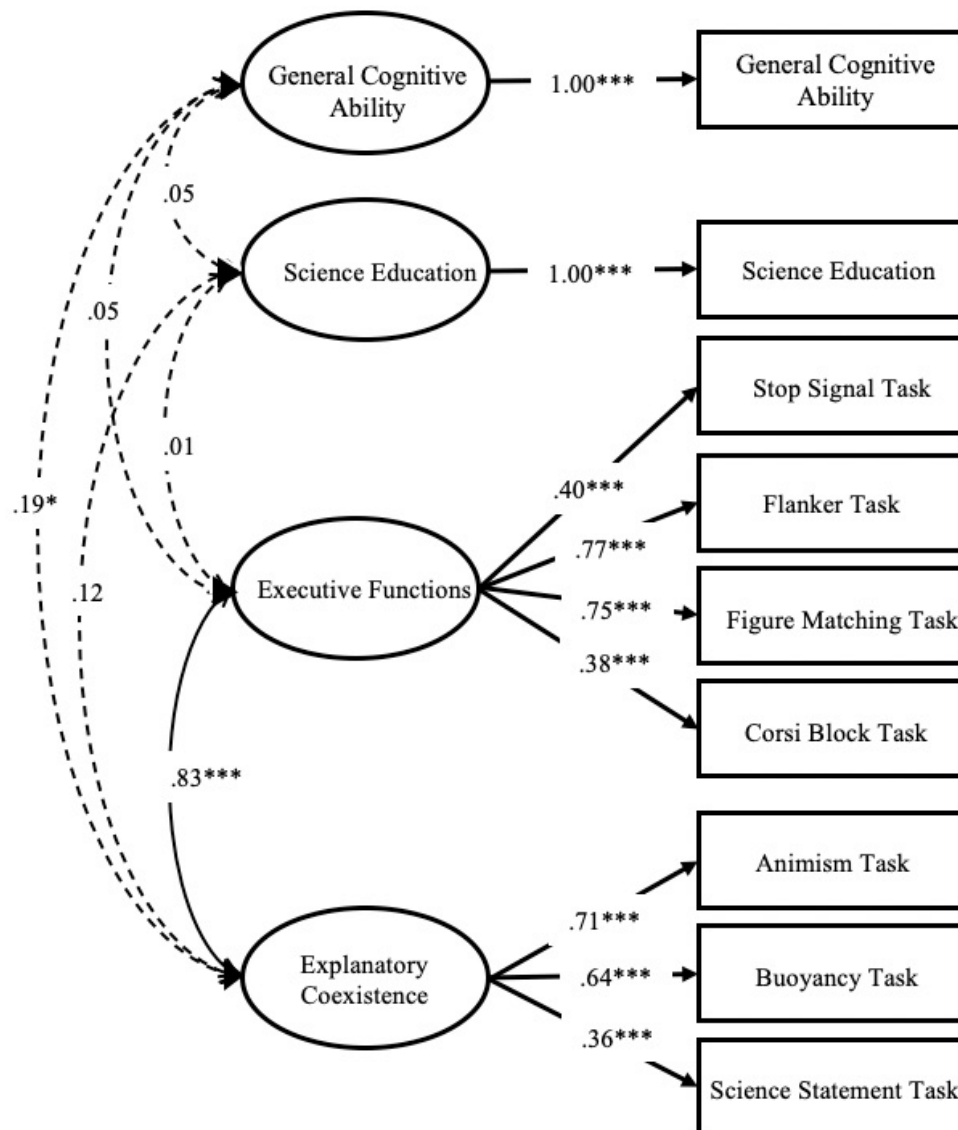
Efficiency for Congruent Trials (China)



Notes. Standardised Path Coefficients were reported. * $p < .05$; ** $p < .01$; *** $p < .001$.

Figure M4

Efficiency for Incongruent Trials (China)



Notes. Standardised Path Coefficients were reported. * $p < .05$; ** $p < .01$; *** $p < .001$.

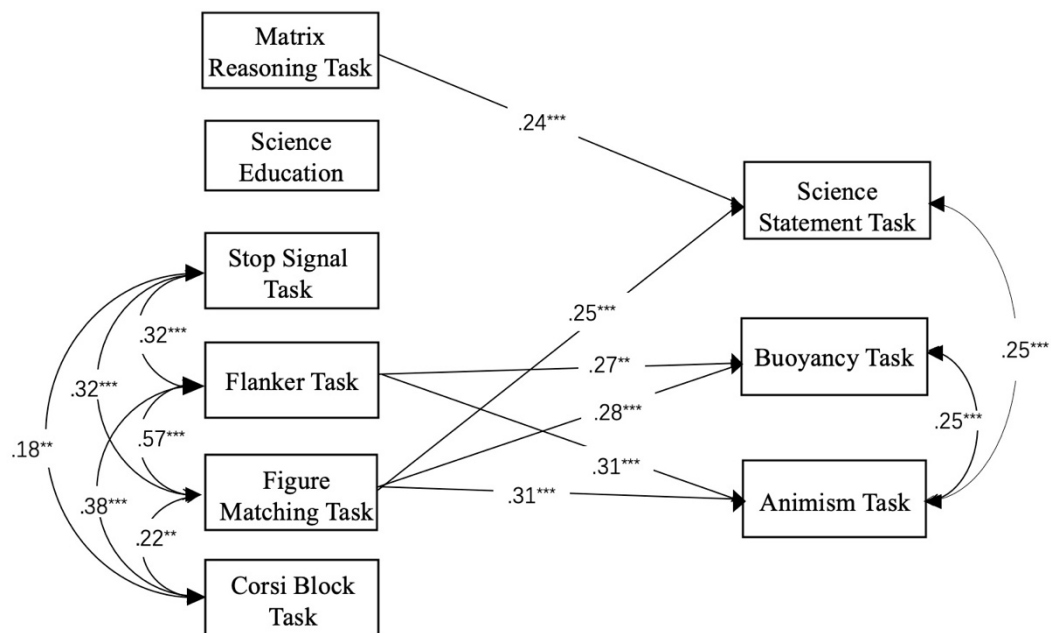
Table 6.11*Correlation Analysis on Efficiency (China)*

Variables	1	2	3	4	5	6	7	8
<i>Overall Trials</i>								
1. Science Education	-							
2. Matrix Reasoning	.06							
3. Stop Signal	.05	.00	-					
4. Figure Matching	.04	.07	.32***	-				
5. Flanker	-.01	-.01	.32***	.57***	-			
6. Corsi Block	-.05	.09	.17**	.22***	.36***	-		
7. Science Statement	.01	.27***	.05	.27***	.16*	.17**		
8. Animism	.09	.04	.22***	.49***	.49***	.16*	.33***	-
9. Buoyancy	.07	.11	.20**	.46***	.45***	.19**	.23***	.46***
<i>Congruent Trials</i>								
1. Science Education	-							
2. Matrix Reasoning	.06	-						
3. Stop Signal	.05	.00	-					
4. Figure Matching	.04	.07	.32***	-				
5. Flanker	-.01	-.01	.32***	.57***	-			
6. Corsi Block	-.05	.09	.17**	.22***	.36***	-		
7. Science Statement	.03	.25***	.07	.33***	.21**	.16*		
8. Animism	.10	.00	.22***	.46***	.49***	.16*	.33***	-
9. Buoyancy	.05	.09	.19**	.42***	.40***	.16*	.24***	.41***
<i>Incongruent Trials</i>								
1. Science Education	-							
2. Matrix Reasoning	.06	-						
3. Stop Signal	.05	.00	-					
4. Figure Matching	.04	.07	.32***	-				
5. Flanker	-.01	-.01	.32***	.57***	-			
6. Corsi Block	-.05	.09	.17**	.22***	.36***	-		
7. Science Statement	.01	.28***	.05	.21**	.11	.15*		
8. Animism	.07	.07	.20**	.49***	.46***	.15*	.30***	-
9. Buoyancy	.05	.13	.18**	.44***	.41***	.21**	.22***	.44***

Notes. * $p < .05$; ** $p < .01$; *** $p < .001$; Correlations $< .30$ are small effect size, between $.30$ and $.50$ are medium effect size, and $> .50$ are large effect size.

Figure 6.3a

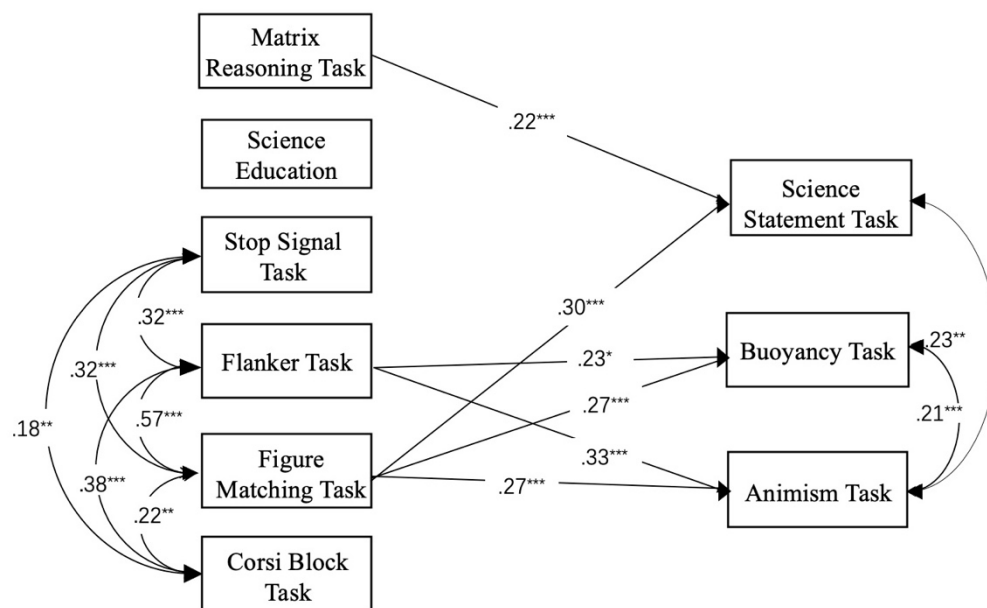
Path Analysis for Overall trials on Efficiency (China)



Notes. Standardised Path Coefficients were reported. * $p < .05$; ** $p < .01$; *** $p < .001$.

Figure 6.3b

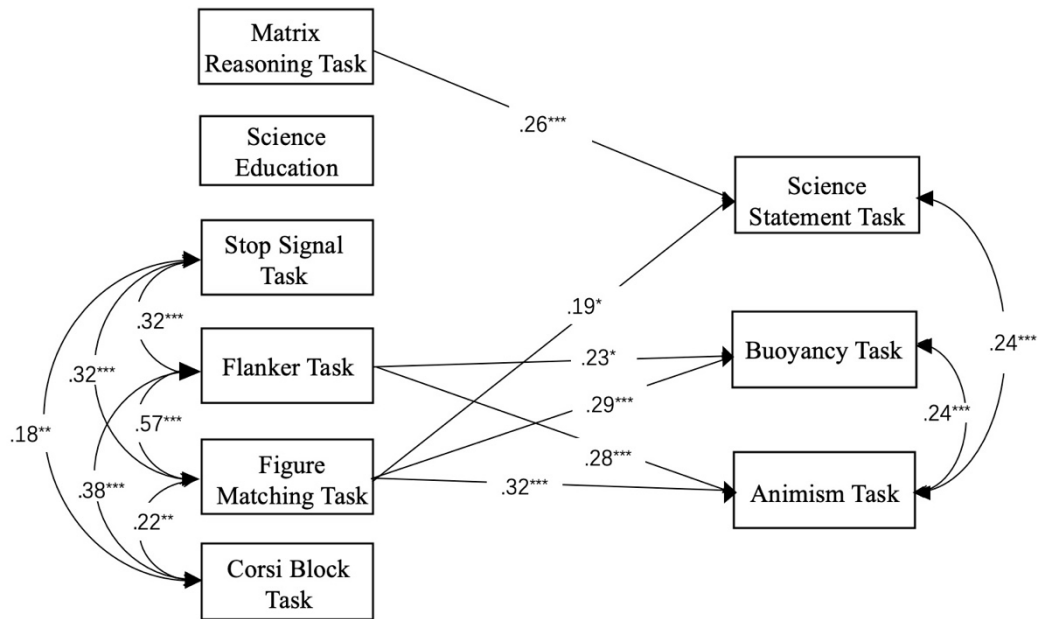
Path Analysis for Congruent trials on Efficiency (China)



Notes. Standardised Path Coefficients were reported. * $p < .05$; ** $p < .01$; *** $p < .001$.

Figure 6.3c

Path Analysis for Incongruent trials on Efficiency (China)



Notes. Standardised Path Coefficients were reported. * $p < .05$; ** $p < .01$; *** $p < .001$.

Table 6.12

Path Analysis Model Fit Indices Summary on Efficiency (China)

Trials	$\chi^2(df)$	CFI	TLI	AIC	BIC	RMSEA	SRMR
Overall	.04(1)	1.00	1.11	5661.19	5844.32	.00	.00
Congruent	.01(1)	1.00	1.13	5682.90	5866.04	.00	.00
Incongruent	.07(1)	1.00	1.11	5684.02	5867.15	.00	.00

Notes. Significant χ^2 was presented with *; * $p < .05$; ** $p < .01$; *** $p < .001$; χ^2 = Chi-square; df = degree of freedom; CFI = Comparative Fit Indices; TLI = Tucker and Lewis' Index; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; RMSEA = Root Mean Square of Approximation; SRMR = Standardized Root Mean Square; Robust CFI; TLI; RMSEA and SRMR were reported in the Table.

Appendix N: Efficiency Analysis (Combined Data)

Table N1

Descriptive statistics for Efficiency (Combined Data)

Variables	<i>M</i>	<i>SD</i>	<i>Skewness</i>	<i>Kurtosis</i>
General Cognitive Ability	0.08	0.06	8.23	92.63
Executive Function				
Stop Signal Task	1.48	0.53	0.17	-0.85
Figure Matching Task	1.10	0.24	0.06	0.00
Fish Flanker Task	2.14	0.37	-1.29	3.17
Corsi Block Task	0.23	0.09	0.65	1.52
Explanatory Coexistence				
Buoyancy Task				
Overall	1.43	0.39	-0.27	0.23
Congruent	1.49	0.41	-0.22	0.13
Incongruent	1.36	0.42	-0.53	0.54
Animism Task				
Overall	1.47	0.30	-0.56	0.20
Congruent	1.55	0.32	-0.55	0.26
Incongruent	1.40	0.30	-0.52	0.11
Science Statement Task				
Overall	0.30	0.11	1.55	8.88
Congruent	0.39	0.14	1.06	4.66
Incongruent	0.23	0.11	3.77	29.93

Table N2*Age adjustment test for Efficiency (Combined Data)*

Variables	Linear	Quadratic	Cubic
General Cognitive Ability	.01	.01	.00
Executive Function skills			
Stop Signal Task	.00	.00	.00
Figure Matching Task	.01*	.01*	.02*
Fish Flanker Task	.01*	.00	.01
Corsi Block Task	.02**	.02**	.02*
Explanatory Coexistence			
Buoyancy Task			
Overall	-.00	-.00	-.00
Congruent	-.00	-.00	-.00
Incongruent	-.00	-.00	-.00
Animism Task			
Overall	.00	-.00	-.00
Congruent	-.00	-.00	-.00
Incongruent	.00	.00	.00
Science Statement Task			
Overall	.00	.00	.00
Congruent	-.00	-.00	-.00
Incongruent	.00	.00	.00

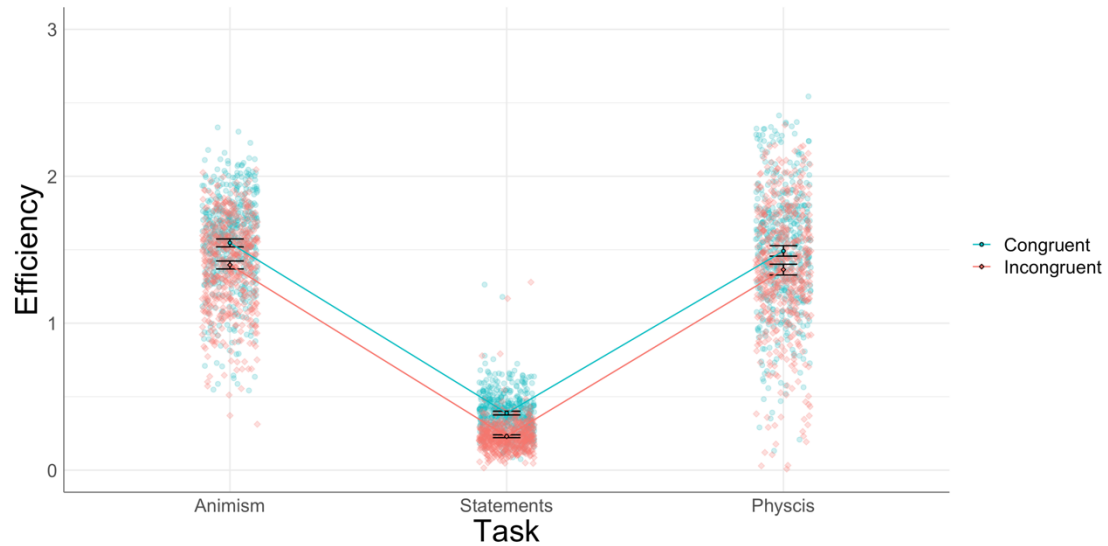
Notes. * $p < .05$; ** $p < .01$; *** $p < .001$.

Table N3*Shapiro-Wilk test for Efficiency (Combined Data)*

Variables	<i>Shapiro-Wilk</i>	<i>p</i>
General Cognitive Ability	.43	< .001
Executive Function		
Stop Signal Task	.98	< .001
Figure Matching Task	1.00	.55
Fish Flanker Task	.92	< .001
Corsi Block Task	.98	< .001
Explanatory Coexistence		
Buoyancy Task		
Overall	.99	.01
Congruent	.99	.02
Incongruent	.98	< .001
Animism Task		
Overall	.98	< .001
Congruent	.98	< .001
Incongruent	.98	< .001
Science Statement Task		
Overall	.92	< .001
Congruent	.95	< .001
Incongruent	.76	< .001

Figure N1

Mean Comparison of Efficiency between Congruent and Incongruent Trials for Naïve Thinking Tasks (Combined)



Notes. In efficiency analysis, significant differences between congruent and incongruent trials were observed in the Animism task ($t = 18.34, p < .001$), the Buoyancy task ($t = 9.79, p < .001$), and the Science Statement task ($t = 36.40, p < .001$); Thanks to Clemens Brunner for sharing the R script used to create this figure in Stricker et al. (2021).

Table N4*Multicollinearity test for Efficiency (Combined Data)*

	Overall		Congruent		Incongruent	
	Tolerance	VIF	Tolerance	VIF	Tolerance	VIF
Regression on Animism						
Science Education	.98	1.02	.98	1.02	.98	1.02
General Cognitive Ability	.98	1.02	.98	1.02	.98	1.02
Stop Signal Task	.85	1.17	.85	1.17	.85	1.17
Figure Matching Task	.66	1.51	.66	1.51	.66	1.51
Fish Flanker Task	.59	1.68	.59	1.68	.59	1.68
Corsi Block Task	.80	1.24	.80	1.24	.80	1.24
Regression on Buoyancy						
Science Education	.98	1.02	.98	1.02	.98	1.02
General Cognitive Ability	.98	1.02	.98	1.02	.98	1.02
Stop Signal Task	.85	1.17	.85	1.17	.85	1.17
Figure Matching Task	.66	1.51	.66	1.51	.66	1.51
Fish Flanker Task	.59	1.68	.59	1.68	.59	1.68
Corsi Block Task	.80	1.24	.80	1.24	.80	1.24
Regression on Science Statement						
Science Education	.98	1.02	.98	1.02	.98	1.02
General Cognitive Ability	.98	1.02	.98	1.02	.98	1.02
Stop Signal Task	.85	1.17	.85	1.17	.85	1.17
Figure Matching Task	.66	1.51	.66	1.51	.66	1.51
Fish Flanker Task	.59	1.68	.59	1.68	.59	1.68
Corsi Block Task	.80	1.24	.80	1.24	.80	1.24

Table N5*Homoscedasticity test for Efficiency (Combined Data)*

	<i>GQ</i>	<i>p</i>
Regression on Animism for Overall trials	1.36	.001
Regression on Animism Congruent trials	1.23	.05
Regression on Animism Incongruent trials	1.27	.03
Regression on Buoyancy Overall trials	0.87	.85
Regression on Buoyancy Congruent trials	0.92	.74
Regression on Buoyancy Incongruent trials	0.79	.96
Regression on Science Statement Overall trials	0.73	.99
Regression on Science Statement Congruent trials	0.74	.99
Regression on Science Statement Incongruent trials	0.44	.99

Table N6*Efficiency Confirmatory Factor Analysis Model Fit Indices (Combined)*

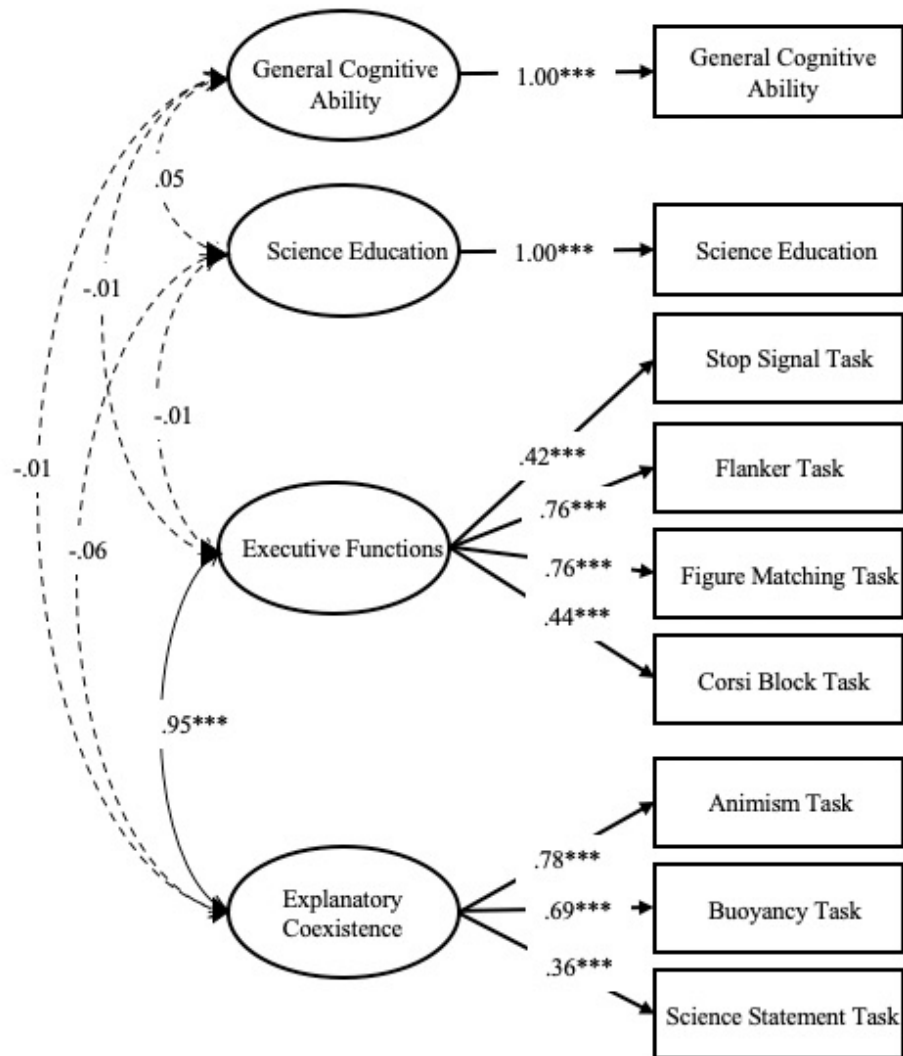
	$\chi^2(df)$	<i>CFI</i>	<i>TLI</i>	<i>AIC</i>	<i>BIC</i>	<i>RMSEA</i>	<i>SRMR</i>
Overall	51.66**	.95	.92	11780.23	11910.64	.07	.04
Congruent	51.30***	.95	.92	11811.73	11942.14	.06	.04
Incongruent	59.01***	.93	.89	11896.19	12026.60	.07	.05

Notes. Significant χ^2 was presented with *; * $p < .05$; ** $p < .01$; *** $p < .001$; χ^2 = Chi-square; *df* = degree of freedom; *CFI* = Comparative Fit Indices; *TLI* = Tucker and Lewis' Index; *AIC* = Akaike Information Criterion; *BIC* = Bayesian Information Criterion; *RMSEA* = Root Mean Square of Approximation; *SRMR* = Standardized

Root Mean Square; Robust *CFI*; *TLI*; *RMSEA* and *SRMR* were reported in the Table.

Figure N2

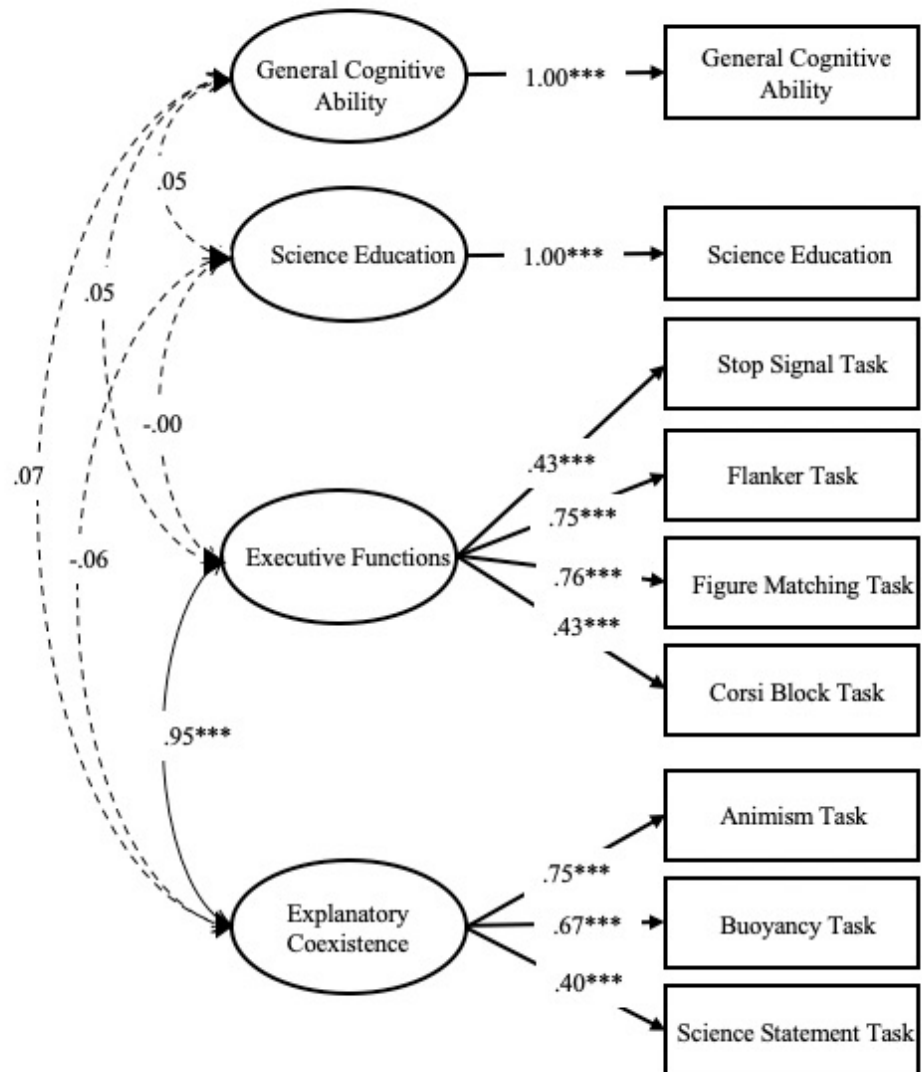
Efficiency for Overall Trials (Combined)



Notes. Standardised Path Coefficients were reported. * $p < .05$; ** $p < .01$; *** $p < .001$.

Figure N3

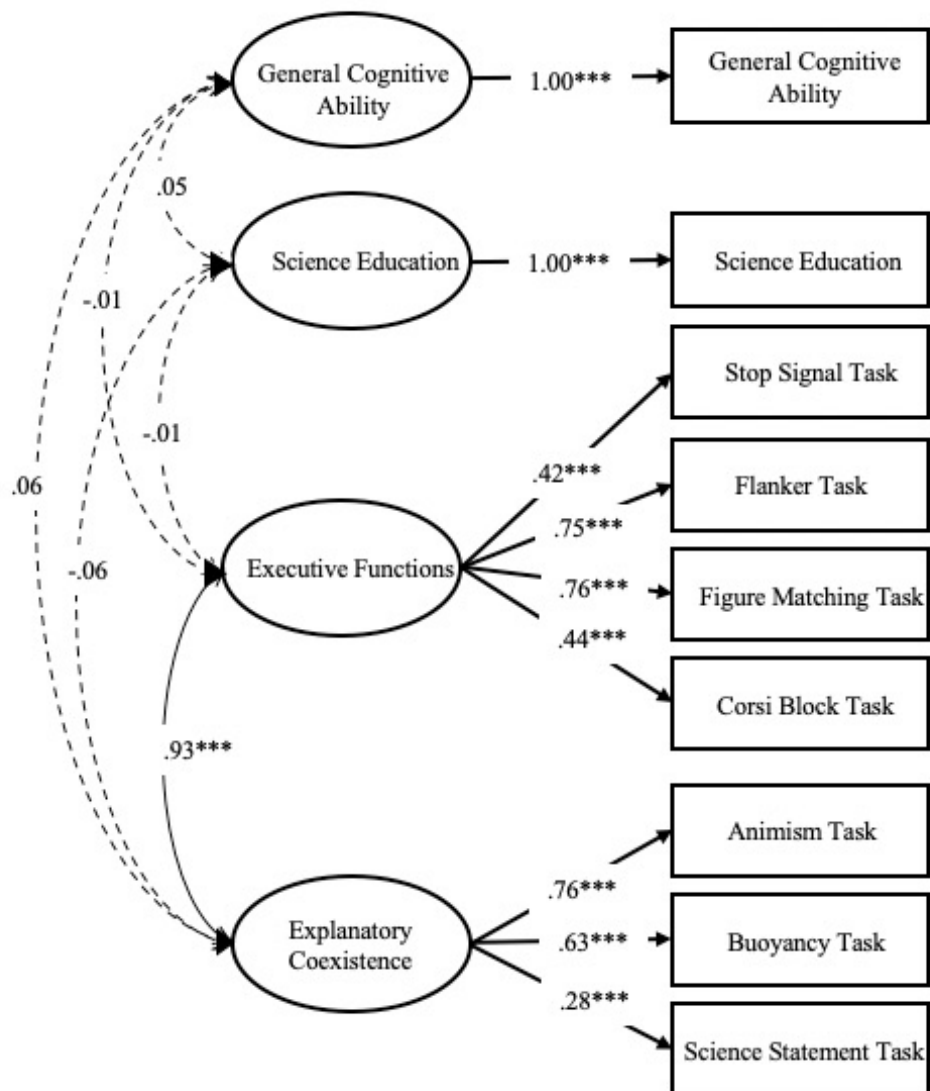
Efficiency for Congruent Trials (Combined)



Notes. Standardised Path Coefficients were reported. * $p < .05$; ** $p < .01$; *** $p < .001$.

Figure N4

Efficiency for Incongruent Trials (Combined)

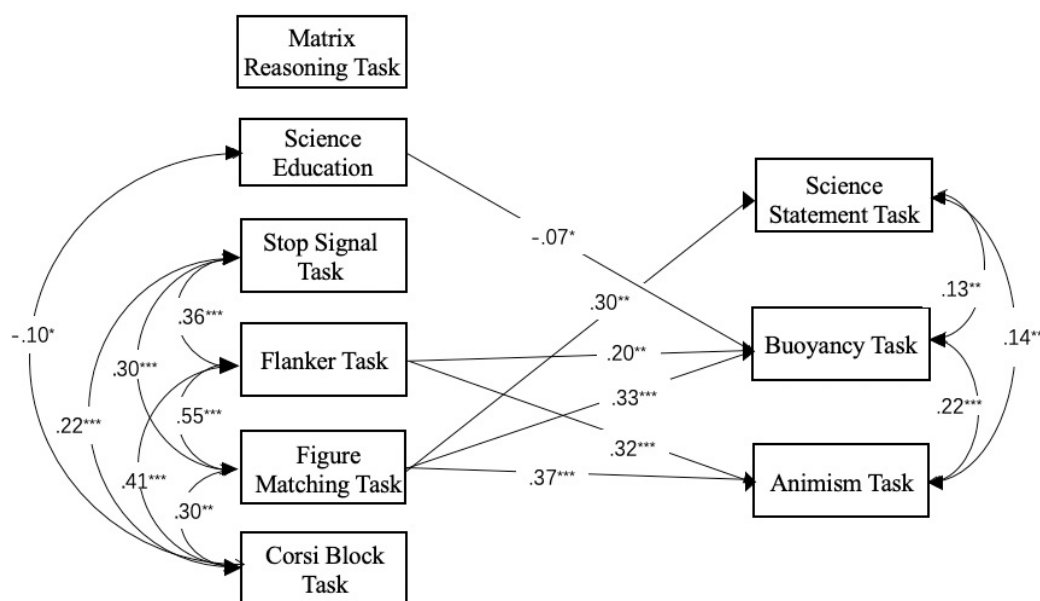


Notes. Standardised Path Coefficients were reported. * $p < .05$; ** $p < .01$; *** $p < .001$.

In Figure N5, I presented the path model analysis of incongruent trials using combined data. The results were similar to those found among the United Kingdom participants, where interference inhibition and switching predicted the Animism task; Science Education (negatively), interference inhibition, and switching predicted the Buoyancy task, with only switching predicting the Science Statement task. In the main text's results section, due to no significant measurement invariance analysis among overall trials and congruent trials, I reported the Chinese data for incongruent trials separately. When reported separately, interference inhibition and switching predicted both the Animism and the Buoyancy tasks; general cognitive ability and switching predicted the Science Statement task.

Figure N5

Path Analysis for Incongruent trials on Efficiency (Combined)

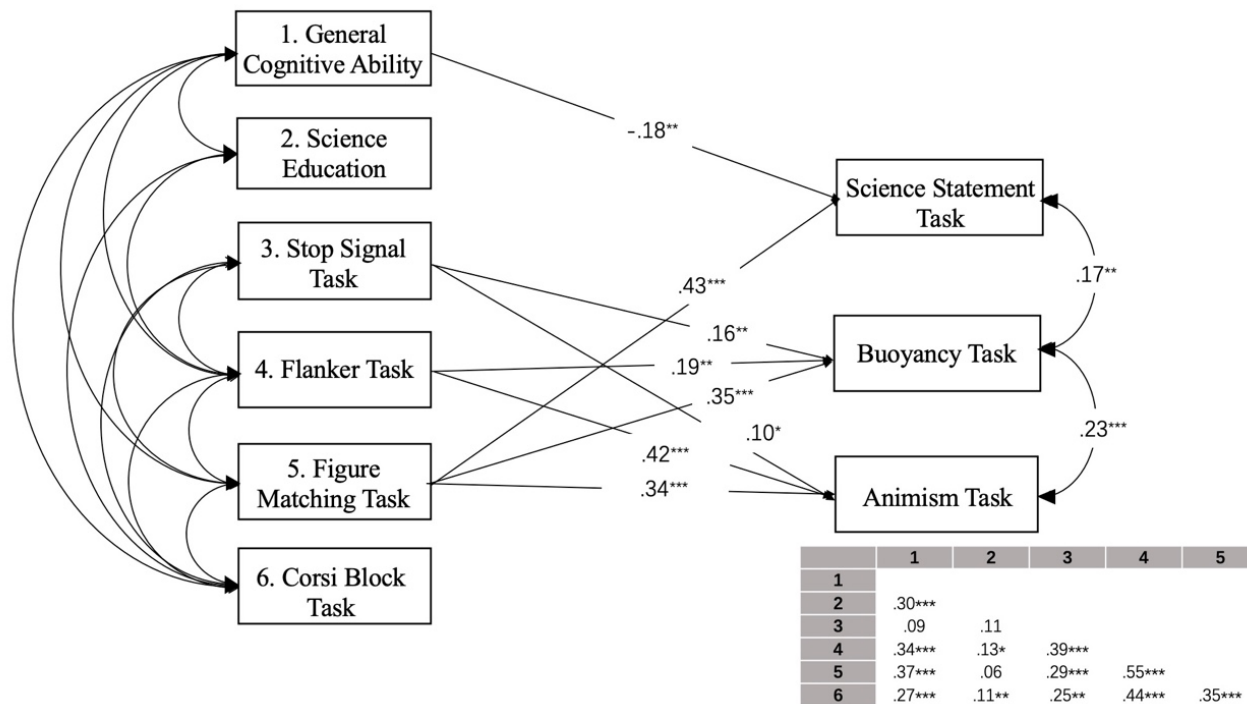


Notes. $\chi^2(df) = 0.14(1)$, $CFI = 1.00$, $TLI = 1.04$, $AIC = 11848.68$, $BIC = 12071.63$, $RMSEA = .00$, $SRMR = .00$; Standardised Path Coefficients were reported; * $p < .05$; ** $p < .01$; *** $p < .001$

Appendix O: Accuracy of General Cognitive Ability and Efficiency for All Other Variables

Figure O1

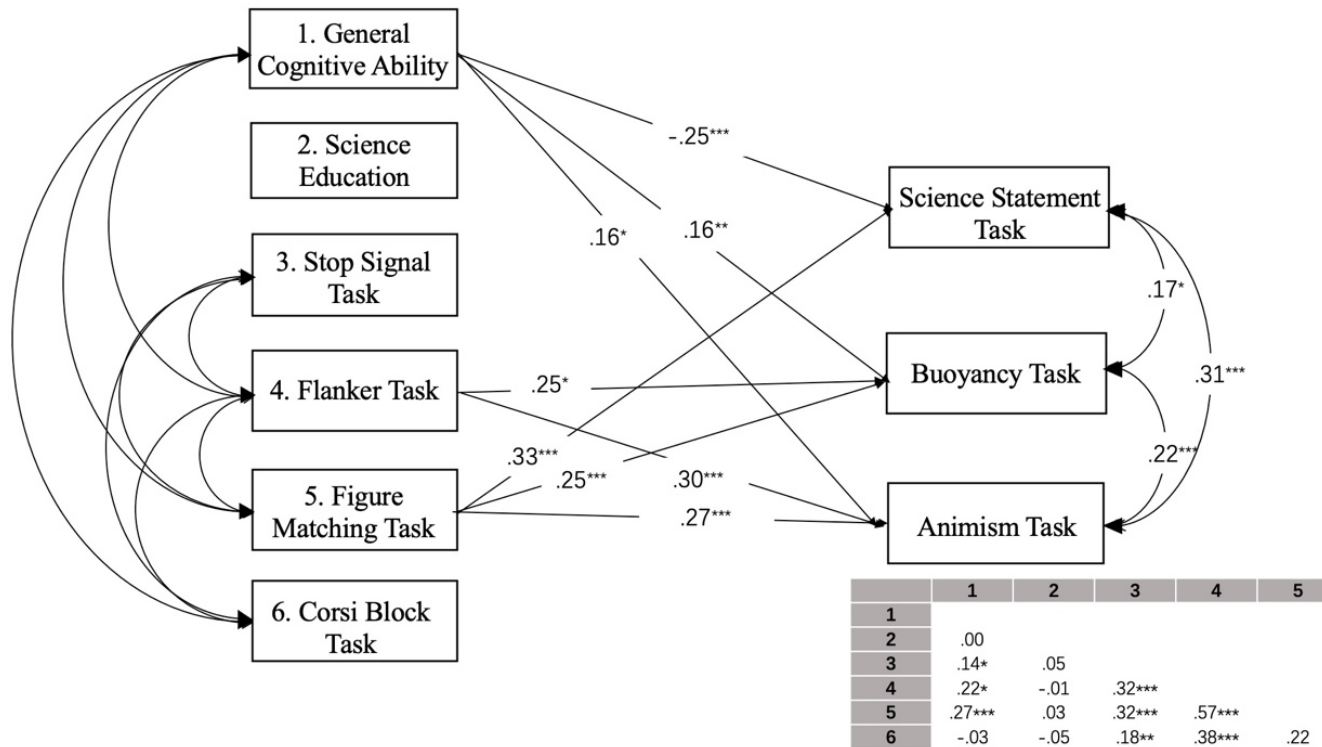
Path Analysis of Overall Trials (the United Kingdom)



Notes. Standardised Path Coefficients were reported. $*p < .05$; $**p < .01$; $***p < .001$.

Figure O2

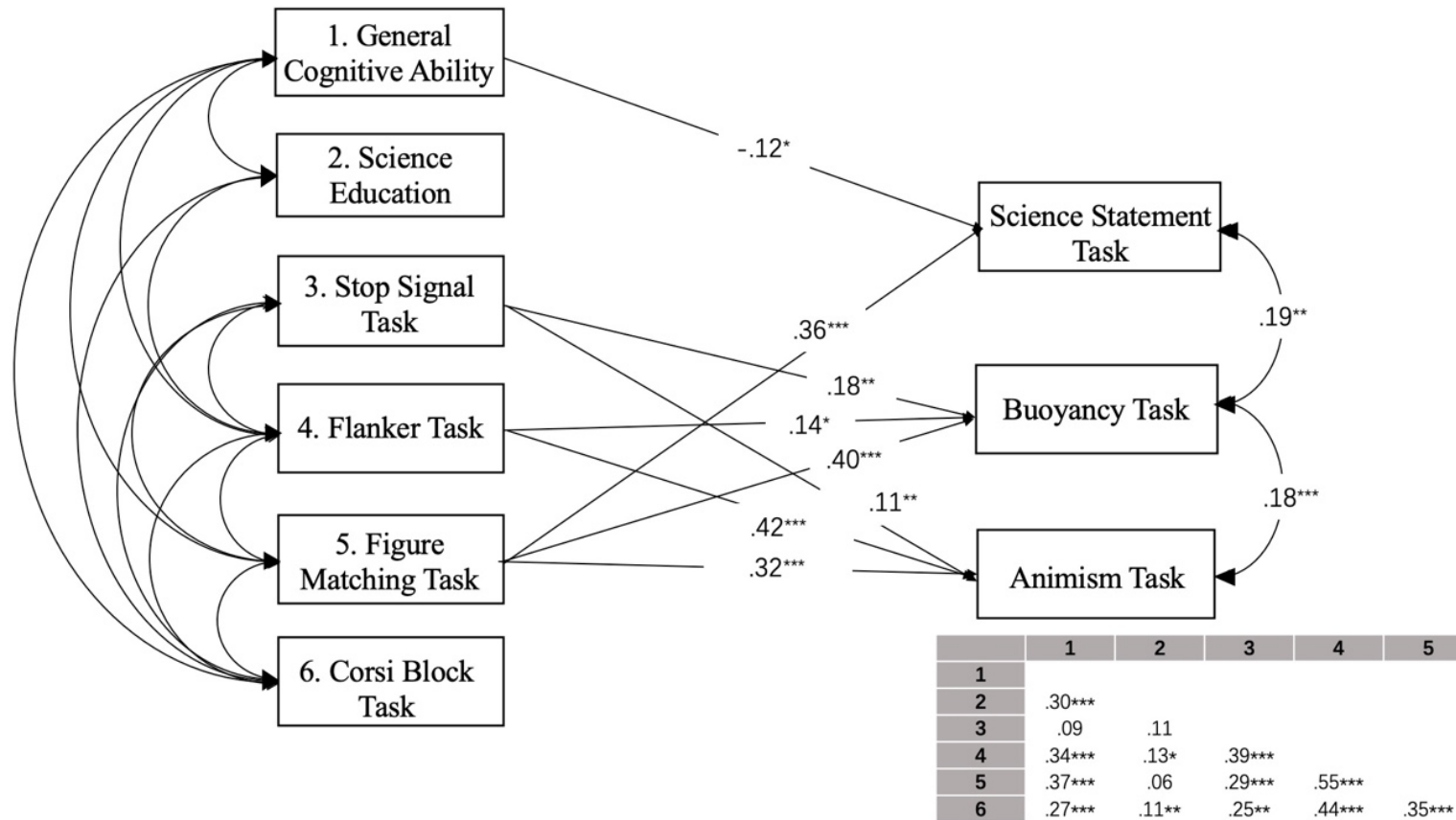
Path Analysis of Overall Trials(China)



Notes. Standardised Path Coefficients were reported. * $p < .05$; ** $p < .01$; *** $p < .001$.

Figure O3

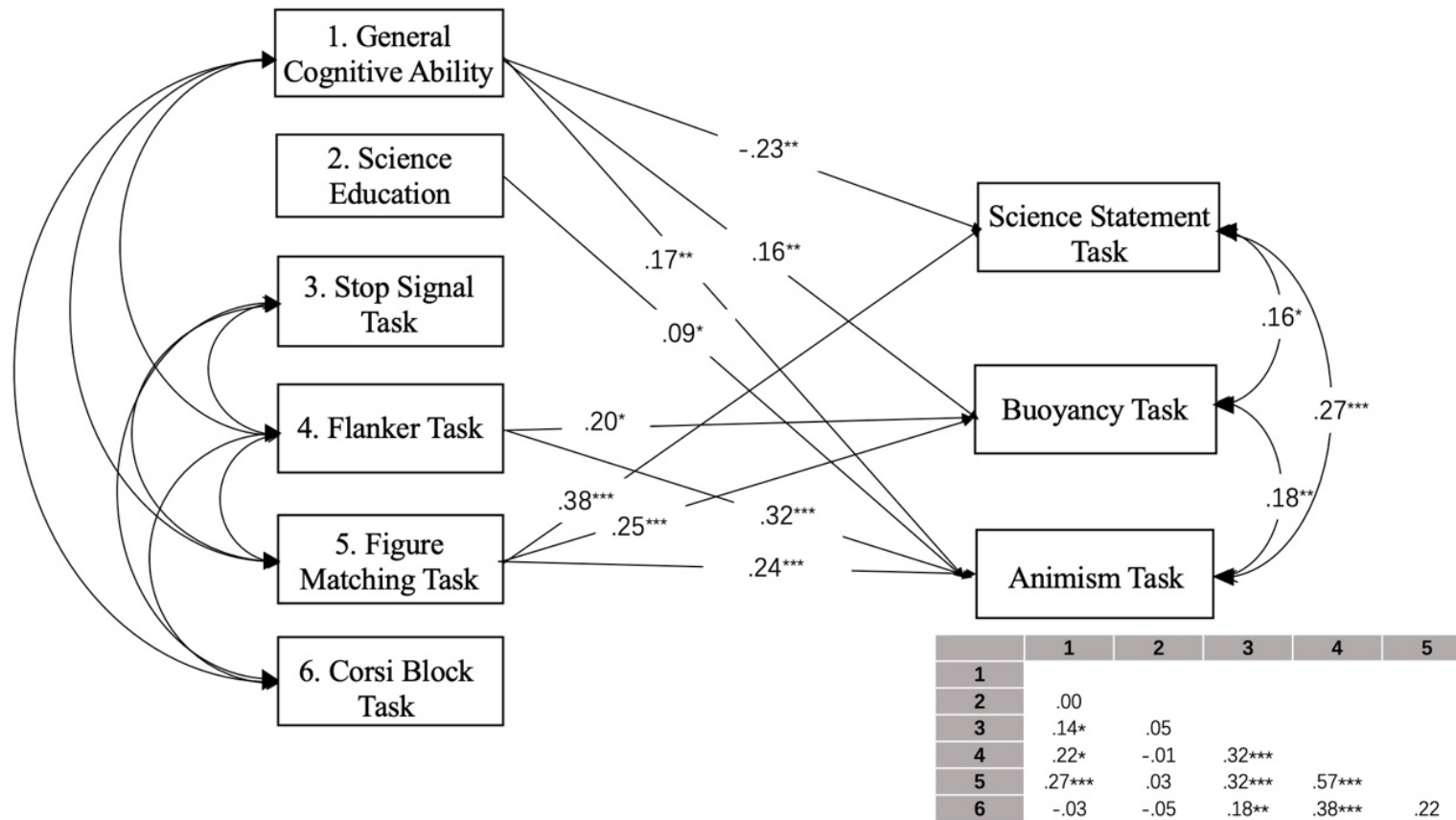
Path Analysis of Congruent Trials (the United Kingdom)



Notes. Standardised Path Coefficients were reported. $*p < .05$; $**p < .01$; $***p < .001$.

Figure O4

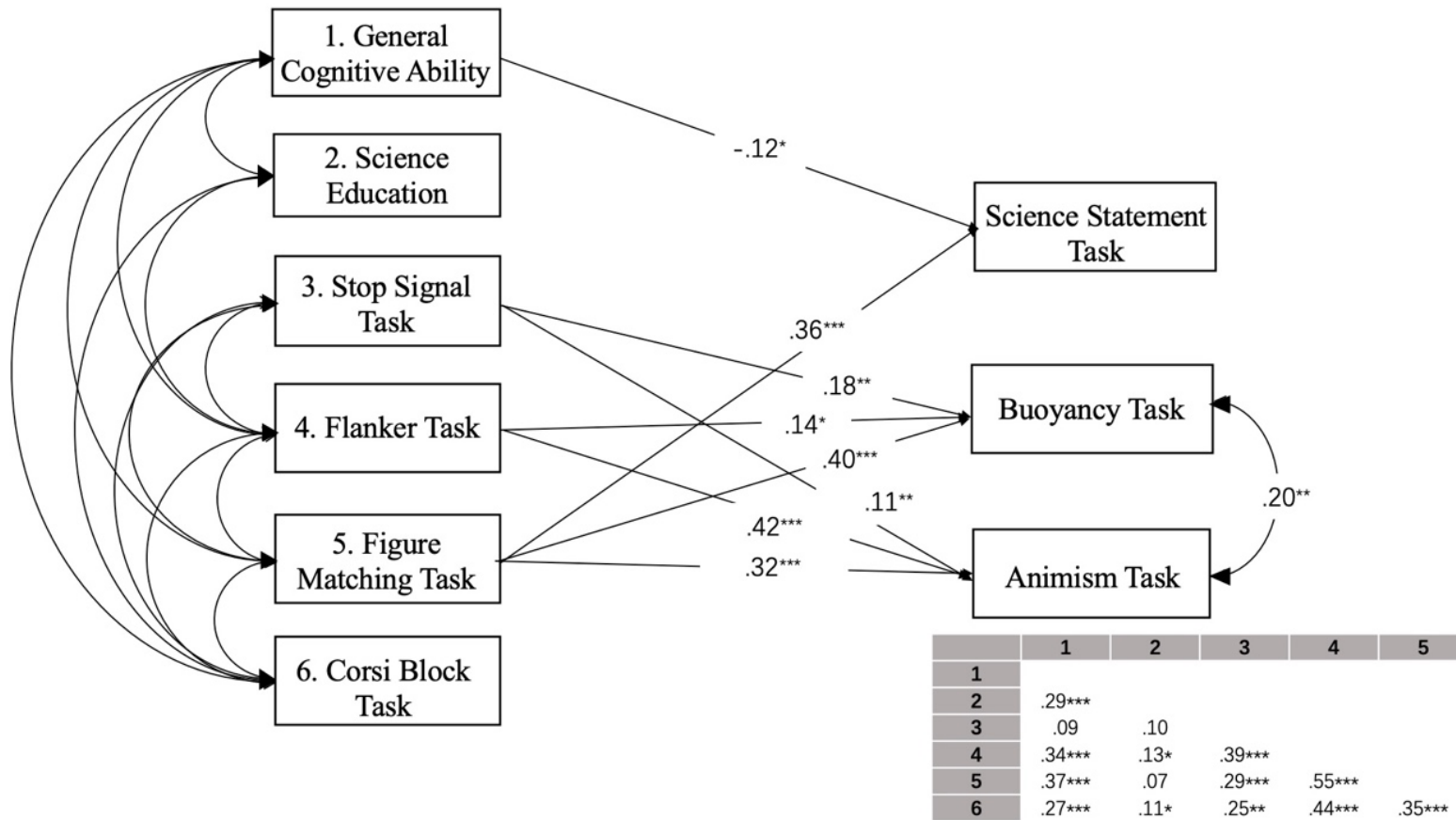
Path Analysis of Congruent Trials(China)



Notes. Standardised Path Coefficients were reported. * $p < .05$; ** $p < .01$; *** $p < .001$.

Figure O5

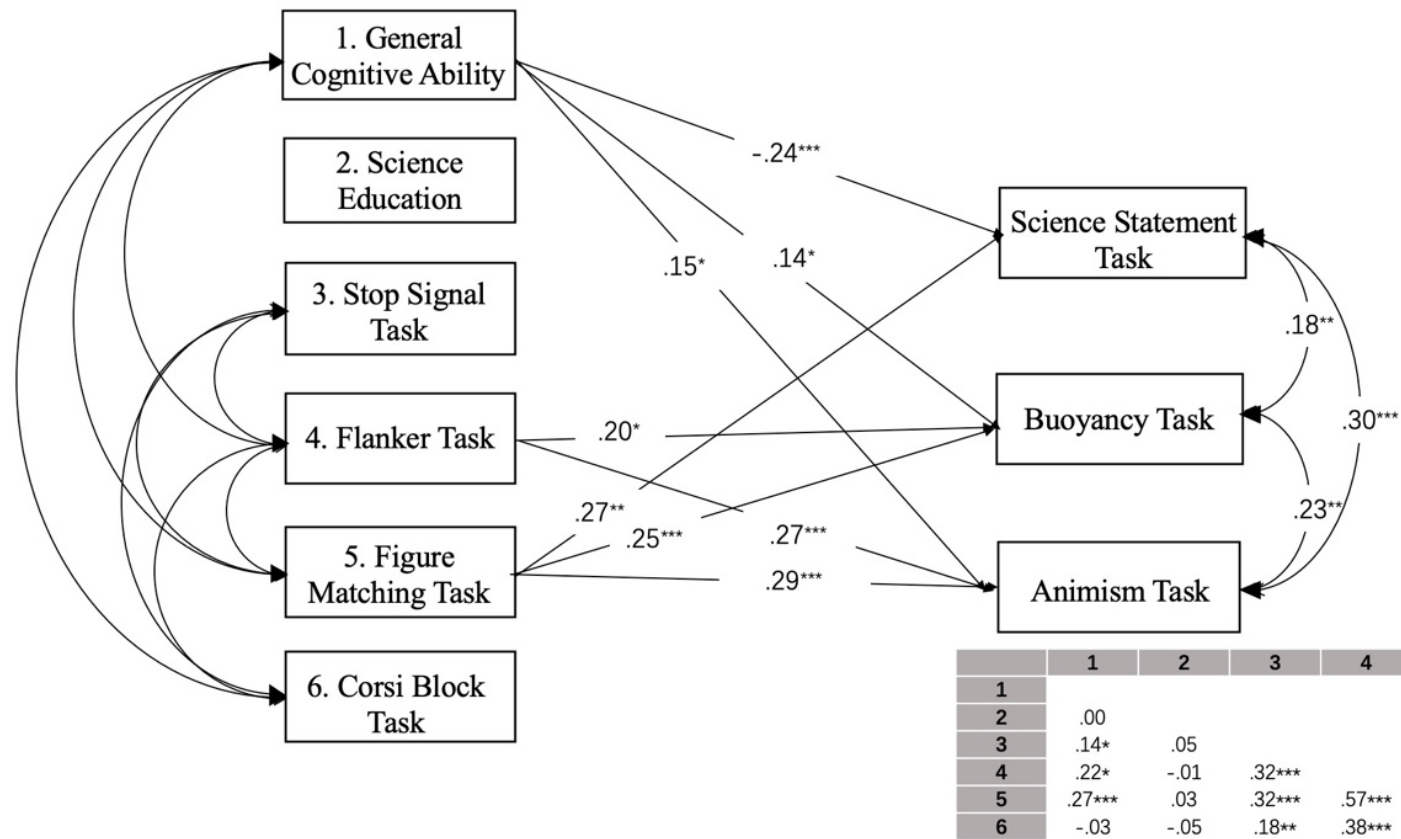
Path Analysis of Incongruent Trials (the United Kingdom)



Notes. Standardised Path Coefficients were reported. $*p < .05$; $**p < .01$; $***p < .001$.

Figure O6

Path Analysis of Incongruent Trials (China)



Notes. Standardised Path Coefficients were reported. $*p < .05$; $**p < .01$; $***p < .001$.

Table O1*Path Analysis Model Fit Indices Summary on Efficiency (the United Kingdom)*

Trials	$\chi^2(df)$	<i>CFI</i>	<i>TLI</i>	<i>AIC</i>	<i>BIC</i>	<i>RMSEA</i>	<i>SRMR</i>
the United Kingdom							
Overall	.47(1)	1.00	1.04	5233.33	5422.45	.00	.00
Congruent	.25(1)	1.00	1.06	5266.09	5455.21	.00	.00
Incongruent	.52(1)	1.00	1.04	5303.66	5492.79	.00	.01
China							
Overall	.04(1)	1.00	1.10	5621.01	5804.14	.00	.00
Congruent	.05(1)	1.00	1.11	5644.09	5827.22	.00	.00
Incongruent	.03(1)	1.00	1.10	5651.80	5834.93	.00	.01

Notes. Significant χ^2 was presented with *; * $p < .05$; ** $p < .01$; *** $p < .001$; χ^2 = Chi-square; df = degree of freedom; *CFI* = Comparative Fit Indices; *TLI* = Tucker and Lewis' Index; *AIC* = Akaike Information Criterion; *BIC* = Bayesian Information Criterion; *RMSEA* = Root Mean Square of Approximation; *SRMR* = Standardized Root Mean Square; Robust *CFI*; *TLI*; *RMSEA* and *SRMR* were reported in the Table.