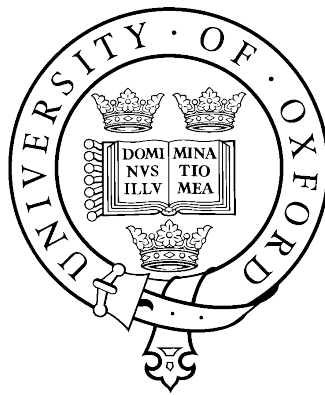


**The Role of Spatial Ability and Problem-Solving Strategies
in Early Arithmetic Word Problem Solving**



Jisoo Seo
St John's College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy
Hilary Term 2026

Supervisors: Professor Gabriel J. Stylianides and Professor Maria Tsitiridou Evangelou

Thesis Abstract

Arithmetic word problem solving in the early primary years is important, not least because it predicts later academic and career success. Factors such as language ability, numerical ability, and fluid reasoning have received considerable attention in research on young pupils' word problem solving. In contrast, spatial ability remains understudied as a potential contributor to word problem solving, despite extensive evidence of a spatial–mathematical link. Specifically, research with older pupils suggests that spatial ability may support word problem solving through the construction of visual representations, a form of problem-solving strategy. However, the extent to which problem-solving strategies are associated with spatial ability and mediate the relationship between spatial ability and word-problem-solving accuracy in young pupils (6–7-year-olds) remains underexplored.

This integrated thesis, comprising three studies, aims to contribute towards addressing these gaps by examining the relationship between young pupils' spatial skills (spatial visualization and visuospatial working memory) and their strategies and accuracy in solving one-step additive word problems.

Theoretically, it proposes a 2×2 conceptual framework for categorizing one-step additive word problems by integrating distinctions between quantity and quantitative relation, methods of quantification, and classifications of dynamic and static relations, yielding four word problem types. It also proposes a complementary framework for categorizing young pupils' problem-solving strategies by integrating arithmetic strategies and representations of quantitative structure, yielding five strategy types.

Empirically, Study 1, using cross-classified mixed-effects logistic regression analyses, showed that neither spatial visualization nor visuospatial working memory uniquely predicted overall problem-solving accuracy. However, the association between spatial visualization, but not visuospatial working memory, and accuracy varied across problem types.

Study 2, using cross-classified mixed-effects multinomial logistic regression analyses, showed that higher spatial visualization was consistently associated with greater relative use of 'Quantitatively Correct Number Fact' strategies compared with 'Quantitatively Incorrect Counting' and 'Quantitatively Correct Counting' strategies. In contrast, higher visuospatial working memory was associated with reduced relative use of 'Non-Quantitative' strategies across most problem types; for comparison problems involving disjoint quantities measured

using an imposed unit, it was instead more consistently associated with reduced relative use of ‘Quantitatively Incorrect Counting’ strategies.

Study 3, using path analyses, showed that spatial visualization had an indirect effect on accuracy via ‘Quantitatively Correct Number Fact’ strategy use for problems requiring coordination and transformation of representations of the problem’s quantitative structure. In contrast, visuospatial working memory showed an indirect effect for problem types requiring the maintenance of these representations.

Together, these findings advance understanding of the distinct and complementary roles of spatial visualization and visuospatial working memory in one-step additive word problem solving, and of their differing relevance based on the problem’s quantitative structure. They provide a nuanced account of these contributions through the proposed frameworks and offer insights into the relationship between spatial skills and arithmetic word problem solving, highlighting problem-solving strategies as a promising target for leveraging the spatial–mathematical link in early educational contexts.

Acknowledgements

Since embarking on this journey, I have always looked forward to the day I would be able to write this acknowledgements section. Now, that day has finally arrived. I am grateful for everything God Almighty has blessed me with. Without His saving grace and guiding hands, I know that I would have no place here, or anywhere in the world. Above all, I thank God for shaping this transformative journey with challenges that I could bear, grow through, and learn from, and for using this experience to mould me into someone I am deeply proud of.

I am profoundly grateful to my supervisors, Professors Gabriel Stylianides and Maria Evangelou. They provided invaluable academic guidance through their extensive expertise, feedback, and professional connections, as well as moral support at crucial moments, while giving me the intellectual freedom to pursue my own interests and develop my ideas independently. Their unwavering belief in me sustained me through every challenge.

I feel very fortunate to have had Professors Terezinha Nunes and Catherine Thevenot as my viva voce examiners; their inspiring work shaped my thesis, and our conversations were intellectually stimulating and genuinely enriching. Likewise, I thank my Transfer of Status and Confirmation of Status assessors, Professors Ariel Lindorff and Lars-Erik Malmberg, for reassuring me that I was on the right track while also challenging me to refine my work.

I also could not have completed this thesis without the headteachers, teachers, and children who welcomed me into their schools. While I cannot name them here for ethical reasons, I sincerely hope that their generosity with time and willingness to participate were not in vain.

I would like to thank my appa, Jaepil Seo, and umma, Younglim Oh, who have always modelled true courage, faith, love, and diligence. Without their sacrifices, I could never have dreamed of embarking on this journey. I also thank my oppa, Minwoo Seo, for his constant support; my unni, Hyewon Jung, for always being someone I could rely on; and my baby niece, Hailey Harin Seo, for featuring in countless photos and videos with her beautiful smiles, masterful roll-overs, and inspirational mukbangs, all of which gave me the boost I needed to reach the final stretch.

I feel blessed for all the friends I have made along this journey—you made this city feel like home. Thank you for sharing countless conversations, meals, laughter, tears, and spiritual connections with me; for listening to my research ideas, coding my data, and proofreading my drafts. Thank you for hosting me in your homes and for the writing retreats and library co-working sessions that carried me all the way to submission. Thank you for the travels, balls,

bops, formals, board game nights, movie nights, and karaoke sessions. Thank you for inspiring and witnessing my growth.

I must express my gratitude to Professors Joan Moss, Beverly Caswell, and Zack Hawes, with whom I had the privilege of working as a research officer for the Robertson Program for Inquiry-Based Teaching in Math and Science at the University of Toronto. Through this work, I developed an interest in understanding young children's mathematical thinking and came to appreciate the importance of conducting responsible and respectful mathematics education research. It was through this experience that I became passionate about developing and disseminating higher-quality, research-based mathematics curricula and pedagogical practices.

I am grateful to Somerville College and St John's College for the opportunity to serve as a Student Welfare Advisor. They not only made my university life financially possible but also gave me a sense of community and made me feel like a valued contributor. I am also grateful to the Department of Education for the teaching and research opportunities that enriched my DPhil experience and supported my professional development.

I am thankful to the University for its libraries, particularly Old Bodleian, Radcliff Camera, and Taylorian, which provided countless hours of study, writing, and reflection throughout my DPhil. In the final stretch of writing, however, the Schwarzman Humanities Library became my second home; it was there that the last two chapters of this thesis were written. When I typed the final full stop, I knew that this library deserved a place in my acknowledgements.

When I began this journey, I had no idea how dynamic and life-changing it would be. I gained far more than a degree, and I lost things I hold too close to my heart to write about in a place so permanent and public. This journey has been about discovering myself, remembering my worth, and embracing my shortcomings. I carry with me an enduring gratitude for the opportunity to have undergone such a humbling experience. With this chapter complete, I stand poised and resolute, ready to embrace the next. Jisoo, I am proud of you for the grace and resilience with which you navigated this journey. I love you for all that you are—and despite all that you are.

Table of Contents

Thesis Abstract.....	i
Acknowledgements	iii
Table of Contents	v
List of Tables	xii
List of Figures	xiv
1 General Introduction.....	1
1.1 Research Background	1
1.2 Research Gaps and Aim.....	2
1.3 Theoretical Foundations	5
1.3.1 The Distinction between Quantitative Reasoning and Arithmetic.....	5
1.3.2 The Distinction between Quantity and Quantitative Relation.....	6
1.3.3 The Distinction between Discrete Quantity and Continuous Quantity	7
1.3.4 The Distinction between Dynamic Relation and Static Relation	8
1.4 Key Concepts	9
1.4.1 One-Step Additive Word Problems.....	9
1.4.2 Problem-Solving Strategies.....	12
1.4.3 Spatial Ability.....	13
1.5 Research Themes and Questions.....	16
1.5.1 Research Theme 1: Spatial Skills and One-Step Additive Word-Problem- Solving Accuracy	16
1.5.2 Research Theme 2: Spatial Skills and Problem-Solving Strategies	17
1.5.3 Research Theme 3: Spatial Skills and One-Step Additive Word-Problem- Solving Accuracy: The Mediating Role of Problem-Solving Strategies	17
1.6 Educational Relevance	17
1.7 Thesis Structure.....	20
2 Methodology	23
2.1 Research Design.....	23
2.2 Participants and Recruitment	23
2.2.1 A Priori Power Analysis.....	24
2.2.2 Recruitment Process	24
2.2.3 School Characteristics	25
2.2.4 Participants	26

2.3 Materials and Procedures.....	27
2.3.1 Overarching Reliability and Validity Considerations	30
2.3.2 Established Measures	32
2.3.2.1 Spatial Ability.....	32
2.3.2.1.1 Spatial Visualization	32
2.3.2.1.2 Visuospatial Working Memory.....	35
2.3.2.2 Language Ability	37
2.3.2.3 Numerical Ability	37
2.3.2.4 Fluid Reasoning.....	38
2.3.3 One-Step Additive Word Problems.....	39
2.3.3.1 Development of the Word Problems	39
2.3.3.1.1 First Pilot Study.....	40
2.3.3.1.2 Second Pilot Study	41
2.3.3.1.3 Selection of Number Facts	42
2.3.3.1.4 Randomization of Number Facts into Four Alternative Forms.....	44
2.3.3.2 Final Word Problem Set	45
2.3.3.3 Administration Procedure.....	51
2.3.3.4 Coding of Problem-Solving Strategies	54
2.3.3.4.1 Coding of Representational Tools and Arithmetic Approaches.....	54
2.3.3.4.2 Derivation of Problem-Solving Strategy Codes.....	57
2.3.3.4.3 Inter-Rater Reliability	58
2.3.3.5 Reliability and Validity of Word Problems.....	58
2.3.3.5.1 Internal Consistency of each Test Form.....	59
2.3.3.5.2 Compatibility of the Four Test Forms.....	59
2.3.3.5.3 Reliability of the Combined Word Problems Measure	61
2.3.3.5.4 Structural Validity of Conceptual Framework	63
2.4 Summary of Analytic Approach	67
2.5 Ethical Considerations	73
2.6 Summary of Methodology Chapter	75
3 Study 1: The Role of Spatial Skills in Young Children’s Arithmetic Word-Problem-	
Solving Accuracy	76
Abstract	76
3.1 Introduction	77
3.2 Literature Review and Conceptual Framework.....	78
3.2.1 First Challenge: Diversity of Spatial Taxonomies and Measures	80

3.2.2 Second Challenge: Limited Use of Structurally Grounded Word Problem	
Categorization in Spatial–Mathematical Research	82
3.2.3 Theoretical Foundations	85
3.2.3.1 Quantity versus Quantitative Relation.....	85
3.2.3.2 Subdivision of Quantity: Discrete versus Continuous.....	86
3.2.3.3 Subdivision of Quantitative Relation: Dynamic versus Static	87
3.2.4 Spatial Ability and Quantitative Structure	89
3.2.5 A 2×2 Conceptual Framework for Categorizing One-Step Additive Word Problems	91
3.3 Methods	93
3.3.1 Participants	93
3.3.2 Measures.....	94
3.3.2.1 Word Problems Task	94
3.3.2.2 Spatial Measures.....	99
3.3.2.3 Covariate Measures	100
3.3.3 Procedure.....	100
3.3.4 Analytic Approach	102
3.4 Findings	103
3.4.1 Preliminary Analyses	104
3.4.2 RQ1: The Relationship between Spatial Skills and Word-Problem-Solving Accuracy	106
3.4.3 RQ2: Interaction Effects of Spatial Skills and Problem Type on Word-Problem-Solving Accuracy	107
3.5 Discussion	110
3.5.1 RQ1: Interpreting the Relationship between Spatial Skills and Word-Problem-Solving Accuracy	111
3.5.2 RQ2: Interpreting Interaction Effects of Spatial Skills and Problem Type on Word-Problem-Solving Accuracy	114
3.5.2.1 Interaction of Spatial Visualization and Problem Type in Predicting Word-Problem-Solving Accuracy.....	114
3.5.2.2 Interaction of Visuospatial Working Memory and Problem Type in Predicting Word-Problem-Solving Accuracy.....	116
3.5.3 Limitations.....	117
3.5.4 Conclusion.....	119
4 Study 2: The Role of Spatial Skills in Young Children’s Arithmetic Word-Problem-Solving Strategies	120

Abstract	120
4.1 Introduction	121
4.2 Literature Review.....	122
4.2.1 Spatial Ability and Visual Representations in Secondary School Pupils.....	123
4.2.2 Spatial Ability and Additive Word-Problem-Solving Strategies in Lower Primary School Pupils.....	125
4.2.2.1 Arithmetic Strategies: Counting-Based versus Number-Fact Strategies.....	126
4.2.2.2 Quantitative Reasoning: Representation of Quantitative Structure.....	130
4.2.2.3 Problem-Solving Strategies: A Holistic Perspective	132
4.2.3 Word Problem Type as a Moderator of the Spatial–Strategy Relationship	133
4.2.4 Present Study	137
4.3 Methods	138
4.3.1 Participants	138
4.3.2 Measures.....	138
4.3.3 Procedure.....	141
4.3.4 Coding of Problem-Solving Strategies.....	142
4.3.5 Analytic Approach	147
4.4 Findings	148
4.4.1 RQ3: The Relationship between Spatial Skills and Problem-Solving Strategy Use	149
4.4.1.1 Preliminary Analyses.....	149
4.4.1.1.1 Problem-Solving Strategy Distribution.....	149
4.4.1.1.2 Between-Pupil Variation in Problem-Solving Strategy Distribution.....	150
4.4.1.1.3 Between-Item Variation in Problem-Solving Strategy Distribution	150
4.4.1.1.4 Descriptive Statistics and Correlations	151
4.4.1.2 Inferential Analyses (Main-Effects Model).....	153
4.4.2 RQ4: Interaction Effects of Spatial Skills and Problem Type on Problem- Solving Strategy Use.....	158
4.4.2.1 Descriptive Statistics	158
4.4.2.2 Inferential Analyses (Interaction Model)	159
4.4.2.2.1 Spatial Visualization and Problem-Solving Strategy by Problem Type	159
4.4.2.2.2 Visuospatial Working Memory and Problem-Solving Strategy by Problem Type..	164
4.5 Discussion	166
4.5.1 RQ3: Interpreting the Relationship between Spatial Skills and Problem-Solving Strategy Use	166
4.5.1.1 Spatial Skills and Arithmetic Strategy.....	167

4.5.1.2 Spatial Skills and Representation of Quantitative Structure	169
4.5.2 RQ4: Interpreting Interaction Effects of Spatial Skills and Problem Type on Problem-Solving Strategy Use.....	171
4.5.2.1 Patterns across Problem Types.....	172
4.5.2.2 Continuous Quantity–Static Relation Problems: A Distinct Pattern	174
4.5.3 Limitations.....	176
4.5.4 Conclusion.....	178
5 Study 3: Young Children’s Problem-Solving Strategies as a Mediator between Spatial Skills and Arithmetic Word-Problem-Solving Accuracy	180
Abstract	180
5.1 Introduction	181
5.2 Literature Review	182
5.2.1 Arithmetic Word Problems	182
5.2.1.1 Quantitative Reasoning.....	183
5.2.1.2 Arithmetic Skills.....	189
5.2.1.3 Spatial Ability and Problem-Solving Strategies	193
5.2.2 Path Analysis.....	195
5.2.3 Present Study	199
5.3 Methods	200
5.3.1 Participants	200
5.3.2 Measures.....	201
5.3.3 Procedure.....	202
5.3.4 Analytic Approach	202
5.4 Findings	204
5.4.1 RQ5: Indirect Effects of Spatial Skills across Word Problem Types.....	204
5.4.1.1 Preliminary Analyses.....	204
5.4.1.1.1 Descriptive Statistics and Correlations	204
5.4.1.1.2 Assumption Checks and Data Structure.....	207
5.4.1.2 Path Analysis.....	208
5.4.1.3 Sensitivity Analysis	210
5.4.2 RQ6: Indirect Effects of Spatial Skills by Word Problem Type	211
5.4.2.1 Descriptive Statistics by Word Problem Type	211
5.4.2.2 Path Analyses by Word Problem Type.....	212
5.5 Discussion	215
5.5.1 RQ5: Interpreting Indirect Effects of Spatial Skills across Word Problem Types.....	215

5.5.2 RQ6: Interpreting Indirect Effects of Spatial Skills by Word Problem Type	216
5.5.2.1 Spatial Visualization Supports the Coordination of Representation	217
5.5.2.2 Visuospatial Working Memory Supports the Maintenance of Representation	218
5.5.2.3 Limits of Strategy Mediation in Discrete Quantity–Dynamic Relation Problems	221
5.5.3 Educational Implications	221
5.5.4 Limitations	223
5.5.5 Conclusion	225
6 General Discussion	226
6.1 Summary of Theoretical Contributions	226
6.2 Summary of Empirical Findings	227
6.2.1 Findings of Research Theme 1: Spatial Skills and One-Step Additive Word- Problem-Solving Accuracy	227
6.2.2 Findings of Research Theme 2: Spatial Skills and Problem-Solving Strategies	228
6.2.3 Findings of Research Theme 3: Spatial Skills and One-Step Additive Word- Problem-Solving Accuracy: The Mediating Role of Problem-Solving Strategies	228
6.3 Synthesis of Findings	230
6.3.1 The Quantitative Structure of the Problem Matters	230
6.3.2 The Type of Spatial Skill Matters	231
6.3.3 The Problem-Solving Strategy Use Matters	232
6.3.4 Contribution to Literature	234
6.4 Educational Implications	234
6.5 Limitations and Suggestions for Future Research	237
6.5.1 Theoretical Limitations and Related Suggestions for Future Research	238
6.5.2 Methodological Limitations and Related Suggestions for Future Research	239
6.6 Concluding Remarks	241
References	243
Appendix 2A: Headteacher Initial Recruitment Email	276
Appendix 2B: Headteacher Follow-Up Recruitment Email	277
Appendix 2C: Headteacher information Sheet and Consent Form	278
Appendix 2D: Teacher Information Sheet and Consent Form	284
Appendix 2E: Parent Information Sheet and Opt-Out Consent Form (Extended Version)	289
Appendix 2F: Parent Information Sheet and Opt-Out Consent Form (Simplified Version)	295
Appendix 2G: Parent Information Sheet and Opt-In Consent Form (Extended Version)	296

Appendix 2H: Parent Information Sheet and Opt-In Consent Form (Simplified Version)	303
Appendix 2I: Item Difficulty and Discrimination Indices	304
Appendix 2J: Standardized Factor Loadings of Confirmatory Factor Analyses.....	305
Appendix 2K: Central University Research Ethics Committee Approval.....	309
Appendix 2L: Child Assent Script	310
Appendix 3A: Assumption Diagnostics for Linearity of the Logit.....	311
Appendix 4A: Assumption Diagnostics of Mixed-Effects Multinomial Logistic Regression Models	313
Appendix 4B: RQ3 Predicted Probabilities of Problem-Solving Strategies as Nonlinear Functions of Numerical Ability	317
Appendix 4C: Discrete Quantity–Dynamic Relation Problems the Key Stage 1 National Curriculum Assessments	318
Appendix 4D: Discrete Quantity–Static Relation Problems from the Key Stage 1 National Curriculum Assessments	319
Appendix 4E: Continuous Quantity–Dynamic Relation problems from the Key Stage 1 National Curriculum Assessments.....	320
Appendix 4F: Continuous Quantity–Static Relation problems from the Key Stage 1 National Curriculum Assessments.....	321
Appendix 5A: Assumption Diagnostics of Path Analysis.....	322
Appendix 5B: Path Coefficients (95% Percentile-Based CI) and Indirect Effects Across Problem Types	324
Appendix 5C: Path Coefficients (95% BCa CI) and Indirect Effects by Problem Type	326
Appendix 5D: Path Coefficients (95% Percentile-Based CI) and Indirect Effects by Problem Type	334

List of Tables

Table 1-1	<i>One-Step Additive Word Problems with Different Compositions</i>	10
Table 1-2	<i>Two Dimensions Underpinning the One-Step Additive Word Problems</i>	
	<i>Conceptual Framework</i>	11
Table 1-3	<i>Problem-Solving Strategies for One-Step Additive Word Problems</i>	13
Table 1-4	<i>Examples of Typical Measures of Spatial Visualization Skills</i>	14
Table 1-5	<i>Types of Visuospatial Working Memory Measures</i>	15
Table 2-1	<i>Characteristics of Participating Schools</i>	26
Table 2-2	<i>Summary of Measures Used in the Thesis</i>	28
Table 2-3	<i>Randomization Process of Number Facts into Groups</i>	44
Table 2-4	<i>Randomization Process of Number-Fact Groups into Alternative Test Forms</i>	45
Table 2-5	<i>Discrete Quantity–Dynamic Relation Problems from Four Test Forms</i>	47
Table 2-6	<i>Discrete Quantity–Static Relation Problems from Four Test Forms</i>	48
Table 2-7	<i>Continuous Quantity–Dynamic Relation Problems from Four Test Forms</i>	49
Table 2-8	<i>Continuous Quantity–Static Relation Problems from Four Test Forms</i>	50
Table 2-9	<i>Representational Tools</i>	55
Table 2-10	<i>Arithmetic Approaches</i>	56
Table 2-11	<i>Problem-Solving Strategies</i>	57
Table 2-12	<i>Group Means for Demographic Characteristics and Performance on</i>	
	<i>Measures</i>	60
Table 2-13	<i>Model Fit Indices of Confirmatory Factor Analyses</i>	64
Table 2-14	<i>Factor Loadings of Second-Order Confirmatory Factor Analyses</i>	66
Table 2-15	<i>Summary of Study 1 Analyses</i>	68
Table 2-16	<i>Summary of Study 2 Analyses</i>	70
Table 2-17	<i>Summary of Study 3 Analyses</i>	72
Table 3-1	<i>Two Dimensions of One-Step Additive Word Problems</i>	88
Table 3-2	<i>Study Measures</i>	95
Table 3-3	<i>Descriptive Statistics and Bivariate Correlations</i>	105
Table 3-4	<i>Mixed-Effects Logistic Regression Predicting Word-Problem-Solving Accuracy</i>	
	<i>(Main-Effects Model)</i>	106
Table 3-5	<i>Mixed-Effects Logistic Regression Predicting Word-Problem-Solving Accuracy</i>	
	<i>(Interaction-Effects Model)</i>	108
Table 4-1	<i>Examples of Visual Representations</i>	124

Table 4-2	<i>Counting-Based Arithmetic Strategies</i>	127
Table 4-3	<i>Number-Fact Arithmetic Strategies</i>	128
Table 4-4	<i>Representations of Quantitative Structure</i>	131
Table 4-5	<i>Problem-Solving Strategies for One-Step Additive Word Problems</i>	133
Table 4-6	<i>Study Measures</i>	139
Table 4-7	<i>Problem-Solving Strategies and Examples</i>	144
Table 4-8	<i>Distribution of Problem-Solving Strategies Across all Item-Level Responses</i>	149
Table 4-9	<i>Between-Pupil Variation in Problem-Solving Strategy Use</i>	150
Table 4-10	<i>Between-Item Variation in Problem-Solving Strategy Use</i>	150
Table 4-11	<i>Descriptive Statistics and Bivariate Correlations Among Variables</i>	152
Table 4-12	<i>RQ3: Cross-Classified Mixed-Effects Multinomial Model Odds Ratios and 95% Confidence Interval</i>	154
Table 4-13	<i>Distribution of Problem-Solving Strategies across Problem Types</i>	158
Table 4-14	<i>RQ4: Cross-Classified Mixed-Effects Multinomial Model with Interaction Terms Odds Ratios and 95% Confidence Interval</i>	160
Table 5-1	<i>Types of Visual Representations for Visual–Spatial Word Problems</i>	184
Table 5-2	<i>Types of Visual Representations for Verbal–Analytical Word Problems</i>	185
Table 5-3	<i>Problem-Solving Strategies for One-Step Additive Word Problems</i>	194
Table 5-4	<i>Necessary Conditions for Causal Interpretation</i>	197
Table 5-5	<i>Descriptive Statistics and Bivariate Correlations</i>	206
Table 5-6	<i>Path Coefficients Predicting ‘Quantitatively Correct Number Fact’ Strategy use and Problem-Solving Accuracy</i>	209
Table 5-7	<i>Indirect Effects of Spatial Skills on Problem-Solving Accuracy</i>	210
Table 5-8	<i>Descriptive Statistics for Problem-Solving Accuracy and Strategy by Word Problem Type</i>	211
Table 5-9	<i>Path Coefficients and Indirect Effects for Mediation Models by Problem Type</i> .	213
Table 6-1	<i>Summary of Main Empirical Findings from each Study</i>	229

List of Figures

Figure 1-1	<i>A 2×2 Conceptual Framework for One-Step Additive Word Problems</i>	11
Figure 1-2	<i>Visual Representation of the Thesis's Research Themes</i>	16
Figure 2-1	<i>Wechsler Abbreviated Scale of Intelligence–Second Edition (WASI–II) Block Design</i>	33
Figure 2-2	<i>Subtasks of the Visual–Spatial Reasoning Task</i>	34
Figure 2-3	<i>PathSpan: A Digitized Corsi Block-Tapping Test</i>	36
Figure 2-4	<i>Iterative Process for Developing One-Step Additive Word Problems</i>	39
Figure 2-5	<i>Elimination of Double (Tie) Number Facts</i>	42
Figure 2-6	<i>Elimination of Number Facts Involving operands 1 to 4</i>	43
Figure 2-7	<i>Elimination of Number Facts Involving 20</i>	43
Figure 2-8	<i>Final Set of 24 Number Facts (Sums 11–20)</i>	44
Figure 2-9	<i>Final Set of One-Step Additive Word Problems</i>	46
Figure 2-10	<i>Randomization of Problem Order within each Scenario</i>	52
Figure 2-11	<i>Randomization of the Scenario Order</i>	52
Figure 2-12	<i>The Layout of Four Word Problems on A4-size Paper</i>	53
Figure 2-13	<i>MathLink Cubes</i>	54
Figure 2-14	<i>Tetrachoric Correlation Coefficients</i>	62
Figure 3-1	<i>A 2 × 2 Conceptual Framework for One-Step Additive Word Problem</i>	92
Figure 3-2	<i>One-Step Additive Word Problems</i>	97
Figure 3-3	<i>24 Number Facts up with Sums Ranging from 11 to 19</i>	98
Figure 3-4	<i>Predicted Probability of Problem-Solving Accuracy as a Function of Spatial Visualization across Word Problem Types</i>	110
Figure 3-5	<i>Conceptual Model of Indirect Pathways from Spatial Ability to Word-Problem-Solving Accuracy via Problem-Solving Strategies</i>	112
Figure 4-1	<i>Example of a Quantitative Structure Amenable to Direct Modeling</i>	134
Figure 4-2	<i>Example of a Quantitative Structure involving the Inverse of a Quantitative Relation</i>	134
Figure 4-3	<i>Conceptual Framework for One-Step Additive Word Problems</i>	136
Figure 4-4	<i>Predicted Probabilities of Problem-Solving Strategy Use across Spatial Visualization</i>	156
Figure 4-5	<i>Predicted Probabilities of Problem-Solving Strategy Use across Visuospatial Working Memory</i>	157

Figure 4-6	<i>Predicted Probabilities of Problem-Solving Strategies Use across Spatial Visualization by Word Problem Type</i>	163
Figure 4-7	<i>Predicted Probabilities of Problem-Solving Strategies Use across Visuospatial Working Memory by Word Problem Type</i>	165
Figure 5-1	<i>Examples of Additive Quantitative Reasoning Problems Used in the ALSPAC Project</i>	186
Figure 5-2	<i>Additive Conceptual Diagrams of Problem Representations from Yip et al. (2020).....</i>	188
Figure 5-3	<i>Hypothesized Mediation Model.....</i>	195
Figure 5-4	<i>Word Problem Types Based on Quantitative Structure</i>	200
Figure 5-5	<i>Mediation Models for each Word Problem Type</i>	214

1 General Introduction

While research with older secondary school pupils suggests that spatial ability may be important for solving certain types of mathematical word problems, few studies have investigated the relationship between spatial ability and arithmetic word problem solving in younger pupils (6–7-year-olds). Furthermore, while a body of research with older secondary school pupils indicates that the relationship between spatial ability and word problem solving success may be mediated by problem-solving strategies, there is scarcity of research with younger primary pupils. This integrated thesis, comprising three studies, aims to contribute towards addressing these gaps in the literature by examining the relationship between young primary school pupils' spatial visualization, visuospatial working memory, and their strategies and accuracy in solving one-step additive word problems.

To contextualize this thesis, [Section 1.1](#) outlines the general research background, followed by [Section 1.2](#), which identifies the research gaps and presents the aim of the thesis. [Sections 1.3](#) and [1.4](#) introduce the theoretical foundations and key concepts, respectively, to help orient the reader to the fundamental constructs used throughout the thesis. [Section 1.5](#) describes the research themes and corresponding research questions. [Section 1.6](#) considers the educational relevance of the thesis in the context of mathematics education. Finally, [Section 1.7](#) outlines the structure of the thesis. Together, they provide the background and context necessary for understanding the three studies presented in Chapters 3–5, which are written in the format of article manuscripts intended for future publication.

1.1 Research Background

Early mathematics skills are predictive of later academic success. In a meta-analytic study of six international longitudinal datasets (i.e., the United States, Great Britain, and Canada), Duncan and colleagues (2007) found that early mathematics skills (at about age 5 or 6) are the best predictor of later primary school academic performance, more so than early reading, attention, and socio-emotional skills. The predictive power of early mathematics skills for later academic achievement persists throughout primary school (Aubrey et al., 2006) and extends to secondary school (Claessens & Engel, 2013), even after accounting for early cognitive skills and family and child characteristics (Watts et al., 2014). Thus, the importance of early mathematics skills is widely recognized, given their association with overall later academic achievement, which ultimately predicts career success and satisfaction (ACT, 2009; Murnane et al., 2001).

One of the fundamental areas of early mathematics is solving arithmetic word problems (National Council of Teachers of Mathematics, 2000; OECD, 2023). As such, arithmetic word problems are emphasized within key education initiatives (e.g., National Centre for Excellence in the Teaching of Mathematics, 2014). Since arithmetic word problems serve as a basic building block upon which advanced mathematics skills are developed, the national curriculum in England dictates that “the principal focus of mathematics teaching in key stage 1 [5–7-year-olds] ... involve working with numerals, words and the four operations” (Department for Education, 2013, p. 5).

Unlike computation tasks, where arithmetic problems are presented as number sentences using numbers and symbols, arithmetic word problems are presented in written text (Wong & Ho, 2017). Also referred to as verbal or story problems, they have been described as “verbal descriptions of problem situations wherein one or more questions are raised that the answer to which can be obtained by the application of mathematical operations [addition, subtraction, multiplication, and division] to numerical data available in the problem statement” (Verschaffel et al., 2000, p. ix).

Even before formal mathematics instructions, children can solve arithmetic word problems using informal methods, such as pretend-play and counting (Nunes & Bryant, 1996; Verschaffel et al., 2020). Still, as solving word problems is a cognitively demanding process, involving the integration of multiple simultaneous processes (Vessonen et al., 2024), children often find that word problems are more challenging than computation tasks (Jitendra et al., 2007). Therefore, word problems are generally considered difficult for all pupils, regardless of their grade level (Vlahović-Štetić et al., 2004).

1.2 Research Gaps and Aim

Spatial ability is one understudied, yet promising lens for understanding young children’s word-problem-solving capabilities. Spatial ability is an umbrella term for a constellation of related skills that concern “the locations of objects, their shapes, their relations to each other, and the paths they take as they move” (Newcombe, 2017, p. 6). Extensive literature indicates that spatial ability plays a fundamental role in the learning and doing of mathematics (Hawes & Ansari, 2020; Lourenco et al., 2018; Mix & Cheng, 2012). While the spatial–mathematical link is considered “one of the most robust and well-established findings in cognitive psychology” (Mix & Cheng, 2012, p. 198), there remains a paucity of research

delineating the relationship between spatial ability and word-problem-solving performance, particularly in young children (Newcombe et al., 2019).

In addition to meta-analytic studies reporting a moderate positive association between spatial skills and mathematical skills (e.g., $r = .36$, Atit et al., 2022; $r = .27$, Xie et al., 2019), a recent meta-analysis by Hawes and colleagues (2022) provides evidence for a causal relationship. This analysis of 29 controlled pre–post studies ($N = 3,765$; ages 3–20 years; intervention durations ranging from 5 minutes to 44 hours) indicates that spatial training improves mathematics outcomes compared to a control condition, with an average effect size of Hedges’s $g = .28$ ($SE = .07$) (Hawes et al., 2022). This effect size can be contextualized in two ways. First, it can be compared to normative pupil growth trajectories in mathematics. From this perspective, the effect size is considerable, comparable to the annual gains that occur in Grades 6–10 (0.25–0.30 SDs), and to 25%–50% of annual gains for younger pupils in Kindergarten–Grade 5 (0.41–1.14 SDs) (Hill et al., 2008). Second, it can be compared to the effects of other cognitive training interventions. For example, working memory training shows effects on mathematics between .06 and .12 SDs (Melby-Lervåg et al., 2016; Schwaighofer et al., 2015); this further suggests that the effect of spatial training is comparatively substantial.

Many studies with older school pupils suggest that spatial ability is linked to solving word problems that are inherently spatial (e.g., Boonen et al., 2013, 2014; Casey et al., 2015; Duffy et al., 2020; Hegarty & Kozhevnikov, 1999; Kozhevnikov & Blazhenkova, 2013; Marschark et al., 2013; Oostermeijer et al., 2014; Van Garderen, 2006; Wang et al., 2022) as well as those that are not inherently spatial (e.g., Casey et al., 2015; Kyttälä & Lehto, 2008; Likhanov et al., 2024; Meyer et al., 2010; Mix et al., 2016, 2017; Wang et al., 2022). However, despite the importance and difficulty of arithmetic word problem solving, and evidence of a possible causal spatial–mathematical link, there have been very few investigations focused on the relationship between spatial ability and arithmetic word problem solving in the early primary education context. The limited research that does exist has yielded mixed findings, with some studies demonstrating a significant association (e.g., Chan & Wong, 2019; Fung & Swanson, 2017; Gilligan et al., 2017; Kyttälä et al., 2014; Mix et al., 2017, 2021; Zhang & Lin, 2015, 2017; Zheng et al., 2011), whereas others report null findings (e.g., Campos et al., 2013; Carr et al., 2008; Ching & Nunes, 2017; Fuchs et al., 2010; Rodán et al., 2019).

One of the suggested ways in which spatial ability facilitates arithmetic word problem solving in older secondary school pupils is through problem-solving strategies. In particular,

literature has focused on one form of problem-solving strategy: the construction of visual representations (Boonen et al., 2013, 2014; de Koning et al., 2002; Hegarty et al., 1995; Ng & Lee, 2009; Van Garderen, 2006). Successful problem solvers tend to produce visual representations that focus on the solution-relevant elements by reducing the situation to its “abstract mathematical gist” (Wong, 2021, p. 20). The production of such useful, accurate schematic visual representations was found to account for 21% of the relationship between spatial ability and word-problem-solving success (Boonen et al., 2013). In contrast, less successful problem solvers tend to produce iconic representations that focus on the vivid and detailed visual appearance of the objects or characters described in the problem, rather than on the solution-relevant elements (Hegarty & Kozhevnikov, 1999). To the best of the researcher’s knowledge, similar research focusing on the link between spatial ability, problem-solving strategies, and word-problem-solving success has not been conducted in the early primary education context (Newcombe et al., 2019).

To summarize, early mathematics skills, especially arithmetic word-problem-solving skills, are important, not least because they predict later academic and career success. Although young children can solve arithmetic word problems even before formal mathematics instruction, these problems are generally considered difficult for pupils of all grade levels. This is because, unlike computation tasks, arithmetic word problems are presented as verbal descriptions and therefore require the involvement of multiple cognitive processes.

Spatial ability is one understudied yet promising lens for understanding young children’s arithmetic word problem solving, as there is evidence of a causal spatial–mathematical link. However, few studies have systematically examined the relationship between spatial ability and arithmetic word problem solving in young primary school pupils. Furthermore, research with older secondary school pupils suggests that problem-solving strategies may be an important cognitive mechanism underlying the relationship between spatial ability and successful word problem solving; however, similar research has not been conducted with younger primary pupils.

This integrated thesis aims to contribute towards addressing these research gaps and to the body of knowledge in both theoretical and empirical ways. First, it proposes a conceptual framework for categorizing arithmetic word problems (more specifically, one-step additive word problems), as well as a complementary framework for categorizing problem-solving strategies used to solve such problems. Second, applying these frameworks, the thesis

empirically examines how spatial skills relate to young primary school pupils' (Key Stage 1, Year 2; 6–7-year-olds) problem-solving strategies and accuracy. By incorporating theoretically grounded variation in word problem types into the empirical investigations, the thesis offers a more fine-grained account of the spatial–mathematical link in the context of early primary mathematics education.

1.3 Theoretical Foundations

One of the biggest challenges in investigating the relationship between spatial ability and arithmetic word problem solving in early primary school pupils is that approaches to categorizing word problems based on *quantitative structure* have rarely been applied in research on the spatial–mathematical link. Therefore, this thesis draws on established theoretical distinctions between quantitative reasoning and arithmetic (Subsection 1.3.1), quantity and quantitative relation (Subsection 1.3.2), discrete and continuous quantities (Subsection 1.3.3), and dynamic and static relations (Subsection 1.3.4), which together provide a principled basis for examining the spatial–mathematical link within the context of arithmetic word problem solving.

1.3.1 The Distinction between *Quantitative Reasoning* and *Arithmetic*

Solving arithmetic word problems requires both quantitative reasoning and arithmetic skills (Nunes et al., 2012). Understanding the distinction between these two components requires distinguishing between quantity and number, which are sometimes erroneously treated as interchangeable. A *quantity* describes “properties of things that exist in the physical world” (Sophian, 2008, p. 3). In contrast, *numbers* are “symbols ... that are a product of our mental representations” (Sophian, 2008, p. 3), used to express “quantitative relationships within and between [quantities]” (Davydov, 1982, p. 227). By this definition, quantities are physical entities that can be measured, whereas numbers are symbols used to represent their measured values. Therefore, as Thompson (1993) points out, “quantities are more concrete than numbers” (p. 166).

Thinking of quantities does not require numerical representation (Nunes et al., 2009a). Davydov (1975) contends that what is most fundamental for mathematical development is not counting *per se*, but rather understanding the basic relationships between quantities—equality, less than, and greater than—which give meaning to the notion of quantity and serve as fundamental building blocks for quantitative reasoning. For example, given two collections of marbles, one can quantitatively reason about which has more without counting. This can be

done through one-to-one matching until one collection is exhausted and the other has unpaired marbles remaining. Likewise, given two sticks, one can reason quantitatively about which one is longer without knowing their exact lengths. Studies show that even 5- and 6-year-olds can engage in such emergent quantitative reasoning, using unenumerated sticks to determine which hole is deeper (Bryant & Kopytynska, 1976).

It is important to note that numbers, while they can function as representational tools of quantities, can also have meaning in and of themselves. For example, the logic of the whole-number system is that each number word (e.g., 9) in the series represents one more unit than the previous one (e.g., 8) such that it encompasses previous numbers additively (e.g., $8 + 1$, $7 + 2$, $6 + 3$, etc.) (Nunes & Bryant, 2015). Distinguishing these two meanings of numbers, representational and analytical, is foundational to understanding the difference between quantitative reasoning and arithmetic (Nunes et al., 2016).

Quantitative reasoning is the ability to analyze a problem situation in terms of its quantitative structure (Thompson, 1993). In quantitative reasoning, numbers have a representational meaning, such that “numbers and numeric relationships are of secondary importance” (Thompson, 1993, p. 165). In contrast, *arithmetic* involve understanding relations between numbers and carrying out operations (i.e., addition, subtraction, multiplication, and division) on them (Guedj, 1997). In arithmetic, numbers have an analytical meaning. In short, quantitative reasoning is based on relations between quantities, whereas arithmetic is based on relations between numbers.

Successful arithmetic word problem solving therefore requires both quantitative reasoning and arithmetic skills to analyze the problem’s underlying quantitative structure and operate on the numbers that represent that structure (Nunes et al., 2015). A problem solver who cannot determine the quantitative structure in a word problem will not be able to solve it correctly, even if they know how to execute the corresponding computation (Nunes et al., 2009a). Likewise, a problem solver who cannot operate “on the numbers to reach conclusions about the quantities” (Nunes et al., 2016, p. 5) will not be able to solve the problem correctly, even if they understand the underlying quantitative structure.

1.3.2 The Distinction between Quantity and Quantitative Relation

As introduced above, quantitative reasoning is the ability to analyze a problem situation in terms of its quantitative structure, that is, as “a network of quantities and quantitative relations” (Thompson, 1993, p. 166). Thus, the ontological distinction between quantities and

quantitative relations underpins quantitative reasoning by enabling representations of the quantitative structure of arithmetic word problems. A *quantity* is an attribute of an object that can be measured, such as the set size or the length (Thompson, 1993). In this regard, a quantity is an intrinsic property as it does not depend on anything other than the object itself (Dos Santos, 2022). A *quantitative relation*, in contrast, describes the connection between quantities (Nunes et al., 2009a). There are two types of quantitative relations: additive (part-whole relation) and multiplicative (one-to-many correspondence or proportions/ratios) (Nunes & Bryant, 2015).

When measured, both quantity and quantitative relation can be represented with numbers. The distinction is that quantitative relation is best represented by “directed numbers” (Vergnaud, 1982, p. 46). Consider the premise, ‘Ann has 8 books. Bob has 3 books. Ann has 5 more books than Bob.’ The first two phrases describe quantities; the quantities of Ann and Bob’s books have been measured and are represented by numbers, 8 and 3, respectively. The last phrase describes the quantitative relation between the two quantities and is best represented by the directed number, +5. Moreover, being a quantitative relation means having an inverse (i.e., Bob has 5 less books than Ann), which can again be represented by the directed number, −5 (Dos Santos, 2022).

1.3.3 The Distinction between Discrete Quantity and Continuous Quantity

Thompson’s (1993) notion of quantity can be subdivided by integrating methods of *measurement*, that is, the process of representing a quantity with a numerical value through the reiteration of a chosen criterion (i.e., a unit; Davydov, 1982; Sophian, 2008). In measurement, the unit defines what counts as ‘one,’ and this can either be *predetermined* or *imposed* (Davydov, 1982)—a distinction used in this study to differentiate between discrete and continuous quantities. Discrete quantities are defined as those for which the measurement unit is predetermined as the whole discrete element itself (Mix et al., 2002). For example, when measuring the numerosity of marbles, what is considered ‘one’ is the whole marble. Therefore, measuring numerosity requires only the *counting of the predetermined unit*, the discrete elements (i.e., each marble). Conversely, measuring continuous quantities requires the *imposition and reiteration of a unit*. For example, when measuring the weight of marbles, one must first determine the unit of measurement (e.g., grams, ounces, or stones), then reiterate the imposed unit to arrive at the numerical value representing the weight.

Put together, the unit of *discrete quantities* are specific and fixed, and measuring discrete quantities involves only counting the discrete units through a one-to-one mapping

between a set of objects and an ordered set of tags (e.g., set size). In contrast, *continuous quantities* describe entities measured on continuous dimensions (e.g., mass, length, distance, area, density, time, etc.; Newcombe et al., 2015), where measurement typically relies on imposed units and the partitioning of the quantity into equal intervals prior to enumeration (e.g., position). For example, 'Ann has 8 books' is a statement presenting a discrete quantity, as the number of books Ann has is a specific, fixed quantity. On the other hand, the statement 'Ann is 8 steps from the wall' presents a continuous quantity because the underlying dimension of Ann's position is continuous and must be measured using an imposed unit (i.e., steps). Both discrete and continuous quantities, when measured, can be integers (e.g., 6 candies and 8 steps) or rational values (e.g., 1/2 a box of candies and 3.5 meters).

1.3.4 The Distinction between Dynamic Relation and Static Relation

Thompson's (1993) notion of quantitative relation can also be subdivided. Quantitative relations are best represented by directed numbers. In Vergnaud's (1982) conceptualization of additive word problems, he contends that two main concepts are not adequately represented by natural numbers: dynamic transformations in state and static relationships between states. These concepts are similar to what Carpenter and Moser (1982) describe as active and static relationships, or what Nesher (1982) refers to as dynamic and comparison texts, or what Riley et al. (1983) describe as change and compare situations. The difference between the two is best explained using word problems that exemplify them. The following is an example of a problem involving a dynamic transformation of state (or, as Vergnaud (1982) describes it, a 'state-transformation-state' problem): 'Ann had 9 candies. She ate 2 candies. How many candies does she have now?' There are two components involved in such problem. One, there is a temporal sequence of events ($T_1 < T_2 < T_3$) (Nesher, 1982). Two, some change in the state of the nature is introduced at T_2 (Carpenter & Moser, 1982), such that the quantity referred to in T_1 is changed (Nesher, 1982). In contrast, the following is an example of a problem involving static relationship linking states (or, as Vergnaud (1982) describes it, a 'state-relationship-state' problem): 'Ann has 9 candies. She has 2 more than Bob. How many candies does Bob have?' Here, no action is implied (Carpenter & Moser, 1982). Instead, two disjoint quantities are compared on the same attribute (Nesher, 1982).

Adapting Vergnaud's (1982) conceptualization, quantitative relation can be subdivided into two categories: dynamic relation and static relation. A *dynamic relation* describes the transformation in the state of nature over time. For example, 'Ann played a game of marbles

and won 10 more marbles,’ describes a dynamic relation, or the change in the quantity of Ann’s marbles (discrete quantity) over time. Similarly, ‘Ann walked 12 steps away from the wall,’ describes a dynamic relation, or the change in Ann’s position (continuous quantity) over time. In essence, a dynamic relation describes the transformation linking the initial quantity and the final quantity.

A *static relation*, in contrast, describes the comparison of disjoint quantities. Simply put, it represents the difference between the referent quantity and the compared quantity. For example, the statement ‘Bob has 11 more marbles than Ann,’ illustrates a static relation, where the number of marbles Bob possesses (i.e., compared quantity) is compared to the number of marbles Ann possesses (i.e., referent quantity). Similarly, the statement ‘Bob is 8 steps farther from the wall than Ann,’ also describes a static relation, with Bob’s position (i.e., compared quantity) being compared to Ann’s position (i.e., referent quantity).

1.4 Key Concepts

Building on these theoretical foundations, this thesis focuses on three key concepts for addressing the research gaps and contributing to the literature: one-step additive word problems ([Subsection 1.4.1](#)), problem-solving strategies ([Subsection 1.4.2](#)), and spatial ability ([Subsection 1.4.3](#)). The first two concepts are grounded in the theoretical distinctions outlined above. Spatial ability is a complementary construct central to examining the spatial–mathematical link. These concepts are introduced here, while more detailed discussions are presented throughout the thesis in the sections where they are most relevant, reflecting the integrated nature of this work.

1.4.1 One-Step Additive Word Problems

Among various mathematical word problems—problems presented in verbal descriptions instead of number sentences (Boonen et al., 2016; Verschaffel et al., 2020; Wong & Ho, 2017)—this thesis focuses on one-step additive word problems. They are *additive* word problems, rather than problems involving multiplicative relations, or both (Nunes & Bryant, 2015). They are additive problems that require only *one step*—either the addition or subtraction of the two values provided in the problem (Verschaffel et al., 2020)—that is, problems involving either finding the cardinal of the whole, given the cardinal of the two parts, or finding the cardinal of a part, given the cardinal of the other part and the whole (Vergnaud, 1997). They are *problems* not necessarily because they are tasks for which the solution is not readily accessible to the problem solver due to the lack of an obvious method or algorithm (Mayer &

Wittrock, 1996), but because solving them involves multiple cognitive processes, including reasoning, comprehension, and decision-making (Vessonen et al., 2024).

More specifically, this thesis focuses on one-step additive word problems comprising *two quantities* and *one quantitative relation*, as they are at an appropriate level of difficulty for young primary pupils in the target age group. Word problems comprising only quantities may be too easy for 6–7-year-olds, such that ceiling effects are likely to occur (Carpenter & Moser, 1982; Riley et al., 1983). In contrast, word problems comprising only quantitative relations may be too difficult for the target participants (Vergnaud, 1982). Table 1-1 summarizes one-step additive word problems with different compositions of quantities and quantitative relations.

Table 1-1

One-Step Additive Word Problems with Different Compositions

Quantitative Structure	Example
Quantities Only	Ann had 5 marbles in her left pocket, and 7 marbles in her right pocket. How many marbles does she have altogether?
2 Quantities & 1 Quantitative Relation	Ann had 5 marbles. She played a game of marbles and won 7 marbles. How many marbles does she have now?
Quantitative Relations Only	Ann played two games of marbles. She won 5 marbles in the first game and won 7 marbles in the second game. What is the change in the number of marbles?

Note. This thesis focuses on one-step additive word problems comprising two quantities and one quantitative relation.

One of the challenges in investigating the association between spatial ability and arithmetic word problem solving is that distinctions in quantitative structure (Thompson, 1993) have rarely been brought into dialogue with research on the spatial–mathematical link. To address this challenge, drawing on the theoretical foundations described above, this thesis proposes a conceptual framework for categorizing one-step additive word problems, structured along two dimensions: (1) quantity (Davydov, 1982) and (2) quantitative relation (Vergnaud, 1982). Each dimension is conceptualized as binary. The quantity dimension comprises *discrete* and *continuous quantities*, while the quantitative relation dimension distinguishes *dynamic* and *static relations*. Table 1-2 summarizes the categories and definitions associated with each dimension.

Table 1-2

Two Dimensions Underpinning the One-Step Additive Word Problems Conceptual Framework

Dimension	Category	Definition
Quantity	Discrete	Quantities that are specific and fixed, such that measuring them involves counting discrete units through a one-to-one correspondence between a set of objects and an ordered set of tags (i.e., set size) (Davydov, 1982)
	Continuous	Quantities that are measured along continuous dimensions (e.g., mass, length, distance, area, density, time), using an imposed unit, such that measurement involves dividing the quantity into equal units before enumeration (i.e., position) (Davydov, 1982; Newcombe et al., 2015)
Quantitative Relation	Dynamic	Quantitative relations that describe a transformation in a state over time (i.e., winning/losing or moving toward/away from) (Vergnaud, 1982)
	Static	Quantitative relations that describe a comparison between disjoint quantities or positions (i.e., more/less or farther/closer) (Vergnaud, 1982)

Together, these two binary dimensions yield a 2×2 conceptual framework (Quantity $\times$ Quantitative Relation) comprising four distinct problem types: *Discrete Quantity–Dynamic Relation*, *Discrete Quantity–Static Relation*, *Continuous Quantity–Dynamic Relation*, and *Continuous Quantity–Static Relation*. The definitions and examples of each word problem type are presented in Figure 1-1. The methodological development, implementation, and validation of the conceptual framework are detailed in the Methodology chapter ([Subsection 2.3.3](#)).

Figure 1-1

A 2×2 Conceptual Framework for One-Step Additive Word Problems

		Quantitative Relation	
		Dynamic Relation	Static Relation
Quantity	Discrete Quantity	Problems involving transformations of quantities measured through one-to-one counting e.g., Ann had 6 marbles. Then she played a game of marbles and won 8 more marbles. How many marbles does she have now?	Problems involving comparisons of disjoint quantities measured through one-to-one counting e.g., Ann has 5 marbles. Bob has 7 more marbles than Ann. How many marbles does Bob have?
	Continuous Quantity	Problems involving transformations of quantities measured using an imposed unit e.g., Ann was 4 steps from the wall. Then she walked 7 steps away from the wall. How far is Ann from the wall now?	Problems involving disjoint comparisons of quantities measured using an imposed unit e.g., Ann is 9 steps from the wall. Bob is 3 steps farther from the wall than Ann. How far is Bob from the wall?

1.4.2 Problem-Solving Strategies

Consistent with prior research characterizing strategies as solution-oriented heuristics and procedures (Lemaire & Siegler, 1995; Schoenfeld, 1985; Tzur et al., 2021), *problem-solving strategies* in this thesis are defined as goal-directed procedures pupils employ to produce a numerical answer to one-step additive word problems, encompassing both formal mathematical heuristics and less structured response approaches. As discussed above, solving arithmetic word problems involve both quantitative reasoning and arithmetic skills (Nunes et al., 2012, 2015). Young pupils may not necessarily solve word problems in the sequential phases of ‘representation’ and ‘computation,’ as often assumed for older pupils and adults (Mayer, 1985), as they may instead rely on informal methods such as direct modelling (e.g., pretend play) and counting (Carpenter & Moser, 1984; Nunes & Bryant, 1996).

Given that young pupils may not proceed through word problems in discrete stages of representation followed by calculation, this thesis proposes a holistic framework for characterizing their problem-solving strategies, along two dimensions: (1) arithmetic strategy (i.e., counting-based versus number-fact strategies; Carpenter & Moser, 1982) and (2) representation of quantitative structure (i.e., whether the problem’s quantitative structure is represented correctly, incorrectly, or not at all; Boonen et al., 2014). In the present thesis, *counting-based* strategies refer to procedures in which children determine answers through sequential numerical progression, whereas *number-fact* strategies refer to solving problems by retrieving arithmetic facts from memory or deriving solutions from related known facts. The distinction between *quantitatively-correct* and *quantitatively-incorrect* strategies refers to the accuracy of the representation of the problem’s quantitative structure, irrespective of whether the final numerical answer is correct. Alternatively, the child’s strategy may be ‘*Non-Quantitative*’ in that it does not demonstrate engagement with the problem’s quantitative structure.

As summarized in Table 1-3, combining the dimensions of arithmetic strategy and representation of quantitative structure yields five distinct problem-solving strategy types: ‘*Non-Quantitative*,’ ‘*Quantitatively Incorrect Counting*,’ ‘*Quantitatively Incorrect Number Fact*,’ ‘*Quantitatively Correct Counting*,’ and ‘*Quantitatively Correct Number Fact*.’ The ‘*Non-Quantitative*’ strategy includes responses in which engagement with the problem’s quantitative structure is not evident in the child’s solution approach. Such responses may

involve guessing, repeating a number from the problem statement, or applying an answer-generation approach that does not reflect an attempt to represent the quantitative structure.

Table 1-3

Problem-Solving Strategies for One-Step Additive Word Problems

Strategy	Definition
Non-Quantitative	Does not appear to actively engage with the problem's quantitative structure (e.g., guessing; repeating a number from the problem statement)
Quantitatively Incorrect Counting	Represents the problem's quantitative structure incorrectly while using a counting-based strategy (e.g., counting-all for a subtraction problem)
Quantitatively Incorrect Number Fact	Represents the problem's quantitative structure incorrectly while using a number-fact strategy (e.g., mentally subtracting for an addition problem)
Quantitatively Correct Counting	Represents the problem's quantitative structure correctly while using a counting-based strategy (e.g., counting-down for a subtraction problem)
Quantitatively Correct Number Fact	Represents the problem's quantitative structure correctly while using a number-fact strategy (e.g., mentally adding for an addition problem)

In contrast, quantitatively-incorrect strategies reflect attempts to engage with the problem's quantitative structure, whether implemented through counting-based or number-fact approaches, but the representation does not align with the problem's underlying quantitative structure. Quantitatively-correct strategies involve engagement with the quantitative structure that accurately represents both the relevant quantities and their relations, whether implemented through counting-based or number-fact approaches.

The strategy coding procedure is described in the Methodology chapter ([Subsection 2.3.3.4](#)). A fuller theoretical justification of this framework is presented in Study 2 ([Subsection 4.2.2](#)), and the rationale for examining 'Quantitatively Correct Number Fact' strategies as a potential pathway linking spatial skills and problem-solving accuracy is elaborated in Study 3 ([Subsection 5.2.1](#)).

1.4.3 Spatial Ability

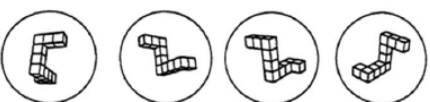
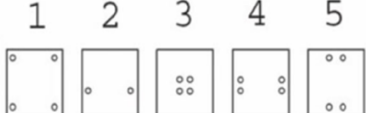
In this thesis, *spatial ability* is an umbrella term for a constellation of related skills that concern “the locations of objects, their shapes, their relations to each other, and the paths they take as they move” (Newcombe, 2017, p. 6). While decades of research have attempted to categorize spatial ability into subfactors, there is no conclusive categorization scheme. In fact, more recent studies indicate that spatial ability may be a unitary construct (Mix et al., 2016,

2017). Given the lack of cohesion and the multiplicity of spatial taxonomies, Schenck and Nathan (2024) suggest that researchers should select spatial skills based on empirical evidence when conducting research on the spatial–mathematical link, to ensure their suitability for the population and mathematical skills under investigation. Following their suggestion, this thesis concentrates on two spatial skills that have been most consistently identified as having a particularly strong association with mathematical thinking: spatial visualization (Hawes & Ansari, 2020) and visuospatial working memory (Allen et al., 2019).

Spatial visualization is “the ability to generate, recall, maintain, and manipulate visual-spatial images and solutions” (Hawes & Ansari, 2020, p. 466). As shown in Table 1-4, typical measures of spatial visualization include 2-dimensional and 3-dimensional mental rotation task, mental composition task, and mental paper-folding task (Hawes & Ansari, 2020). An extensive body of literature suggests that the relationship between spatial visualization and mathematical performance is correlational (e.g., Ganley & Vasilyeva, 2011; Rhode & Thompson, 2007; Sorby & Panther, 2020), possibly causal (e.g., Cheng & Mix, 2014; Cheung et al., 2020; Newcombe et al., 2019), and even bi-directional (e.g., Geer et al., 2019).

Table 1-4

Examples of Typical Measures of Spatial Visualization Skills

Task Description	Example Item
2D Mental Rotation^a ‘Which shape will you get when you combine these two pieces (<i>left</i>) together?’	 
3D Mental Rotation^b ‘Which two figures are identical to the target figure on the left (<i>just seen from different angles</i>)?’	 
Mental Composition^b ‘Which three shapes can be combined to compose the target shape on the far left?’	 
Mental Paper Folding^b ‘A piece of paper has been folded and hole-punched (<i>left</i>). Which image on the right corresponds to the instructions on the left?’	 

Note. ^aAdapted from “Relations between numerical, spatial, and executive function skills and mathematics achievement: A latent-variable approach,” by Z. Hawes, J. Moss, B. Caswell, J. Seo, and D. Ansari, 2019b. *Cognitive Psychology* 109, p. 74 (<https://doi.org/10.1016/j.cogpsych.2018.12.002>). Copyright 2019 by Science Direct. Reprinted with permission.

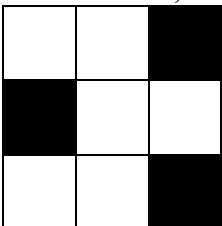
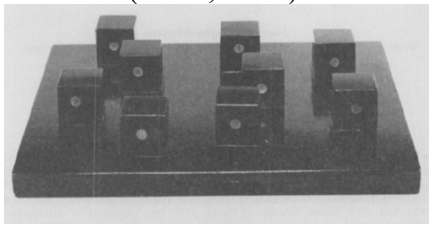
^bAdapted from “What explains the relationship between spatial and mathematical skills? A review of evidence from brain and behavior,” by Z. Hawes and D. Ansari, 2020, *Psychonomic Bulletin & Review* 27(3), p. 467 (<https://doi.org/10.3758/s13423-019-01694-7>). Copyright 2020 by Springer Nature. Reprinted with permission.

Visuospatial working memory is the ability to temporarily store and manipulate visuospatial information (Baddeley & Hitch, 1974). It is closely associated with executive function processes (Miyake et al., 2000; for a recent review, see Miller-Cotto & Gordon, 2025), and reflects both storage and processing demands within the working memory system (Shah & Miyake, 1996). However, it also overlaps substantially with spatial ability (Mix & Cheng, 2012). Reflecting this conceptual overlap, visuospatial working memory has been treated as either a component of executive function or spatial ability in the literature (Hawes et al., 2019b). In this thesis, visuospatial working memory is conceptualized as part of spatial ability and is therefore treated as a spatial skill.

As shown in Table 1-5, there are two types of visuospatial working memory: (1) simultaneous, and (2) sequential (for a review, see Allen et al., 2019). The distinction between the two types of visuospatial working memory is based on how the stimuli are presented. In simultaneous visuospatial working memory tasks, all information is presented at the same time. In contrast, in sequential visuospatial working memory tasks, the presentation of stimuli is in a sequence.

Table 1-5

Types of Visuospatial Working Memory Measures

Simultaneous	Sequential
Visual Patterns Task (Della Sala et al., 1999)	Corsi Block-Tapping Test (Corsi, 1972) ^a
	

Note. In the Visual Patterns Task (left), the participant is presented with a stimulus grid with some internal squares filled in and others left blank. Once the stimulus grid is taken away, the participant is given a blank grid on which they must replicate the pattern. In the Corsi Block-Tapping Test (right), the examiner taps a sequence of blocks, and the participant must either repeat in the same order (forward span) or in reverse order (backward span). ^a*Reprinted from* “Differential course of development of spatial and verbal memory span: A normative study,” by E. B. Isaacs, & F. Vargha-Khadem, 1989, *British Journal of Developmental Psychology*, 7(4), p. 378 (<https://doi.org/10.1111/j.2044-835X.1989.tb00814.x>). Copyright 1989 by The British Psychological Society. *Reprinted with permission.*

A recent systematic review on the relationship between visuospatial working memory and mathematical performance indicates that sequential visuospatial working memory may be particularly promising in examining the spatial–mathematical link (Allen et al., 2019).

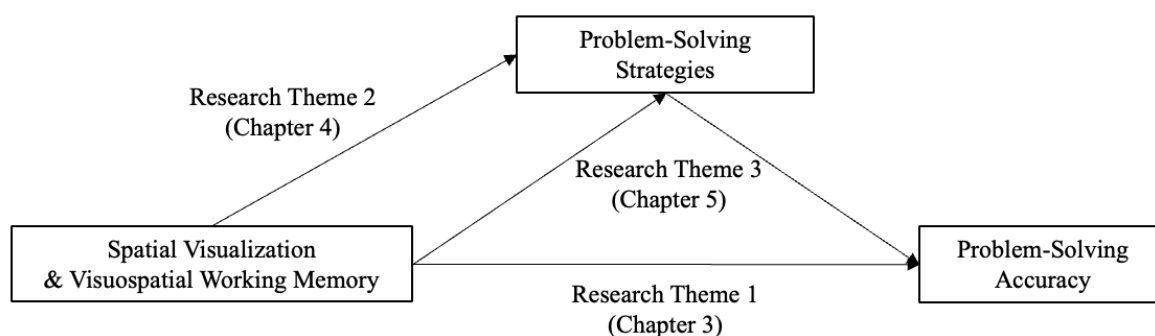
However, studies have only investigated the influence of sequential visuospatial working memory on numerical operation (i.e., computation tasks), as opposed to mathematical reasoning (i.e., word problems); this gap highlights “an evident lack in the literature of such focused work” (Allen et al., 2019, p. 526). Therefore, this thesis will focus on sequential visuospatial working memory and how it relates to one-step additive word problem solving.

1.5 Research Themes and Questions

To investigate the relationship between young pupils’ spatial visualization, visuospatial working memory, and their strategies and accuracy in solving one-step additive word problems, this integrated thesis is organized around three research themes. As shown in Figure 1-2, these themes correspond to Chapters 3–5, each of which comprises a manuscript intended for journal publication. Each of these addresses two research questions, outlined in [Subsections 1.5.1](#), [1.5.2](#), and [1.5.3](#).

Figure 1-2

Visual Representation of the Thesis’s Research Themes



1.5.1 Research Theme 1: Spatial Skills and One-Step Additive Word-Problem-Solving Accuracy

The first research theme, presented in [Chapter 3](#) (Study 1), examines the extent to which spatial skills predict one-step additive word-problem-solving accuracy. The corresponding manuscript is titled *The Role of Spatial Skills in Young Children’s Arithmetic Word-Problem-Solving Accuracy*. The research questions are:

RQ1) To what extent do spatial visualization and visuospatial working memory predict young primary school pupils’ (6–7-year-olds’) accuracy on one-step additive word problems?

RQ2) Do the relationships between spatial visualization, visuospatial working memory, and problem-solving accuracy vary across different word problem types?

1.5.2 Research Theme 2: Spatial Skills and Problem-Solving Strategies

The second research theme, presented in [Chapter 4](#) (Study 2), examines the extent to which spatial skills predict problem-solving strategy use. The corresponding manuscript is titled *The Role of Spatial Skills in Young Children's Arithmetic Word-Problem-Solving Strategies*. The research questions are:

- RQ3)** To what extent do spatial visualization and visuospatial working memory predict young primary school pupils' (6–7-year-olds') problem-solving strategy use?
- RQ4)** Do the relationships between spatial visualization, visuospatial working memory, and problem-solving strategy use vary across different word problem types?

1.5.3 Research Theme 3: Spatial Skills and One-Step Additive Word-Problem-Solving Accuracy: The Mediating Role of Problem-Solving Strategies

The third, and final, research theme, presented in [Chapter 5](#) (Study 3), examines whether problem-solving strategies—specifically, the use of the 'Quantitatively Correct Number Fact' strategy—mediate the relationship between spatial skills and one-step additive word-problem-solving accuracy. The corresponding manuscript is titled *Young Children's Problem-Solving Strategies as a Mediator between Spatial Skills and Arithmetic Word-Problem-Solving Accuracy*. The research questions are:

- RQ5)** To what extent do spatial visualization and visuospatial working memory exert indirect effects on young primary school pupils' (6–7-year-olds') accuracy on one-step additive word problems through 'Quantitatively Correct Number Fact' strategy use?
- RQ6)** Do the indirect effects of spatial visualization and visuospatial working memory on problem-solving accuracy through 'Quantitatively Correct Number Fact' strategy use vary across different word problem types?

1.6 Educational Relevance

Research on arithmetic word problem solving through the lens of spatial ability is sorely needed in the context of early primary education in England. While evidence suggests that spatial reasoning may be a valuable tool for mathematical problem solving (Gilligan-Lee et al., 2022), it is underrepresented in England's early years (ages 2 to 5) and primary mathematics curriculum (ages 5 to 11). Within the English Early Years Foundation Stage (EYFS) statutory framework (ages 2 to 5), spatial reasoning has been relatively de-emphasized following the

2021 reforms, which removed the ‘*shape, space and measures*’ Early Learning Goal (ELG) and replaced it with a second ‘*numbers*’ ELG (Department for Education, 2021). In the English national mathematics curriculum for Key stages 1 and 2 (ages 5 to 11), the word ‘spatial’ appears once. Notably, this occurs in the non-statutory notes to describe how Year 2 pupils should be encouraged to “represent [larger numbers] in different ways, including spatial representations” (Department for Education, 2013, p. 11), rather than in reference to spatial reasoning itself.

While England may not be unique in this respect, as “contemporary school mathematics curricula tend to include few opportunities to develop, apply, and integrate spatial reasoning” (Davis et al., 2015a, p. 47), there is evidence to suggest that pupils in England exhibit different patterns of mathematical performance compared with their international peers. In PISA 2022, whereas, on average, 15-year-olds across OECD countries performed lowest on the ‘*change and relationships*’ domain (one of the ‘*quantity*,’ ‘*uncertainty and data*,’ ‘*change and relationships*,’ and ‘*space and shape*’ domains), pupils in England performed lowest in the ‘*space and shape*’ domain (Ingram et al., 2023). Similarly, in the Trends in International Mathematics and Science Study (TIMSS), most of the highest-performing countries performed strongly in the ‘*measurement and geometry*’ domain (one of the ‘*number*,’ ‘*measurement and geometry*,’ and ‘*data*’ domains) that emphasizes spatial reasoning. In contrast, Year 5 pupils (aged 9–10) in England performed most weakly in this domain in 2023, as they also did in 2015 and 2019 (Golding et al., 2024). Such patterns are consistent with arguments that pupils’ mathematical performance reflects the emphases and structures of the curriculum they experience (e.g., Resnick et al., 1989). Although performance on large international assessments should not be the primary driver of the mathematics curriculum, such performance provides some indication of its impact.

There are two main approaches for harnessing the spatial–mathematical link in education: (1) directly training spatial ability to yield mathematical gains and (2) *spatializing* the mathematics curriculum (Newcombe, 2017). The former approach, directly training spatial ability to yield mathematical gains, has been the subject of ongoing debate in the literature. Broadly, a recent meta-analysis indicates that spatial training improves mathematics outcomes compared to a control condition, an improvement that is comparable to the annual gains made in regular educational contexts and greater than effects of other cognitive training interventions (Hawes et al., 2022). Part of the argument of the proponents for spatial training is that spatial

ability is malleable, transferable and durable (Newcombe, 2017; Uttal et al., 2013), which translates to the idea that spatial skills can be improved through training, gains can be transferred across different sets of spatial tasks, and the benefits of spatial training last over time.

More radical proponents of curriculum reform claim that “spatial reasoning should be developed for the sake of better spatial reasoning—that is, not just in the service of better number sense or in anticipation of its utility for teaching algebra” (Davis et al., 2015b, p. 145). There may be some justification for this position, given rigorous evidence of the importance of spatial reasoning for entry into and success in science, technology, engineering, and mathematics (STEM)-related fields (Newcombe, 2017; Shea et al., 2001; Uttal & Cohen, 2012; Wai et al., 2009). Specifically, spatial skills are especially crucial for novice pupils in STEM-related fields as they “serve as a gateway or barrier for entry into STEM fields” (Uttal & Cohen, 2012, p. 148). Thus, the failure to cultivate spatial reasoning may have a negative impact on equitable access to STEM-related fields. Numerous anecdotal accounts, such as Francis Crick and James Watson’s envisioning of the 3D structure of deoxyribonucleic acid (DNA) from Rosalind Franklin’s flat x-ray diffraction images, suggest that spatial reasoning is foundational to creative innovations in STEM-related fields (Newcombe, 2017; Uttal & Cohen, 2012). The link between spatial reasoning and STEM-related fields is particularly relevant in an age of technological change, in which “technology-related roles are the fastest-growing jobs in percentage terms” and “AI and big data top the list of fastest-growing skills, followed closely by networks and cybersecurity as well as technology literacy” (World Economic Forum, 2025, p. 5).

The latter approach, spatializing the mathematics curriculum, may seem a more efficient and pragmatic method of harnessing the spatial–mathematical link, rather than simply adding spatial reasoning as an additional topic. However, this approach requires clear understanding of the cognitive mechanisms underlying the spatial–mathematical link. Given that both mathematical and spatial abilities comprise multiple skills (Lourenco et al., 2018; Mix & Cheng, 2012), identifying such mechanisms is a challenging endeavor that requires careful investigation of specific types of spatial and mathematical skills.

As such, this thesis’s focus on specific spatial skills (i.e., spatial visualization and visuospatial working memory) and a specific mathematical skill (i.e., one-step additive word problem solving) may have educational relevance in terms of spatializing the mathematics

curriculum. Furthermore, the thesis situates these investigations within the English educational context, where spatial reasoning is underrepresented in the EYFS statutory framework and primary mathematics curriculum, thereby highlighting a potential opportunity to strengthen early mathematics education through the integration of spatial approaches.

1.7 Thesis Structure

This integrated thesis consists of six chapters: General Introduction, Methodology, Study 1, Study 2, Study 3, and General Discussion. The three manuscripts (corresponding to Chapters 3–5) are structured in accordance with the conventional format of journal articles, including introduction, literature review, methods, findings, and discussion sections. However, they currently exceed the typical length limits for journal submissions. This is because each manuscript includes a more comprehensive literature review than is customary, given that the thesis does not contain a separate, stand-alone literature review chapter. The findings and discussion sections are likewise more detailed in order to reflect the scope and expectations of a doctoral thesis. Some material—particularly theoretical and methodological content—is necessarily revisited across the manuscripts to ensure that each study is self-contained. Cross-references are included throughout the thesis, including within the manuscripts, to enhance overall cohesion.

For journal submission, the manuscripts will be revised and condensed to meet the specific requirements of the target journals. The references for the thesis are presented collectively to minimize repetition and improve coherence; however, for journal submission, the relevant references will be extracted from this master list. Likewise, for consistency across the thesis, the research questions within each manuscript are numbered to align with the overall thesis structure; for journal submission, they will be renumbered as appropriate. The manuscripts are written in American English, consistent with the conventions of many relevant publishers, and the thesis adopts American English spelling throughout for consistency.

The contents of each chapter are as follows:

Chapter 1: General Introduction

To contextualize this thesis, the introduction chapter outlines the importance of arithmetic word problem solving in the early primary years. It introduces extensive evidence of a spatial–mathematical link, and identifies spatial ability as an understudied yet promising lens for understanding young pupils’ arithmetic word problem solving. The chapter then

Chapter 1: General Introduction

presents the theoretical foundations of the thesis, namely the distinctions between quantitative reasoning and arithmetic, quantity and quantitative relation, discrete and continuous quantities, and dynamic and static relations. It also introduces and briefly describes the three key concepts central to the investigation: one-step additive word problems, problem-solving strategies, and spatial ability. The chapter considers the educational relevance of the thesis. Finally, it orients the reader to the overall structure of the thesis.

Chapter 2: Methodology

This chapter presents the methodology used to investigate the three research themes. Although each manuscript contains its own methods section, this chapter provides a more comprehensive account of the research design, recruitment process, participants, materials, procedures, and ethical considerations. It details the steps taken to ensure the reliability and validity of the one-step additive word problems, an original measure developed for this thesis. In addition, it provides an overview of the analytic approaches used in each study, while detailed descriptions are reported in the individual manuscripts. The methods sections of the manuscripts draw on relevant content from this chapter.

Chapter 3: The Role of Spatial Skills in Young Children's Arithmetic Word-Problem-Solving Accuracy (Study 1 Manuscript)

This manuscript aims to make both theoretical and empirical contributions to literature. Theoretically, the study proposes a conceptual framework for categorizing one-step additive word problems by integrating Thompson's (1993) distinction between quantity and quantitative relation, Davydov's (1982) discussion of different methods of quantification, and Vergnaud's (1982) classification of dynamic and static relations. In doing so, the study examines how variations in quantitative structure may relate to spatial reasoning, thereby extending research on the spatial–mathematical link to early quantitative reasoning.

Empirically, the study investigates whether spatial visualization and visuospatial working memory predict problem-solving accuracy on one-step additive word problems (RQ1), and whether these relationships vary across word problem types (RQ2). RQ1 is examined using a cross-classified mixed-effects logistic regression model. RQ2 is examined by incorporating interaction terms between spatial skills and word problem type to test for moderation.

Chapter 4: The Role of Spatial Skills in Young Children's Arithmetic Word-Problem-Solving Strategies (Study 2 Manuscript)

This manuscript aims to make both theoretical and empirical contributions to literature. Theoretically, the study proposes a holistic framework for categorizing the problem-solving strategies that young pupils use when solving one-step additive word problems by integrating two dimensions: arithmetic strategies and representations of quantitative structure (Nunes et al., 2012).

Empirically, the study investigates whether spatial visualization and visuospatial working memory predict problem-solving strategies (RQ3), and whether these relationships vary across word problem types (RQ4). RQ3 is examined using a cross-classified mixed-effects multinomial logistic regression model. RQ4 is examined by incorporating interaction terms between spatial skills and word problem type to test for moderation.

Chapter 5: Young Children's Problem-Solving Strategies as a Mediator between Spatial Skills and Arithmetic Word-Problem-Solving Accuracy (Study 3 Manuscript)

This manuscript aims to advance understanding of the mechanisms underlying the relationship between children's spatial skills and problem-solving accuracy. Using path analysis, the study investigates whether 'Quantitatively Correct Number Fact' strategy use mediates the relationship between spatial visualization, visuospatial working memory, and problem-solving accuracy (RQ5). This mediating strategy is defined as representing the problem's quantitative structure correctly while using a number-fact strategy. The study also examines whether the indirect effects of spatial skills on problem-solving accuracy, mediated through 'Quantitatively Correct Number Fact' strategies, vary across word problem types (RQ6). RQ5 is examined using a mediation model estimated through path analysis, and RQ6 is examined by estimating separate mediation models through path analysis for each word problem type.

Chapter 6: General Discussion

This final chapter summarizes the thesis's main theoretical and empirical contributions and synthesizes the key findings into overarching conclusions. In light of these findings, educational implications are discussed. Finally, limitations of the thesis and directions for future research are presented.

2 Methodology

This chapter presents the methodology used in investigating the three research themes. [Subsection 2.1](#) outlines the research design of this thesis. [Subsection 2.2](#) describes the recruitment process and participants. [Subsection 2.3](#) details the materials and testing procedures, and [Subsection 2.4](#) outlines the analytic approaches used in the three manuscripts. Finally, [Subsection 2.5](#) discusses ethical considerations, and [Subsection 2.6](#) summarizes the Methodology chapter.

2.1 Research Design

To explore the three research themes and their corresponding research questions, this thesis employs a correlational and cross-sectional design using quantitative methods. A *correlational* design is used to examine the relationships between young primary school pupils' spatial skills, problem-solving strategies, and accuracy in solving one-step additive word problems (Punch & Oancea, 2014)—an under-researched area in early primary education. The *cross-sectional* design enables examination of existing differences in children's cognitive abilities and in their problem-solving accuracy and strategies at a single point in time (Cohen et al., 2018). *Quantitative* methods are employed to describe the associations among the variables, offering insights into the *what*, *when*, and *where* of these relationships (Cohen et al., 2018; Field, 2018).

There are two different temporal perspectives for understanding the spatial–mathematical link: one concerns how the two domains relate in real time, and the other concerns how they relate across developmental time (Mix, 2019). This thesis focuses on the former, examining how spatial ability, problem-solving strategies, and accuracy in solving word problems relate in real time. Importantly, the correlational, cross-sectional design of the thesis precludes definitive causal claims about these relationships.

2.2 Participants and Recruitment

This subsection first describes the *a priori* power analysis conducted to determine the thesis's sample size ([Subsection 2.2.1](#)). It then outlines the two-step recruitment process based on non-probabilistic convenience sampling ([Subsection 2.2.2](#)). Next, the characteristics of the participating schools ([Subsection 2.2.3](#)) and pupils ([Subsection 2.2.4](#)) are presented.

2.2.1 *A Priori Power Analysis*

An *a priori* power analysis was conducted in G*Power (Version 3.1.9.6; Faul et al., 2009) using the ‘Linear multiple regression: Fixed model, single regression coefficient’ procedure. The analysis was based on a model containing seven predictors and was designed to determine the sample size required to detect a medium effect. The anticipated effect size was informed by a benchmark from the spatial literature available at the time (e.g., Uttal et al., 2013; $g \approx 0.47$), which was converted to the corresponding regression effect size ($f^2 \approx 0.055$) for use in G*Power. Assuming $\alpha = .05$ and power of .80, the resulting minimum sample size was 145 participants. This anticipated effect size is broadly consistent with meta-analytic studies reporting moderate positive associations between spatial and mathematical skills (e.g., $r = .27$, Xie et al., 2019; $r = .36$, Atit et al., 2022).

Subsequent meta-analytic work has reported a range of effect sizes depending on the outcome domain and age group. For example, Yang et al. (2020) reported relatively large effects of spatial training on spatial outcomes in 0–8-year-olds ($g \approx 0.96$), whereas Hawes et al. (2022) reported smaller average effects on mathematical outcomes ($g \approx 0.28$), with the effect size increasing as participants’ age increased from 3 to 20 years. Given that the present study examines the association between spatial ability and mathematical word problem solving, the latter benchmark may be more directly comparable, although the age range of the participants aligns more closely with the former. Furthermore, the final analytic approach incorporated mixed-effects and multinomial models, which differ from the simpler associations assumed in the initial power calculation. Accordingly, although the present sample was likely sufficient to detect moderate effects, power to detect smaller effects and interaction terms may be more limited. These considerations should be taken into account when interpreting non-significant findings reported in the manuscripts, as such findings may reflect insufficient statistical power rather than the absence of an effect.

2.2.2 *Recruitment Process*

A non-probabilistic convenience sampling method was used to recruit participants due to accessibility and resource constraints (Cohen et al., 2018; Punch & Oancea, 2014). A two-step recruitment process was used whereby schools were recruited first, followed by the pupils. Primary schools within a county in South East England were identified based on geographical convenience. The headteachers of the identified schools were invited to participate in the study via email ([Appendices 2A](#) and [2B](#)) in three waves over the course of three months (between

January 2023 and April 2023). Once consent from the headteachers ([Appendix 2C](#)) was obtained, the class teachers were contacted to obtain their consent ([Appendix 2D](#)) and to negotiate testing dates. A total of 67 schools were contacted, and six schools consented to participate in the study; however, due to a scheduling conflict, only five schools ultimately participated.

Schools were offered a choice of four parental consent approaches: extended opt-out ([Appendix 2E](#)), simplified opt-out ([Appendix 2F](#)), extended opt-in ([Appendix 2G](#)), or simplified opt-in ([Appendix 2H](#)). The simplified versions were one page in length that summarized the key points from the extended forms. Four of the five schools used the opt-out parental consent approach, and one school used the opt-in parental consent approach. Out of the four schools that used the opt-out parental consent approach, two of them chose the simplified version while the other two chose the extended version. The school that used the opt-in parental consent approach chose the simplified version.

2.2.3 School Characteristics

As shown in Table 2-1, the five participating schools varied in size and pupil background characteristics (Department for Education, 2024). The school population ranged from just over 100 pupils to just under 450 pupils. The percentage of pupils from a minority ethnic background (i.e., of any origin other than White British) ranged from 35.0% to 81.0%, compared to the national average of 36.1%. The percentage of pupils whose first language is known or believed to be other than English ranged from 14.6% to 43.4%, compared to the national average of 22.0%. The percentage of pupils eligible for free school meals ranged from 12.4% to 46.0%, compared to the national average of 23.8%.

Additionally, the five participating schools also varied in their Special Education Needs (SEN) provision. There are two types of SEN provision: SEN support and Education, Health and Care (EHC) plan. SEN support is given to pupils who have a learning difficulty or disability and need extra or different support than what is usually provided in the school curriculum (Department for Education, 2024). EHC plan is implemented when a pupil's needs extend beyond the scope of SEN support. The percentage of pupils receiving SEN support ranged from 17.7% to 25.8%, compared to the national average of 13.5%. The percentage of pupils receiving an EHC plan ranged from 1.9% to 3.7%, compared to the national average of 2.5%.

Table 2-1*Characteristics of Participating Schools*

School	Number of Pupils Enrolled in the School	Percentage of Pupils				
		From a Minority Ethnic Background	Whose First Language is not English	Eligible for Free School Meals	Receiving SEN Support	Receiving EHC Plan
A	201 to 300	67.9%	43.4%	23.4%	18.5%	2.3%
B	201 to 300	59.1%	24.8%	45.8%	20.7%	3.7%
C	101 to 200	35.0%	14.6%	12.4%	17.7%	2.1%
D	401 to 500	47.9%	26.7%	46.0%	18.5%	1.9%
E	301 to 400	81.0%	33.3%	30.0%	25.8%	2.9%
National Average		36.1%	22.0%	23.8%	13.5%	2.5%

Note. The statistics of the schools (2023-2024) are taken from <https://explore-education-statistics.service.gov.uk/find-statistics/school-pupils-and-their-characteristics>. SEN = Special Educational Needs. EHC = Education, Health and Care.

2.2.4 Participants

A total of 149 pupils participated in the study between February 2023 and June 2023. Of these, 12 pupils did not complete the study: 11 experienced difficulty understanding tasks, and one chose to discontinue participation. The final sample therefore comprised 137 pupils (50.4% female, 49.6% male; age range: 6.52–7.74 years; $M = 7.10$ years).

Of the final sample, 27 pupils (19.7%) were reported by schools as having SEN, and 21 pupils (15.3%) were identified as English as an Additional Language (EAL) learners. These children were retained in the analyses, as schools indicated that their level of SEN support within regular classrooms was minimal or that their English proficiency was sufficient to engage meaningfully with the tasks.

Although SEN and EAL information was not consistently available across all schools, the inclusion of pupils with diverse learning profiles reflects the natural variability present in mainstream classrooms. This enhances the ecological validity and generalizability of the findings, as the sample represents a broad range of learners rather than a narrowly selected subgroup. While more comprehensive documentation of SEN and EAL status would have enabled a more detailed characterization of the sample and the examination of potential differences across learner groups, the diversity of the sample strengthens the relevance of the findings to typical English primary school contexts.

2.3 Materials and Procedures

As summarized in Table 2-2, each child completed a total of seven tasks measuring various abilities, including spatial, language, numerical, fluid reasoning, and one-step additive word problem solving. All tasks were administered individually in a quiet space, either in the hallway or in an open area near the child's classroom, to ensure the presence of a familiar adult. The tasks were completed over two testing sessions, each lasting approximately 30 minutes and typically spaced at least one hour apart (but no less than 30 minutes), an arrangement informed by pilot studies (see [Subsection 2.3.3.1.1](#)).

[Subsection 2.3.1](#) outlines the overarching reliability and validity considerations that guided the selection and design of measures. [Subsection 2.3.2](#) then describes the established measures (i.e., spatial ability, language ability, numerical ability, and fluid reasoning measures) including their administration procedures and evidence of reliability and validity. [Subsection 2.3.3](#) provides a detailed account of the development, implementation, and validation of the one-step additive word problems.

Table 2-2*Summary of Measures Used in the Thesis*

Variable	Task	Description	Scoring
Spatial Visualization	Wechsler Abbreviated Scale of Intelligence–Second Edition (WASI–II) Block Design subtest	While viewing a constructed model or a picture in the Stimulus Book, the participant uses red-and-white cubes to recreate the design within a specified time limit of 30 to 120 seconds per design. There are 11 items for children aged 6 to 8 years; testing is discontinued after two consecutive scores of zero.	• Scores range from 0 to 2 for the first four items, and 0 to 7 for the final seven items. • The total raw score is the sum of all item scores. • The raw score ranges from 0 to 57 for children aged 6 to 8 years.
	Visual–Spatial Reasoning Task	Four subtasks (five items per subtask) require reasoning and visualizing solutions to problems involving parts of objects in relation to their whole: missing-center problems, missing-border problems, two-dimensional shape composition/decomposition problems, and paper-folding problems. There are 20 items in total, and all items are administered.	• Each item is scored as correct (1) or incorrect (0). • The total raw score is the sum of all item scores. • The raw score ranges from 0 to 20.
Visuospatial Working Memory	PathSpan	A digitized version of the Corsi Block-Tapping Task administered on an iPad, in which the participant replicates a sequence of dots that light up one at a time by tapping them in the same order. The sequence length begins at two, with two trials per sequence length. The sequence length increases by one if the participant correctly reproduces the pattern on at least one of the two trials at a given length. Testing is discontinued if the participant fails both trials at a given sequence length. The maximum sequence length is nine.	• The score is the highest sequence length accurately reproduced at least once. • The raw score ranges from 0 to 9.

Variable	Task	Description	Scoring
Language Ability	Wechsler Individual Achievement Test–Second UK Edition (WIAT–II ^{UK}) Listening Comprehension subtest	Three subtasks involve selecting the picture that matches a word (Receptive Vocabulary; 16 items) or sentence (Sentence Comprehension; 10 items) and generating a word that matches a picture and oral description (Expressive Vocabulary; 15 items). Each subtask requires a basal set of three consecutive correct responses and is discontinued after six consecutive scores of zero.	• Each item is scored as correct (1) or incorrect (0). • The total raw score is the sum of all item scores. • The raw score ranges from 0 to 41.
Numerical Ability	Wechsler Individual Achievement Test–Second UK Edition (WIAT–II ^{UK}) Numerical Operations subtest	Tasks include identifying and writing numbers, counting using one-to-one correspondence, and solving written calculation problems and simple equations involving addition, subtraction, multiplication, and division. There are 54 items in total. A basal set of three consecutive correct responses is required; testing is discontinued after six consecutive scores of zero.	• Each item is scored as correct (1) or incorrect (0). • The total raw score is the sum of all item scores. • The raw score ranges from 0 to 54.
Fluid Reasoning	Wechsler Abbreviated Scale of Intelligence–Second Edition (WASI–II) Matrix Reasoning subtest	While viewing an incomplete matrix or series, the participant selects the response option that completes the matrix or series from five alternatives. There are 24 items for children aged 6 to 8 years; testing is discontinued after three consecutive scores of zero.	• Each item is scored as correct (1) or incorrect (0). • The total raw score is the sum of all item scores. • The raw score ranges from 0 to 24 for children aged between 6 to 8 years.
Word-Problem-Solving Accuracy	Word Problems	Additive word problems that require either addition or subtraction of the values provided in the problem. Various representational tools such as counters (MathLink cubes) and paper and pencil are provided. There are 24 items in total, and all items are administered.	• The problem-solving accuracy of each item is scored as correct (1) or incorrect (0). • The total raw score is the sum of all item scores. • The raw score ranges from 0 to 24.
Word-Problem-Solving Strategy			• The problem-solving strategy of each item is coded using dimensions of arithmetic strategy and representation of quantitative structure (see Table 1-3).

2.3.1 Overarching Reliability and Validity Considerations

In selecting and designing the tasks, due diligence was exercised to address concerns regarding reliability and validity. Validity refers to whether a task measures what it is intended to measure, and reliability refers to the consistency with which it does so (De Vaus, 2001). First, to enhance construct validity (i.e., the extent to which a measure accurately represents the theoretical construct it is intended to measure, De Vaus, 2014), subtests from standardized instruments were used wherever possible (Johnson & Christensen, 2020). The standardized tests employed were the Wechsler Abbreviated Scale of Intelligence–Second Edition (WASI–II) and the Wechsler Individual Achievement Test–Second UK Edition (WIAT–II^{UK}). The researcher obtained written permission to use these standardized tests from a professional psychologist. Prior to administering the instruments, the administration procedures were reviewed thoroughly to ensure standardized implementation and support the reliability and validity of the assessment results.

The WASI–II was developed as a valid, effective, and brief assessment of intelligence for individuals aged 6 to 90 years (Wechsler, 2011). It comprises four subtests (Vocabulary, Similarities, Block Design, and Matrix Reasoning), which are similar in format to their counterparts in the Wechsler Intelligence Scale for Children–Fourth Edition and the Wechsler Adult Intelligence Scale–Fourth Edition. The WIAT–II^{UK} assesses academic skills in children and adolescents aged 4 years to 16 years, 11 months (Wechsler, 2005), and comprises nine subtests contributing to four composite scores: Reading, Mathematics, Written Language, and Oral Language.

Note that although the WIAT–III^{UK} is the most recent UK edition, it was not accessible to the researcher. Therefore, the WIAT–II^{UK} was used to assess language and numerical abilities. Differences between the WIAT–II^{UK} and WIAT–III^{UK}—such as updated norms, revised subtests, and additional content—may limit the direct comparability of scores and findings with studies that have used the latest edition. However, the WIAT–II^{UK} was considered acceptable for two reasons: (1) raw scores, rather than standardized scores, were used in the analyses, reducing the importance of updated norms; and (2) the selected subtests were used as covariates rather than primary predictor or outcome variables, reducing the likelihood that differences between editions influenced the study’s primary findings.

Second, when standardized tests were not available, tasks that had been widely used in prior research were employed. To enhance construct validity, tasks were selected that closely

aligned with the theoretical definitions of the constructs. For instance, the Visual–Spatial Reasoning task was selected as one of the measures of spatial visualization because it has been widely used in prior research with young primary school pupils, albeit in a Canadian context, and because it aligns closely with the present thesis’s definition of spatial visualization. Similarly, PathSpan, a digitized version of the Corsi Block-Tapping task, was selected because it has been widely used in prior research.

Third, several steps were taken to ensure the reliability and validity of the one-step additive word problems during the development phase (for details, see [Subsection 2.3.3.5](#)). The word problems were developed through an iterative process by adapting them from the literature. These problems underwent three rounds of review and revision, including two rounds of pilot testing and one round of expert consultation with a psychologist specializing in children’s mathematical development.

Given that language and numerical features influence word-problem-solving success (Verschaffel et al., 2020), efforts were made to control for these factors. To minimize the effect of language features, the word problems were developed with minimal variation in wording and structure. The continuous-quantity problems were designed to align closely with their discrete-quantity counterparts to ensure consistency without compromising clarity. Likewise, the dynamic-relation problems were designed to align closely with their static-relation counterparts. The word problems were condensed to make them as concise as possible without compromising situational understanding (Stern & Lehrndorfer, 1992).

Likewise, to minimize the effect of numerical features (Daroczy et al., 2015; Vessonen et al., 2024), 24 unique number facts with sums ranging from 11 to 19 were systematically randomized to create four parallel versions (Form A, Form B, Form C, and Form D) of the word problems. The wording and structure of the problems were identical across all forms; the only difference lay in the specific numerical values presented within each problem. This approach was intended to mitigate the potential influence of numerical values on item difficulty and problem-solving strategies. The selection of number facts and the randomization of number facts into four alternative test forms are detailed in [Subsection 2.3.3.1](#).

Furthermore, the validity of the problem-type categorization was examined through expert review. A mathematics educator with a PhD in mathematics education was invited to classify the 24 word problems according to the two binary dimensions of problem categorization (i.e., the 2×2 conceptual framework, introduced in [Subsection 1.4.1](#)): (1)

quantity (discrete quantity versus continuous quantity), and (2) quantitative relation (static relation versus dynamic relation). Inter-rater agreement was 100%, indicating strong congruence between the expert's classification and the intended categorization of the problems.

2.3.2 Established Measures

This subsection describes the established measures used to assess spatial ability ([Subsection 2.3.2.1](#)), language ability ([Subsection 2.3.2.2](#)), numerical ability ([Subsection 2.3.2.3](#)), and fluid reasoning ([Subsection 2.3.2.4](#)). Spatial visualization was measured using the WASI-II Block Design subtest and the Visual-Spatial Reasoning task. Visuospatial working memory was measured using a digitized version of the Corsi Block-Tapping task. Language ability and numerical ability were measured using the WIAT-II^{UK} Listening Comprehension and Numerical Operations subtests, respectively. Fluid reasoning was measured using the WASI-II Matrix Reasoning subtest. Detailed descriptions of the measures, including their reliability, validity, and administration procedures, are provided below.

2.3.2.1 Spatial Ability

Three measures of spatial ability were used. The WASI-II Block Design subtest and the Visual-Spatial Reasoning task were used as measures of spatial visualization, and PathSpan was used as a measure of visuospatial working memory. As introduced in Chapter 1 ([Subsection 1.4.3](#)), spatial visualization (Hawes & Ansari, 2020) and visuospatial working memory (Allen et al., 2019) are two spatial skills that have been most consistently identified as having particularly strong associations with mathematical thinking. Spatial visualization refers to “the ability to generate, recall, maintain, and manipulate visual-spatial images and solutions” (Hawes & Ansari, 2020, p. 466). Visuospatial working memory refers to the ability to temporarily store and manipulate visuospatial information (Baddeley & Hitch, 1974).

2.3.2.1.1 Spatial Visualization

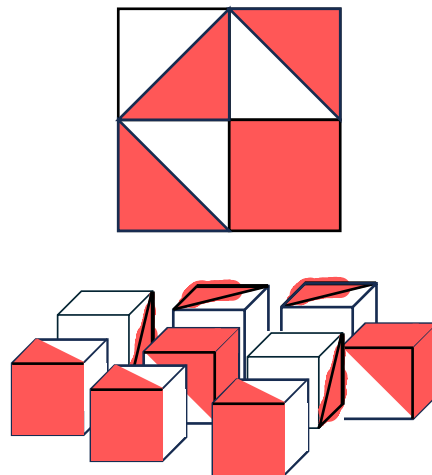
The WASI-II Block Design subtest has been widely used in the literature as a measure of spatial visualization (Atit et al., 2022). As shown in Figure 2-1, while viewing a constructed model or a picture in the Stimulus Book, the examinee uses red-and-white cubes to recreate the design within a specified time limit. Each red-and-white cube has two red faces, two white faces, and two half-red and half-white faces. The number of cubes required to recreate the design ranges from two to nine. The score depends on the time window within which the design is recreated, such that faster completion yields a higher score. No points are awarded if the recreation exceeds the time limit, which ranges from 30 to 120 seconds per design. There are

11 items for children aged 6 to 8 years; testing is discontinued after two consecutive scores of zero. Scores range from 0 to 2 for the first four items, and from 0 to 7 for the final seven items. The total raw score is the sum of all item scores. The raw score ranges from 0 to 57 for children aged between 6 to 8 years.

The WASI-II Block Design subtest has demonstrated high reliability and validity (Wechsler, 2011). As reported in the test manual, the split-half reliability coefficient for children aged 6–7 years is $r = .83$; the test-retest reliability coefficient for the 6–11-year-old age band, calculated using Pearson's product-moment correlation, is $r = .80$; and interscorer agreement is $r = .98$. Convergent validity has been examined in relation to other standardized measures, including the Wechsler Intelligence Scale for Children–Fourth Edition Block Design subtest ($r = .83$) and the Kaufman Brief Intelligence Test–Second Edition Nonverbal composite score ($r = .69$).

Figure 2-1

Wechsler Abbreviated Scale of Intelligence–Second Edition (WASI-II) Block Design

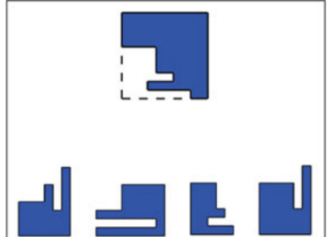
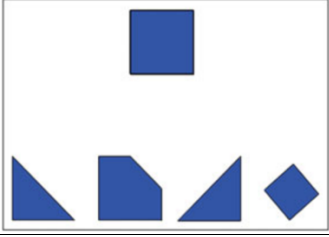
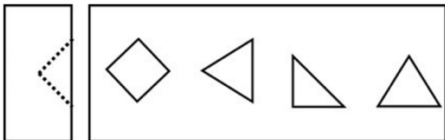


The Visual–Spatial Reasoning task is a measure of spatial visualization adopted from Hawes et al. (2017). This task has been widely used with young primary school pupils in Canadian schools (e.g., Hawes et al., 2015, 2017, 2019b; Whitehead & Hawes, 2023); however, to the best of the researcher's knowledge, it has not previously been used in England. This measure was selected because it aligns closely with the present thesis's definition of spatial visualization. As shown in Figure 2-2, the task comprises four subtasks that require reasoning and visualizing solutions to problems involving parts of objects in relation to their whole: missing-center, missing-border, two-dimensional shape composition/decomposition, and paper-folding problems. There are 20 items in total, with five items per subtask. Each correct

response yields one point. All items are administered. The total raw score is the sum of all items scores. The raw score ranges from 0 to 20.

Figure 2-2

Subtasks of the Visual–Spatial Reasoning Task

Missing Centre ‘Which of these pieces fits perfectly into this white space?’	
Missing Border ‘Which of these pieces is the missing piece?’	
Two-Dimensional shape Composition/Decomposition ‘Which two shapes can be put together to make the target shape?’	
Paper Folding (Presented with a folded piece of paper with the outline of a shape along the fold) ‘Which shape would you get if you cut along the dotted lines and opened the paper?’	

*Note. Adapted from “Enhancing children’s spatial and numerical skills through a dynamic spatial approach to early geometry instruction: Effects of a 32-week intervention,” by Z. Hawes, J. Moss, B. Caswell, S. Naqvi, & S. MacKinnon, *Cognition and Instruction*, 35(3), p. 251 (<https://doi.org/10.1080/07370008.2017.1323902>). Copyright 2017 by Taylor & Francis. Reprinted with permission.*

Two measures of spatial visualization—WASI-II Block Design and Visual–Spatial Reasoning task—were used to address the limitations of each measure. The WASI-II Block Design is a standardized test with high reliability and validity (Wechsler, 2011) and is widely used as a measure of spatial visualization because it involves “visualizing and/or mentally manipulating objects or figures” (Atit et al., 2022, p. 708). However, in some studies, the WASI-II Block Design has been used more broadly as a measure of nonverbal problem-solving ability (e.g., Fuchs et al., 2005). Conversely, while the Visual–Spatial Reasoning task

has primarily been used with pupils in Canadian schools, its four age-appropriate subtasks align closely with this thesis's definition of spatial visualization.

Consistent with prior research (Casey et al., 2015; Song et al., 2013; Schenck & Nathan, 2024), a composite score of spatial visualization was computed from the two spatial visualization measures. Principal Component Analysis (PCA) was selected as a data-reduction technique to derive a parsimonious summary index of shared variance across the two measures, rather than to estimate a latent measurement model (De Pauw et al., 2009; Song et al., 2013). Given that only two indicators were available, factor analysis would have been just-identified, precluding the evaluation of model fit and providing no additional information beyond the observed correlations.

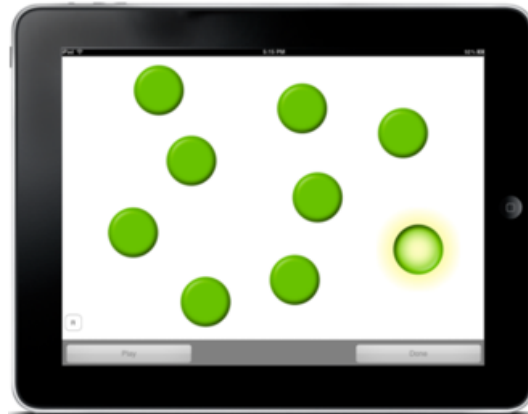
Preliminary analyses indicated that raw WASI-II Block Design scores were positively skewed, whereas scores on the Visual-Spatial Reasoning task were approximately normally distributed. Given the non-normality of one indicator, a Spearman rank-order correlation was conducted, revealing a strong positive association between the two measures ($\rho = .61, p < .001$; Cohen, 1988). PCA, based on the correlation matrix, yielded a first principal component with an eigenvalue of 1.65, accounting for 82.6% of the total variance. Both measures loaded strongly and equally on this component (loading = .91). The resulting component scores were saved and used as a continuous measure of spatial visualization in Studies 1–3.

2.3.2.1.2 Visuospatial Working Memory

PathSpan (Hume, 2019), an iPad-based application, was used as a measure of sequential visuospatial working memory. It is a digitized version of the Corsi Block-Tapping task (Corsi, 1972). PathSpan has been used with young children in various research studies (e.g., Coolen & Castronovo, 2023; Hawes et al., 2019b; Rittle-Johnson et al., 2019; Valcan et al., 2020). As shown in Figure 2-3, participants are presented with nine dots arranged irregularly, some of which light up one at a time in a random sequence. Participants then replicate the sequence by touching the dots in the same order in which they were illuminated.

Figure 2-3

PathSpan: A Digitized Corsi Block-Tapping Test



Note. From *Screenshot of PathSpan Interface* [Image], by T. Hume & S. Hume (2014), <https://hume.ca/ix/pathspan/>

The forward span was used as the measure of visuospatial working memory. While both the forward and backward span conditions recruit visuospatial working memory resources (Berch et al., 1998; Kessels et al., 2008; Vandierendonck et al., 2004), the two conditions differ in their additional cognitive demands. The forward span relies primarily on the visuospatial sketchpad, whereas the backward span places greater demands on the central executive due to the requirement to mentally reverse the sequence (Vandierendonck et al., 2004; Hester et al., 2004). To ensure that the task primarily captured visuospatial working memory as a spatial skill, rather than the domain-general executive demands associated with manipulating sequence, only the forward span was used.

To ensure that participants understood the instructions, a practice round with a sequence length of two was repeated until the participant responded correctly. The actual test also began with a sequence length of two. For each sequence length, there were two trials with different patterns. The sequence length increased by one if the participant correctly repeated the pattern on at least one of the two trials at a given length. Testing was discontinued if the participant failed both trials at a given sequence length. The final score was the highest sequence length that the child accurately repeated at least once before the task ended. The maximum sequence length (i.e., the highest possible score) was nine. The sequencing patterns were identical across participants.

Participants were required to press the ‘Start’ button to initiate the flashing sequence. All buttons were disabled while the sequence was flashing. When the flashing sequence ended, a tone signaled the participant to reproduce the sequence. After completing the reproduction,

the participant pressed the ‘Done’ button. Although PathSpan can be configured manually, the default settings (i.e., layout of dots, sequencing patterns, and 600 ms of flashes followed by a 500 ms delay between flashes) were used, as they were based on the ‘Count Me In’ longitudinal study with over 500 children, which reported acceptable internal consistency reliability (Cronbach’s $\alpha = .70$; LeFevre et al., 2010).

2.3.2.2 Language Ability

The WIAT-II^{UK} Listening Comprehension subtest (Wechsler, 2005) was used as a measure of language ability. In this thesis, listening comprehension was considered a more relevant measure of language ability than reading comprehension because the one-step additive word problems were not only presented in written form but also read aloud to the participants (Fuchs et al., 2006; Kyttälä et al., 2014; Spencer et al., 2022). The Listening Comprehension subtest assesses the ability to listen for details. It comprises three subtasks: Receptive Vocabulary (16 items), Sentence Comprehension (10 items), and Expressive Vocabulary (15 items). The subtasks include selecting the picture that matches a word or sentence and generating a word that corresponds to a picture and oral description. Each subtask requires a basal set of three consecutive correct responses and is discontinued after six consecutive incorrect responses. Each correct response yields one point. The total raw score is the sum of all item scores. The total raw score ranges from 0 to 41.

The WIAT-II^{UK} Listening Comprehension subtest has demonstrated high reliability and validity (Wechsler, 2005). As reported in the test manual, the split-half reliability coefficient for the child sample (aged 6–7 years) is $r = .83$; the test-retest reliability coefficient for the 6–9-year-old age band, corrected for the variability of the standardization sample, is $r = .88$. Because this subtest is scored objectively (i.e., responses are scored as either correct or incorrect), interscorer agreement approaches 100%. Convergent validity has been examined in relation to other standardized measures, including the Peabody Picture Vocabulary Test–Third Edition ($r = .44$) and the Academic Competence Evaluation Scales Reading/Language Arts Subscale ($r = .55$).

2.3.2.3 Numerical Ability

The WIAT-II^{UK} Numerical Operations subtest (Wechsler, 2005) was used as a measure of numerical ability. This pencil-and-paper task assesses the ability to identify and write numbers, count using one-to-one correspondence, and solve calculation problems and simple equations involving the basic operations of addition, subtraction, multiplication, and division.

This subtest requires a basal set of three consecutive correct responses. Although there are 54 items in total, testing is discontinued after six consecutive incorrect responses. Each correct response yields one point, and the total raw score is the sum of all item scores. The maximum possible raw score is 54, and the minimum possible score is zero.

The WIAT-II^{UK} Numerical Operations subtest has demonstrated high reliability and validity (Wechsler, 2005). As reported in the test manual, the split-half reliability coefficient for the child sample (aged 6–7 years) is $r = .85$; the test-retest reliability coefficient for the 6–9-year-old age band, corrected for variability in the standardization sample, is $r = .92$. Because this subtest is scored objectively (i.e., responses are scored as either correct or incorrect), interscorer agreement approaches 100%. Convergent validity has been examined in relation to other standardized measures, such as the Wide Range Achievement Test–Third Edition Arithmetic subtest ($r = .77$) and the Differential Ability Scales Basic Number Skills subtest ($r = .75$).

2.3.2.4 Fluid Reasoning

Non-verbal fluid reasoning was measured using the WASI-II Matrix Reasoning subtest (for a review, see Atit et al., 2022). Although fluid reasoning tests may involve visualization or visuospatial working memory, they are distinguished from purely spatial ability tests because they require inductive and deductive reasoning (Schneider & McGrew, 2012). Each item comprises an incomplete matrix or series. The participant selects the response option that completes the matrix or series from five alternatives. Each correct response yields one point. There are 24 items in total for children aged 6 to 8 years; testing is discontinued after three consecutive scores of zero. The raw score ranges from 0 to 24 for children aged 6 to 8 years.

The WASI-II Matrix Reasoning subtest has demonstrated high reliability and validity (Wechsler, 2011). As reported in the test manual, the split-half reliability coefficient for children aged 6–7 years is $r = .89$; the test-retest reliability coefficient for the 6–11-year-old age band, calculated using Pearson's product-moment correlation, is $r = .81$; and interscorer agreement is $r = .99$. Convergent validity has been examined in relation to other standardized measures, such as the Wechsler Intelligence Scale for Children–Fourth Edition (WISC-IV) Matrix Reasoning subtest ($r = .73$) and the Kaufman Brief Intelligence Test–Second Edition (KBIT-2) Nonverbal composite score ($r = .80$).

2.3.3 One-Step Additive Word Problems

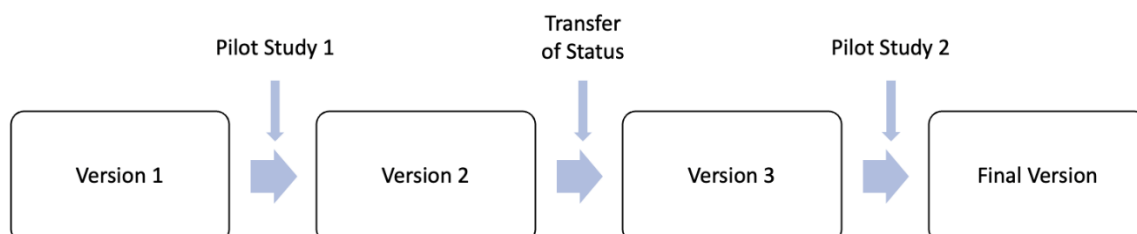
The one-step additive word problems, developed following the 2×2 conceptual framework introduced in [Subsection 1.4.1](#), are an original measure developed for this thesis. This subsection focuses on the methodological development and implementation of the word problems and the associated coding procedures used to assess solution accuracy and problem-solving strategies. Specifically, it first describes the word problem development process, including two pilot studies, the selection of number facts, and the randomization of number facts into four alternative test forms ([Subsection 2.3.3.1](#)), before presenting the final set of 24 word problems across these four forms ([Subsection 2.3.3.2](#)). The administration procedure is then detailed ([Subsection 2.3.3.3](#)), followed by a description of the coding schemes ([Subsection 2.3.3.4](#)). Finally, evidence for the reliability and validity of the one-step additive word problem measure is presented ([Subsection 2.3.3.5](#)).

2.3.3.1 Development of the Word Problems

As shown in Figure 2-4, one-step additive word problems based on the proposed 2×2 conceptual framework were developed through an iterative process. An initial version (Version 1) was designed and tested in a pilot study, resulting in minor revisions to wording and administration procedures and leading to Version 2. Following feedback from the Transfer of Status assessors, the original focus on quantity-only problems was replaced with a framework centered on problems involving two quantities linked by a quantitative relation. Consequently, the conceptual framework and the word problems were reformulated, resulting in Version 3. A second pilot study was then conducted to evaluate Version 3, leading to further refinements and the development of the final version.

Figure 2-4

Iterative Process for Developing One-Step Additive Word Problems



[Subsection 2.3.3.1.1](#) presents the first pilot study and the revisions it informed, followed by [Subsection 2.3.3.1.2](#), which presents the second pilot study and its resulting revisions. Next,

[Subsection 2.3.3.1.3](#) outlines the process used to select the 24 unique number facts included in the word problems. Finally, [Subsection 2.3.3.1.4](#) describes the randomization of these number facts to create four alternative forms of the word problem sets.

2.3.3.1.1 First Pilot Study

The first pilot study involved five Year 2 children in South East England, recruited through their parents via a pre-existing personal network. It had two main aims: (1) to determine the overall administration procedure for the seven tasks, and (2) to evaluate the appropriateness of the one-step additive word problems (Version 1 in Figure 2-4).

Findings from this pilot study indicated that administering all tasks required approximately one hour and that a break was necessary. These observations informed the decision to conduct two testing sessions, each lasting approximately 30 minutes and typically spaced at least one hour apart (but no less than 30 minutes). For the word problems, one minute per problem was found to be sufficient—rather than the three minutes typically allocated for older children in comparable studies (e.g., Boonen et al., 2013, 2014; Hegarty & Kozhevnikov, 1999). It also became evident that repeating the scenario before each problem to establish situational context (Stern & Lehrndorfer, 1992) was unnecessary and overly time-consuming. Accordingly, the procedure was revised so that each scenario was stated only once before presenting its associated problems together in randomized order.

The initial versions of the problems (Versions 1 and 2 in Figure 2-4) were adapted from Sophian and McCorgay (1994) and Nunes et al. (2009a). These versions focused exclusively on problems involving three quantities and used number facts up to 10. Following the pilot study, minor revisions were made for clarity; for example, temporal markers such as ‘first’ and ‘then’ were introduced to clarify the sequence of events within each problem.

Following feedback from the Transfer of Status assessors—specifically, that quantity-only problems might produce ceiling effects for Year 2 pupils—the one-step additive word problems were revised. The original emphasis on quantity-only problems was replaced with a focus on problems involving two quantities linked by a quantitative relation. Accordingly, the conceptual framework was reformulated, and 24 word problems were developed (Version 3 in Figure 2-4). These revised problems were adapted from Carpenter et al. (1999) and Nunes et al. (2009a). Version 3 aligns with the thesis’s final 2×2 conceptual framework (Quantity $\times$ Quantitative Relation) for categorizing one-step additive word problems.

2.3.3.1.2 Second Pilot Study

Version 3 (in Figure 2-4) was evaluated in a second pilot study involving fourteen Year 2 children recruited through a school in South East England. This pilot study had three aims: (1) to evaluate the appropriateness of the revised word problems, (2) to refine the operationalization of problem-solving strategies, and (3) to refine the administration procedure of the 24 word problems.

First, observations indicated that the word problems in Version 3 (i.e., comprising two quantities and one quantitative relation) were at an appropriate level of difficulty for 6–7-year-olds. However, pilot testing, together with consultations with classroom teachers, suggested that number facts up to 20 were more appropriate than limiting them to 10. This was because a wider range of problem-solving strategies was observed for number facts above 10. This also aligned with the English national curriculum requirement that Year 2 pupils “recall and use addition and subtraction facts to 20 fluently” (Department for Education, 2013, p. 108). Furthermore, pilot observations suggested that using alternative problem forms could help reduce the influence of specific numerical characteristics (Daroczy et al., 2015). Accordingly, the final version of the word problems was constructed using 24 unique number facts with sums ranging from 11 to 19, which were randomized to create four alternative test forms (for details, see [Subsection 2.3.3.1](#)).

Second, most children did not spontaneously use pencil and paper as a representational tool, in contrast to older children in similar studies (e.g., Boonen et al., 2013, 2014; Hegarty & Kozhevnikov, 1999). When asked how they arrived at their answers, most focused on their computation process (e.g., ‘I added,’ ‘three plus six is nine’). These findings suggested that focusing solely on visual representations was too narrow and that a broader conceptualization of problem-solving strategies was required. In particular, it became clear that problem-solving strategies should integrate both representations of quantitative structure and arithmetic strategies, rather than treating representational and arithmetic approaches separately.

Third, following the second pilot study, a number of refinements to the administration procedure were made. Presenting a single word problem per A4 sheet was deemed unnecessary; presenting four problems on a single sheet was considered sufficient should the child choose to write/draw. Moreover, checking comprehension through clarifying questions (e.g., ‘What does it mean to walk towards something?’) was found to be more effective than presenting illustrative pictures, which appeared to encourage the production of iconic visual

representations that focused on the problem's surface details rather than its quantitative structure.

Furthermore, the order of problem administration was refined. For instance, the pilot testing suggested that fully randomizing all 24 word problems required children to repeatedly reorient themselves to different problem types, which was cognitively demanding and time-consuming. It was also observed that children were more engaged when they were initially presented with easier problems that allowed early success (Hattie, 2009). Informed by these pilot observations, the final administration order with predefined sequencing constraints was established and is detailed in [Subsection 2.3.3.3](#).

2.3.3.1.3 Selection of Number Facts

To arrive at the final set of 24 unique number facts, a systematic elimination process was undertaken. First, as shown in Figure 2-5, from the 75 unique number facts with sums ranging from 11 to 20, four double (tie) number facts (e.g., 6, 6, 12) were eliminated because research indicates that doubles are often retrieved more fluently than other number facts, reflecting a tie effect in arithmetic fact retrieval (Ashcraft & Fierman, 1982). This first round of elimination reduced the pool to 71 unique number-fact combinations.

Figure 2-5

Elimination of Double (Tie) Number Facts

1, 19, 20	1, 18, 19	2, 16, 18	3, 14, 17	5, 11, 16	7, 8, 15	3, 10, 13	1, 10, 11
2, 18, 20	2, 17, 19	3, 15, 18	4, 13, 17	6, 10, 16	1, 13, 14	4, 9, 13	2, 9, 11
3, 17, 20	3, 16, 19	4, 14, 18	5, 12, 17	7, 9, 16	2, 12, 14	5, 8, 13	3, 8, 11
4, 16, 20	4, 15, 19	5, 13, 18	6, 11, 17	8, 8, 16	3, 11, 14	6, 7, 13	4, 7, 11
5, 15, 20	5, 14, 19	6, 12, 18	7, 10, 17	1, 14, 15	4, 10, 14	1, 11, 12	5, 6, 11
6, 14, 20	6, 13, 19	7, 11, 18	8, 9, 17	2, 13, 15	5, 9, 14	2, 10, 12	
7, 13, 20	7, 12, 19	8, 10, 18	1, 15, 16	3, 12, 15	6, 8, 14	3, 9, 12	
8, 12, 20	8, 11, 19	9, 9, 18	2, 14, 16	4, 11, 15	7, 7, 14	4, 8, 12	
9, 11, 20	9, 10, 19	1, 16, 17	3, 13, 16	5, 10, 15	1, 12, 13	5, 7, 12	
10, 10, 20	1, 17, 18	2, 15, 17	4, 12, 16	6, 9, 15	2, 11, 13	6, 6, 12	

Second, as shown in Figure 2-6, number facts involving operands from one to four were eliminated, as children were less likely to use observable external strategies with smaller numbers. In such cases, it was difficult to determine whether a child was using a counting-based strategy or a number-fact strategy. Moreover, there is debate over whether recalling number facts involving small operands (≤ 4) reflects true retrieval from long-term memory

(Ashcraft & Fierman, 1982) or an automatized counting-based procedure that has become rapid and unconscious (Barrouillet & Thevenot, 2013). This second round of elimination reduced the pool to 30 unique number-fact combinations.

Figure 2-6

Elimination of Number Facts Involving operands 1 to 4

1, 19, 20	1, 18, 19	2, 16, 18	3, 14, 17	5, 11, 16	7, 8, 15	3, 10, 13	1, 10, 11
2, 18, 20	2, 17, 19	3, 15, 18	4, 13, 17	6, 10, 16	1, 13, 14	4, 9, 13	2, 9, 11
3, 17, 20	3, 16, 19	4, 14, 18	5, 12, 17	7, 9, 16	2, 12, 14	5, 8, 13	3, 8, 11
4, 16, 20	4, 15, 19	5, 13, 18	6, 11, 17	8, 8, 16	3, 11, 14	6, 7, 13	4, 7, 11
5, 15, 20	5, 14, 19	6, 12, 18	7, 10, 17	1, 14, 15	4, 10, 14	1, 11, 12	5, 6, 11
6, 14, 20	6, 13, 19	7, 11, 18	8, 9, 17	2, 13, 15	5, 9, 14	2, 10, 12	
7, 13, 20	7, 12, 19	8, 10, 18	1, 15, 16	3, 12, 15	6, 8, 14	3, 9, 12	
8, 12, 20	8, 11, 19	9, 9, 18	2, 14, 16	4, 11, 15	7, 7, 14	4, 8, 12	
9, 11, 20	9, 10, 19	1, 16, 17	3, 13, 16	5, 10, 15	1, 12, 13	5, 7, 12	
10, 10, 20	1, 17, 18	2, 15, 17	4, 12, 16	6, 9, 15	2, 11, 13	6, 6, 12	

Third, number facts involving 20 were excluded, as multi-digit number processing draws on “additional representations, mental manipulations, and processing of inter-digit relations beyond ... single-digit numbers” (Nuerk et al., 2015, p. 107). For example, addition problems may require carrying over and regrouping two tens, whereas subtraction problems may require borrowing from the tens column, with one ten still remaining. As shown in Figure 2-7, this third round of elimination reduced the pool to 25 unique number-fact combinations.

Figure 2-7

Elimination of Number Facts Involving 20

1, 19, 20	1, 18, 19	2, 16, 18	3, 14, 17	5, 11, 16	7, 8, 15	3, 10, 13	1, 10, 11
2, 18, 20	2, 17, 19	3, 15, 18	4, 13, 17	6, 10, 16	1, 13, 14	4, 9, 13	2, 9, 11
3, 17, 20	3, 16, 19	4, 14, 18	5, 12, 17	7, 9, 16	2, 12, 14	5, 8, 13	3, 8, 11
4, 16, 20	4, 15, 19	5, 13, 18	6, 11, 17	8, 8, 16	3, 11, 14	6, 7, 13	4, 7, 11
5, 15, 20	5, 14, 19	6, 12, 18	7, 10, 17	1, 14, 15	4, 10, 14	1, 11, 12	5, 6, 11
6, 14, 20	6, 13, 19	7, 11, 18	8, 9, 17	2, 13, 15	5, 9, 14	2, 10, 12	
7, 13, 20	7, 12, 19	8, 10, 18	1, 15, 16	3, 12, 15	6, 8, 14	3, 9, 12	
8, 12, 20	8, 11, 19	9, 9, 18	2, 14, 16	4, 11, 15	7, 7, 14	4, 8, 12	
9, 11, 20	9, 10, 19	1, 16, 17	3, 13, 16	5, 10, 15	1, 12, 13	5, 7, 12	
10, 10, 20	1, 17, 18	2, 15, 17	4, 12, 16	6, 9, 15	2, 11, 13	6, 6, 12	

Finally, as only 24 number facts were required, the number combination 5, 10, 15 was removed because pilot findings suggested it was disproportionately easier than the number facts containing other number combinations. Figure 2-8 presents the remaining 24 number facts that were used to construct four alternative versions of the word problems.

Figure 2-8

Final Set of 24 Number Facts (Sums 11–20)

5, 14, 19	9, 10, 19	8, 10, 18	8, 9, 17	6, 9, 15	5, 8, 13
6, 13, 19	5, 13, 18	5, 12, 17	5, 11, 16	7, 8, 15	6, 7, 13
7, 12, 19	6, 12, 18	6, 11, 17	6, 10, 16	5, 9, 14	5, 7, 12
8, 11, 19	7, 11, 18	7, 10, 17	7, 9, 16	6, 8, 14	5, 6, 11

2.3.3.1.4 Randomization of Number Facts into Four Alternative Forms

The 24 unique number facts were subjected to a randomization process to create four alternative forms of the word problem sets (Forms A–D). As shown in Table 2-3, the number facts were first grouped into four sets through a series of three steps. First, one of the four number facts involving 10 was assigned to each group. Second, the ten number facts containing two two-digit numbers were distributed across the groups. Finally, the remaining ten number facts (each containing a single two-digit number) were allocated across the groups.

Table 2-3

Randomization Process of Number Facts into Groups

	Number-Fact Group 1	Number-Fact Group 2	Number-Fact Group 3	Number-Fact Group 4
Step 1: Assign the four number facts containing 10 to each group.	6, 10, 16	7, 10, 17	8, 10, 18	9, 10, 19
Step 2: Assign the ten number facts with two two-digit numbers.	8, 11, 19 5, 13, 18 5, 12, 17	7, 12, 19 6, 12, 18 5, 11, 16	6, 13, 19 7, 11, 18	5, 14, 19 6, 11, 17
Step 3: Assign the remaining ten number facts with a single two-digit number.	6, 7, 13 7, 8, 15	8, 9, 17 5, 6, 11	5, 9, 14 7, 9, 16 5, 7, 12	6, 9, 15 6, 8, 14 5, 8, 13

Next, as shown in Table 2-4, all four number-fact groups were included in each of the four alternative test forms and were randomly assigned to the different word problem types (i.e., problem scenarios). Moreover, within each problem type, each of the six number facts in a number-fact group was randomly paired with one of the six problems. Additional considerations were made when incorporating the number facts into each problem. If the two smaller numbers from a number fact were used in the problem, the smaller of the two was always presented first to allow for the observation of whether the participant counted on from

the smaller or the larger number. If the largest number from a number fact had to be used, then between the two smaller numbers, the one-digit number was always selected to ensure that no problem presented two two-digit numbers. If both smaller numbers were one-digit numbers, one was chosen at random to be used in the problem.

Table 2-4

Randomization Process of Number-Fact Groups into Alternative Test Forms

	Number-Fact Group 1	Number-Fact Group 2	Number-Fact Group 3	Number-Fact Group 4
Form A	Scenario 2	Scenario 3	Scenario 4	Scenario 1
Form B	Scenario 4	Scenario 1	Scenario 2	Scenario 3
Form C	Scenario 3	Scenario 2	Scenario 1	Scenario 4
Form D	Scenario 1	Scenario 4	Scenario 3	Scenario 2

2.3.3.2 Final Word Problem Set

Figure 2-9 presents the final set of 24 word problems developed through the iterative process described above. In each problem, x and y represent two values, with x denoting the smaller value and y the larger value. The addition and subtraction signs indicate whether the correct answer is obtained by adding or subtracting the two values. Four alternative versions of the problem set (Form A, Form B, Form C, and Form D) were created using the number-fact randomization process described above. Table 2-5, Table 2-6, Table 2-7, and Table 2-8 present the specific values used in each problem across the four test forms.

A different scenario (shown in blue cells of Figure 2-9) was associated with each word problem type. The Discrete Quantity–Dynamic Relation problems (i.e., the top-left quadrant in Figure 2-9; Items 1–6) involved Ann playing a game of marbles by herself. These problems required solving for the number of marbles Ann had either at the end of the game (Items 1–2), at the beginning of the game (Items 3–4), or the number of marbles she won or lost during the game (Items 5–6).

The Discrete Quantity–Static Relation problems (i.e., the top-right quadrant in Figure 2-9; Items 7–12) involved Ann and Bob playing a game of marbles together. These problems required solving either for the number of marbles Bob had (Items 7–10) or for the difference between the number of marbles Ann and Bob had (Items 11–12). Within Items 7–10, the referent quantity (i.e., the quantity used as the basis for comparison) differed: in Items 7–8, Ann’s quantity served as the referent, whereas in Items 9–10, Bob’s quantity served as the referent.

Figure 2-9*Final Set of One-Step Additive Word Problems*

		Quantitative Relation					
		Dynamic Relation			Static Relation		
Quantity	Discrete Quantity	“Ann was playing a game of marbles. Before the game she had some marbles. Then she played a game of marbles where she either won more marbles or lost some of her marbles.”			“Ann and Bob are going to play a game of marbles. So, Ann brought some marbles, and Bob also brought some marbles to play with.”		
		1	Ann had x marbles. Then she played a game of marbles and won y more marbles. How many marbles does she have now?	+	7	Ann has x marbles. Bob has y more marbles than Ann. How many marbles does Bob have?	+
		2	Ann had y marbles. Then she played a game of marbles and lost x marbles. How many marbles does she have now?	−	8	Ann has y marbles. Bob has x marbles less than Ann. How many marbles does Bob have?	−
		3	Ann had some marbles. Then she played a game of marbles and won x more marbles. Now she has y marbles. How many marbles did she have to begin with?	−	9	Ann has y marbles. She has x more marbles than Bob. How many marbles does Bob have?	−
		4	Ann had some marbles. Then she played a game of marbles and lost x marbles. Now she has y marbles left. How many marbles did she have to begin with?	+	10	Ann has x marbles. She has y marbles less than Bob. How many marbles does Bob have?	+
		5	Ann had x marbles. Then she played a game of marbles. Now she has y marbles. How many marbles did she win?	−	11	Ann has y marbles. Bob has x marbles. How many less marbles than Ann does Bob have?	−
		6	Ann had y marbles. Then she played a game of marbles. Now she has x marbles left. How many marbles did she lose?	−	12	Ann has x marbles. Bob has y marbles. How many more marbles than Ann does Bob have?	−
	Continuous Quantity	“Ann was playing a game. She first stood some distance from the wall. Then she played a game where she either walked towards or walked away from the wall.”			“Ann and Bob were playing Red Light. Green Light. The wall is the finish line.”		
		13	Ann was x steps from the wall. Then she walked y steps away from the wall. How far is Ann from the wall now?	+	19	Ann is x steps from the wall. Bob is y steps farther from the wall than Ann. How far is Bob from the wall?	+
		14	Ann was y steps from the wall. Then she walked x steps towards the wall. How far is Ann from the wall now?	−	20	Ann is y steps from the wall. Bob is x steps closer to the wall than Ann. How far is Bob from the wall?	−
		15	Ann was some steps from the wall. Then she walked x steps away from the wall. Now she is y steps from the wall. How far was Ann from the wall to begin with?	−	21	Ann is y steps from the wall. She is x steps farther from the wall than Bob. How far is Bob from the wall?	−
		16	Ann was some steps from the wall. Then she walked x steps towards the wall. Now she is y steps from the wall. How far was Ann from the wall to begin with?	+	22	Ann is x steps from the wall. She is y steps closer to the wall than Bob. How far is Bob from the wall?	+
		17	Ann was x steps from the wall. Then she walked some steps. Now she is y steps from the wall. How many steps did Ann walk away from the wall?	−	23	Ann is y steps from the wall. Bob is x steps from the wall. How many steps closer to the wall is Bob than Ann?	−
		18	Ann was y steps from the wall. Then she walked some steps. Now she is x steps from the wall. How many steps did Ann walk towards the wall?	−	24	Ann is x steps from the wall. Bob is y steps from the wall. How many steps is Bob farther from the wall than Ann?	−

Note. The addition and subtraction signs indicate whether the correct answer can be obtained by adding or subtracting the x and y values.

Table 2-5*Discrete Quantity–Dynamic Relation Problems from Four Test Forms*

Problem	Form A			Form B			Form C			Form D		
	x value	y value	Answer	x value	y value	Answer	x value	y value	Answer	x value	y value	Answer
1. Ann had x marbles. Then she played a game of marbles and won y more marbles. How many marbles does she have now?	9	10	19	5	11	16	7	9	16	5	12	17
2. Ann had y marbles. Then she played a game of marbles and lost x marbles. How many marbles does she have now?	6	14	8	5	11	6	7	12	5	6	16	10
3. Ann had some marbles. Then she played a game of marbles and won x more marbles. Now she has y marbles. How many marbles did she have to begin with?	9	15	6	9	17	8	8	18	10	8	15	7
4. Ann had some marbles. Then she played a game of marbles and lost x marbles. Now she has y marbles left. How many marbles did she have to begin with?	6	11	17	7	12	19	6	13	19	5	13	18
5. Ann had x marbles. Then she played a game of marbles. Now she has y marbles. How many marbles did she win?	5	19	14	6	18	12	7	18	11	8	19	11
6. Ann had y marbles. Then she played a game of marbles. Now she has x marbles left. How many marbles did she lose?	8	13	5	10	17	7	5	14	9	7	13	6

Note. Items 1–2 are final-quantity unknown problems. Items 3–4 are initial-quantity unknown problems. Items 5–6 are dynamic-relation known problems.

Table 2-6*Discrete Quantity–Static Relation Problems from Four Test Forms*

Problem	Form A			Form B			Form C			Form D		
	x value	y value	Answer	x value	y value	Answer	x value	y value	Answer	x value	y value	Answer
7. Ann has x marbles. Bob has y more marbles than Ann. How many marbles does Bob have?	6	10	16	5	7	12	6	12	18	6	9	15
8. Ann has y marbles. Bob has x marbles less than Ann. How many marbles does Bob have?	5	18	13	7	16	9	7	19	12	5	19	14
9. Ann has y marbles. She has x more marbles than Bob. How many marbles does Bob have?	8	19	11	6	19	13	5	16	11	8	13	5
10. Ann has x marbles. She has y marbles less than Bob. How many marbles does Bob have?	7	8	15	5	9	14	5	6	11	6	8	14
11. Ann has y marbles. Bob has x marbles. How many less marbles than Ann does Bob have?	6	13	7	8	18	10	7	17	10	9	19	10
12. Ann has x marbles. Bob has y marbles. How many more marbles than Ann does Bob have?	5	17	12	7	18	11	8	17	9	6	17	11

Note. Items 7–8 are compared-quantity unknown problems. Items 9–10 are referent-quantity unknown problems. Items 11–12 are static-relation known problems.

Table 2-7*Continuous Quantity–Dynamic Relation Problems from Four Test Forms*

Problem	Form A			Form B			Form C			Form D		
	x value	y value	Answer	x value	y value	Answer	x value	y value	Answer	x value	y value	Answer
13. Ann was x steps from the wall. Then she walked y steps away from the wall. How far is Ann from the wall now?	5	9	14	6	7	13	6	11	17	7	10	17
14. Ann was y steps from the wall. Then she walked x steps towards the wall. How far is Ann from the wall now?	8	18	10	7	15	8	9	15	6	8	17	9
15. Ann was some steps from the wall. Then she walked x steps away from the wall. Now she is y steps from the wall. How far was Ann from the wall to begin with?	7	12	5	6	16	10	8	14	6	5	11	6
16. Ann was some steps from the wall. Then she walked x steps towards the wall. Now she is y steps from the wall. How far was Ann from the wall to begin with?	7	11	18	5	12	17	5	14	19	6	12	18
17. Ann was x steps from the wall. Then she walked some steps. Now she is y steps from the wall. How many steps did Ann walk away from the wall?	6	19	13	5	18	13	5	13	8	7	19	12
18. Ann was y steps from the wall. Then she walked some steps. Now she is x steps from the wall. How many steps did Ann walk towards the wall?	9	16	7	8	19	11	9	19	10	5	16	11

Note. Items 13–14 are final-quantity unknown problems. Items 15–16 are initial-quantity unknown problems. Items 17–18 are dynamic-relation known problems.

Table 2-8*Continuous Quantity–Static Relation Problems from Four Test Forms*

Problem	Form A			Form B			Form C			Form D		
	x value	y value	Answer	x value	y value	Answer	x value	y value	Answer	x value	y value	Answer
19. Ann is x steps from the wall. Bob is y steps farther from the wall than Ann. How far is Bob from the wall?	5	6	11	5	14	19	5	13	18	7	11	18
20. Ann is y steps from the wall. Bob is x steps closer to the wall than Ann. How far is Bob from the wall?	5	16	11	8	13	5	8	19	11	6	19	13
21. Ann is y steps from the wall. She is x steps farther from the wall than Bob. How far is Bob from the wall?	7	19	12	6	17	11	5	17	12	7	16	9
22. Ann is x steps from the wall. She is y steps closer to the wall than Bob. How far is Bob from the wall?	7	10	17	9	10	19	6	7	13	5	7	12
23. Ann is y steps from the wall. Bob is x steps from the wall. How many steps closer to the wall is Bob than Ann?	8	17	9	6	14	8	8	15	7	5	14	9
24. Ann is x steps from the wall. Bob is y steps from the wall. How many steps is Bob farther from the wall than Ann?	6	18	12	9	15	6	6	16	10	8	18	10

Note. Items 19–20 are compared-quantity unknown problems. Items 21–22 are referent-quantity unknown problems. Items 23–24 are static-relation known problems.

The Continuous Quantity–Dynamic Relation problems (i.e., the bottom-left quadrant in Figure 2-9; Items 13–18) involved Ann playing a game by herself in which she could walk either toward or away from the wall. These problems required solving for Ann’s position either at the end of the game (Items 13–14), at the beginning of the game (Items 15–16), or the number of steps she took during the game (Items 17–18).

The Continuous Quantity–Static Relation problems (i.e., the bottom-right quadrant in Figure 2-9; Items 19–24) involved Ann and Bob playing a game of red light, green light. These problems required solving either for Bob’s position relative to the finish line (Items 19–22) or for the difference between Ann’s and Bob’s positions (Items 23–24). Within Items 19–22, the referent position differed: in Items 19–20, Ann’s position served as the referent, whereas in Items 21–22, Bob’s position served as the referent.

2.3.3.3 Administration Procedure

Participants were randomly assigned to one of the four forms of word problems (Forms A–D). Within each form, half of the word problems (i.e., 12 items) were administered in the first testing session, and the remaining half (i.e., the other 12 items) in the second. The administration order of the 24 problems was randomized subject to predefined sequencing constraints in two ways. First, all six problems belonging to the same scenario (i.e., in the same quadrant in Figure 2-9, and belonging to the same word problem type) were administered consecutively; however, the order of these six problems was randomized. Second, the order in which the four scenarios were administered was also randomized.

Within each scenario, the order of the six problems was randomized subject to predefined sequencing constraints to minimize potential order effects, such as practice, fatigue, and carryover effects (Punch & Oancea, 2014). The easier problem pairs in each quadrant were always administered first. These included Items 1–2 and 13–14 from the left two quadrants (Langhorst et al., 2012; Rosenthal & Resnick, 1974), and Items 7–8 and 19–20 from the right two quadrants (Hegarty et al., 1995; Mayer & Hegarty, 1996).

After the easier pair, the remaining two problem pairs within the same scenario were presented in randomized order. Within each pair, the internal order of the two items was also randomized (i.e., Items 3–4, 5–6, 9–10, 11–12, 15–16, 17–18, 21–22, and 23–24). For example, within Scenario 1, the easier pair (Items 1 and 2) was administered first, with its internal order randomized (i.e., either 1–2 or 2–1). This was followed by either the 3–4 pair or Items 5–6 pair. Within the selected pair, the item order was randomized (e.g., 5–6 or 6–5). Finally, the

remaining pair was administered, again with its internal order randomized (e.g., 3–4 or 4–3). As shown in Figure 2-10, this structured randomization process yielded 16 possible problem-order sequences within each scenario ($2 \times 2 \times 2 \times 2 = 16$).

Figure 2-10

Randomization of Problem Order within each Scenario

Number of Options:	2	×	2	×	2	×	2
Decisions:	Order of problems 1–2 (1–2 or 2–1)		Which problem pair goes next (3–4 or 5–6)		Order within selected pair (e.g., 5–6 or 6–5)		Order within final pair (e.g., 3–4 or 4–3)

Second, the order of the four scenarios was systematically randomized. In any given testing session, Scenarios 1 and 2 were never administered together, nor were Scenarios 3 and 4. This constraint ensured variation in quantity type (i.e., discrete versus continuous) within each session, thereby reducing potential monotony due to repetition. In the first testing session, the first scenario was either a discrete quantity or a continuous quantity scenario (i.e., four possible options). The second scenario was selected from those involving the alternative quantity type (i.e., two options). In the second testing session, the remaining two scenarios were administered in either order (i.e., two options). As shown in Figure 2-11, this resulted in 16 possible scenario-order combinations ($4 \times 2 \times 2 \times 1 = 16$) across the two testing sessions. Taken together, the within-scenario problem-order and scenario-order randomization procedures substantially reduced the likelihood that systematic order effects influenced the findings.

Figure 2-11

Randomization of the Scenario Order

Number of Options:	4	×	2	×	2	×	1
Options:	Either a Discrete Quantity or a Continuous Quantity Scenario		One of the scenarios involving the quantity type not previously administered		One of the two remaining scenarios		The one remaining scenario
Testing Session:	First Testing Session				Second Testing Session		

The interview-style procedure for administering word problems was adapted from Hegarty and Kozhevnikov (1999) and Boonen et al. (2013, 2014). Before presenting the first problem from each scenario (i.e., word problem type), the scenario context was introduced by reading the script shown in the blue cells in Figure 2-9. This was intended to provide situational context (Stern & Lehrndorfer, 1992). Pupils' comprehension of the scenario was then checked using clarifying questions (e.g., 'What does it mean to walk towards something?').

Problems were presented on A4-size paper, with four problems arranged in each quadrant of the page, as shown in Figure 2-12. Each problem was read aloud to minimize the influence of reading fluency. Problems were read slowly and repeated if necessary to reduce memory load (Riley et al., 1983). Children were provided with manipulatives (i.e., 30 MathLink cubes, see Figure 2-13), and a pencil to support their work (Riley et al., 1983). At the start of the task, children were informed that they could use the manipulatives and pencil and paper if they found them helpful. Whenever a child experienced difficulty with a particular problem for more than one minute (e.g., stating that they did not know the answer or remaining silent), they were reminded that they could use the available supports ('Think about how you would solve this problem. You can draw on the paper and use these counters, if you would like.').

Figure 2-12

The Layout of Four Word Problems on A4-size Paper

1) Ann has 6 marbles. Bob has 10 more marbles than Ann. How many marbles does Bob have?	2) Ann has 18 marbles. Bob has 5 marbles less than Ann. How many marbles does Bob have?
3) Ann has 19 marbles. She has 8 more marbles than Bob. How many marbles does Bob have?	4) Ann has 7 marbles. She has 8 marbles less than Bob. How many marbles does Bob have?

A-1

While each child was solving a problem, detailed observation notes of overt behavior were recorded (e.g., the pattern and order in which MathLink cubes were placed on the desk, the order in which fingers were used, AND subvocal or vocal counting), as no photographic or video recordings were made. After each problem was solved, children were asked to explain how they arrived at their answer: 'What were you thinking in your mind?' Children's retrospective reports of their reasoning were also noted verbatim.

Figure 2-13

MathLink Cubes



Note. MathLink cubes were selected because they can connect on all faces, which affords greater flexibility in representation compared to linear-only manipulatives (e.g., Unifix cubes) and non-connective materials (e.g., counters).

2.3.3.4 Coding of Problem-Solving Strategies

Problem-solving strategies for solving word problems were coded in three stages. The first stage of coding ([Subsection 2.3.3.4.1](#)) occurred during testing and considered children's representational tools and arithmetic approaches. The second stage of coding ([Subsection 2.3.3.4.2](#)) used field notes—containing initial codes, additional notes on observable behavior, and verbatim retrospective reports from pupils—to apply the final problem-solving strategy codes. The final stage of coding ([Subsection 2.3.3.4.3](#)) involved a second researcher independently coding 20% of the field notes to establish inter-rater reliability.

2.3.3.4.1 Coding of Representational Tools and Arithmetic Approaches

The first stage of coding occurred during testing. For each word problem item, children's use of representational tools and arithmetic approaches was coded based on observable behavior and retrospective reports. Combining observations with retrospective reporting was considered appropriate, as prior research indicates that this approach validly captures children's strategies (Carr et al., 2008; Cho et al., 2011).

As summarized in Table 2-9, *representational tools* refer to the implements used to enact the problem situation, track counting steps, or represent quantities and quantitative relations. Initially, a separate code for *number line* was included. However, only four instances were observed. Upon review, one instance functioned more closely as *marks/drawings on paper*, whereas the remaining three functioned as *number sentences*. Accordingly, the number-

line code was removed, and the four instances were reassigned to the most appropriate categories.

Table 2-9

Representational Tools

Representational Tools	Definition
None	Not using any externally visible representational tool to model the problem situation or to keep track of the number of counting steps
Fingers	Using their own fingers to either directly model the problem situation or to keep track of the number of counting steps
Counters	Using the MathLink cubes provided to either directly model the problem situation or to keep track of the number of counting steps
Figural Units (e.g., nods, eye-tracking, gesturing)	Using figurative representations (e.g., nodding or pointing in the air, eye-tracking, or gesturing) to either directly model the problem situation or to keep track of the number of counting steps
Marks/Drawings on Paper	Making marks or drawing using the pencil and paper provided to either directly model the problem situation or to keep track of the number of counting steps
Number Sentence	Constructing a number sentence to represent the quantitative structure in the problem

As summarized in Table 2-10, *arithmetic approaches* capture the methods used to carry out the calculation, with or without explicit awareness of the associated operation. These codes were refined iteratively. First, the *count-on* strategy was initially divided into *count-on-from-the-smaller-value* and *count-on-from-the-larger-value*. However, these were combined because it was not always possible to distinguish between them, particularly when a child inappropriately applied a count-on strategy (i.e., using a count-on strategy for a problem that cannot be solved in this way).

Second, *derived number-fact* and *recall number-fact* were initially separate codes. During coding, it became evident that these approaches were often indistinguishable, especially when computation occurred internally without visible representation. The two were therefore merged into a single *Number Fact* code. For instance, if a child responded immediately and reported that they ‘just knew’ the answer, it was easier to categorize the arithmetic approach as a recalled number fact. Conversely, when a response was delayed, it was difficult to discern whether the latency was due to the use of a derived number-fact strategy or some other reason (e.g., difficulty comprehending the problem). This was further exacerbated by the fact that pupils’ retrospective reports did not always corroborate the observations. Since response times

were not formally recorded, latency could not be used to definitively distinguish between recall and derived strategies.

Table 2-10*Arithmetic Approaches*

Arithmetic Approaches	Definition
Count-All	Counting from one to determine the total of both numbers
Count-On	Counting on from one of the numbers and continuing for as many steps as indicated by the other number
Count-Up	Counting up from the smaller number until reaching the larger number
Separate-From	Counting the remaining units after taking away units corresponding to the smaller number from the larger number
Count-Down	Counting backward from the larger number, producing as many counting number words as indicated by the smaller number
Comparing/Matching	Pairing objects from two groups in a one-to-one manner to determine the difference between the two numbers
Iterations	Using fingers, counters, eye-tracking or nods in a repetitive manner to determine the distance between two numbers
Number Fact	Solving the problem by either recalling a known arithmetic fact from memory or deriving an unknown fact from a known one
Other/Unsystematic	Using an arithmetic approach that is undiscernible, or appears to involve guessing rather than a systematic strategy

Third, a new code—*Iterations* (using fingers, counters, gestures, or nodding)—was introduced after observing that some children used repetitive movements to solve continuous-quantity problems. Therefore, this code was incorporated into the final arithmetic approaches codes.

Finally, *random guesses*, *values from the problem*, and *other/undiscernible* arithmetic approaches were collapsed into a single *Other/Unsystematic* category. Reflecting the well-documented challenges of accessing young children's mental processes (Siegler & Jenkins, 1989), in many instances, a child's precise approach could not be confidently determined, even with retrospective reporting. For example, responses that appeared to be random guesses were not necessarily random, as they sometimes reflected underlying reasoning that the child was unable to fully articulate; in some cases, children repeated a value from the problem because they genuinely believed the answer was identical to that value, due to incorrect reasoning.

Conversely, elaborate explanations did not always accurately reflect the child's actual reasoning process. Collapsing these subcategories served to reduce interpretive ambiguity.

Two additional coding principles were applied. First, when discrepancies arose between verbal reports and observable behavior (e.g., a child used fingers to count but reported 'I just knew'), observable behavior was prioritized, consistent with prior research (Casey et al., 2017; Laski et al., 2013). Second, multiple codes were permitted if each contributed meaningfully to the solution. However, if an initial strategy was abandoned and did not contribute to the final answer, only the strategy that led to the solution was retained.

2.3.3.4.2 Derivation of Problem-Solving Strategy Codes

In the second stage, field notes—including representational tool codes, arithmetic approach codes, additional observations of behavior, and retrospective reports—were used to assign a single problem-solving strategy per item per child. The five mutually exclusive problem-solving strategies are summarized in Table 2-11. Importantly, problem-solving strategies were coded independently of solution accuracy. Thus, even if the final numerical answer was incorrect, the strategy was coded as 'Quantitatively Correct Counting' or 'Quantitatively Correct Number Fact' if the quantitative structure was represented accurately.

Table 2-11

Problem-Solving Strategies

Strategy	Definition
Non-Quantitative	Does not appear to actively engage with the problem's quantitative structure (e.g., guessing; repeating a number from the problem statement)
Quantitatively Incorrect Counting	Represents the problem's quantitative structure incorrectly while using a counting-based strategy (e.g., counting-all for a subtraction problem)
Quantitatively Incorrect Number Fact	Represents the problem's quantitative structure incorrectly while using a number-fact strategy (e.g., mentally subtracting for an addition problem)
Quantitatively Correct Counting	Represents the problem's quantitative structure correctly while using a counting-based strategy (e.g., counting-down for a subtraction problem)
Quantitatively Correct Number Fact	Represents the problem's quantitative structure correctly while using a number-fact strategy (e.g., mentally adding for an addition problem)

Note. Reproduced from Table 1-3 (Chapter 1).

When multiple arithmetic approaches were observed, the final problem-solving strategy code was determined by considering relative sophistication. Consider the following example: ‘Ann had 14 marbles. Then she played a game of marbles and lost 6 marbles. How many marbles does she have now?’ A child can solve this by first recalling that six plus six is 12, then counting-on using fingers to reach 14, to finally determine that the answer is eight. In this case, the final problem-solving code was deemed to be ‘Quantitatively Correct Number Fact’ rather than ‘Quantitatively Correct Counting’ because decomposing the problem was the critical step that enabled the child to arrive at the answer (Laski et al., 2013).

2.3.3.4.3 Inter-Rater Reliability

In the third stage, inter-rater reliability was established. A second rater independently coded 20.4% of the dataset (28 participants $\times$ 24 items) for problem-solving strategies, based on the researcher’s field notes of observable behavior and retrospective reports. The 28 participants were selected at random from the total sample ($N = 137$). The two raters showed 97.2% agreement, and Cohen’s κ was .96, 95% CI [.95, .98], $p < .001$, indicating excellent reliability (Siegel & Castellan, 1988).

2.3.3.5 Reliability and Validity of Word Problems

Following data collection, various statistical analyses were conducted to evaluate the reliability and validity of the word problems and the underlying 2×2 conceptual framework. All analyses were performed using R version 4.4.0 (R Core Team, 2024) and RStudio version 2023.12.1.402 (Posit Team, 2024). First, the internal consistency of each form was examined independently using the Kuder-Richardson 20 (KR-20) analysis ([Subsection 2.3.3.5.1](#)). Second, the equivalence of the four forms—designed to be parallel through systematic randomization of numerical values—was assessed by comparing average item difficulty and mean item-total point-biserial correlations across forms using one-way analyses of variance (ANOVAs) ([Subsection 2.3.3.5.2](#)). Third, once the compatibility of these forms was established, they were treated as equivalent and aggregated. The reliability of this combined dataset was then examined using tetrachoric correlation, KR-20 analysis, and overall item difficulty and discrimination indices ([Subsection 2.3.3.5.3](#)). Finally, the structural validity of the 2×2 conceptual framework was evaluated through confirmatory factor analyses (CFAs) ([Subsection 2.3.3.5.4](#)). Each assessment step is detailed below.

2.3.3.5.1 Internal Consistency of each Test Form

Using the *psych* package in R (Revelle, 2026), the reliability of each form was estimated separately through the Kuder-Richardson 20 (KR-20) coefficient, which is the preferred estimator for dichotomous items (McGartland Rubio, 2005). The four forms yielded KR-20 values of .92 (95% CI [0.87, 0.95]), .94 (95% CI [0.91, 0.96]), .94 (95% CI [0.91, 0.96]), and .94 (95% CI [0.91, 0.96]), respectively, indicating strong internal consistency within each test form (Field, 2018).

2.3.3.5.2 Compatibility of the Four Test Forms

The compatibility of the four alternative word problem test forms was examined by comparing item difficulty and item discrimination indices. The item difficulty index represents the proportion of test-takers who answered an item correctly, ranging from 0 (indicating that no one answered the item correctly, i.e., very difficult) to 1 (indicating that everyone answered the item correctly, i.e., very easy), with a value of .50 indicating that half the test-takers answered the item correctly. Although the literature varies regarding threshold values, it is generally accepted that values between 0.1 and 0.3 indicate difficult items, values between 0.7 and 0.9 indicate easy items, and values between 0.3 and 0.7 indicate moderate difficulty (El-Hamamsy et al., 2022). The item discrimination index, on the other hand, represents how effectively an item differentiates between higher- and lower-performing participants. An item discrimination index $\geq .30$ is considered reasonably good, and $\geq .40$ is considered very good (Ebel & Frisbie, 1991).

The average item difficulty indices of the four forms were .50 ($SD = 0.18$), .49 ($SD = 0.16$), .47 ($SD = 0.18$), and .46 ($SD = 0.19$), respectively, all falling within the moderate level of difficulty range. The average item discrimination index for each form exceeded the .40 threshold: .49 ($SD = 0.12$), .50 ($SD = 0.11$), .55 ($SD = 0.15$), and .55 ($SD = 0.14$), respectively. [Appendix 2I](#) presents detailed item-level difficulty and discrimination indices across the four forms, including: (1) item-level indices for each form, (2) average indices per item across forms, (3) average indices per form, and (4) overall averages across all items and forms.

One-way ANOVAs were conducted to assess whether there were statistically significant differences in mean item difficulty and mean item discrimination (point-biserial correlation) across the four forms. Prior to interpreting the results, assumptions of normality and homogeneity of variances were evaluated. For item difficulty, the Shapiro-Wilk test indicated that residuals were normally distributed ($W = .99$, $p = .36$), supported by visual

inspection of the histogram and Q-Q plot. Levene's test indicated homogeneity of variances ($F(3, 92) = 0.07, p = .98$). The one-way ANOVA revealed no significant differences in mean item difficulty across test forms ($F(3,92) = 0.26, p = .86$). For item discrimination, the Shapiro-Wilk test also supported normality ($W = .98, p = .08$), and Levene's test confirmed homogeneity of variances ($F(3, 92) = 0.73, p = .54$). The one-way ANOVA indicated no significant differences in item discrimination across test forms ($F(3, 92) = 1.57, p = .20$).

To rule out the possibility that pre-existing group differences could confound comparisons across forms, the four groups were compared on demographic characteristics and performance on all measures. Table 2-12 summarizes group means. A chi-square test for homogeneity of proportions was conducted for categorical variables (i.e., gender), while one-way ANOVAs were conducted for all continuous variables.

Table 2-12

Group Means for Demographic Characteristics and Performance on Measures

	Form A <i>M (SD)</i>	Form B <i>M (SD)</i>	Form C <i>M (SD)</i>	Form D <i>M (SD)</i>	<i>F-Test</i> <i>(p-value)</i>
Age	7.10 (0.30)	7.06 (0.32)	7.18 (0.28)	7.08 (0.31)	0.90 (.44)
Gender	0.52 (0.51)	0.35 (0.49)	0.62 (0.49)	0.50 (0.51)	-
WASI-II Block Design	12.71 (6.28)	10.15 (7.22)	12.68 (7.65)	9.97 (4.25)	1.83 (.15)
Visual-Spatial Reasoning	11.10 (3.19)	10.09 (3.38)	10.12 (3.72)	9.59 (3.57)	1.06 (.37)
PathSpan	4.74 (0.97)	4.09 (1.31)	4.47 (1.35)	4.15 (1.08)	2.11 (.10)
WIAT-II ^{UK} Listening Comprehension	18.19 (3.24)	17.21 (4.22)	17.59 (4.02)	17.65 (3.88)	0.36 (.79)
WIAT-II ^{UK} Numerical Operations	13.23 (3.22)	12.47 (2.85)	12.97 (3.90)	11.85 (2.80)	1.17 (.33)
WASI-II Matrix Reasoning	10.26 (3.75)	8.74 (4.56)	8.32 (4.19)	8.59 (3.85)	1.43 (.24)
Word Problem Accuracy	10.19 (6.68)	8.59 (7.00)	8.94 (7.19)	9.21 (7.17)	0.31 (.82)

Note. $N_{\text{Form A}} = 34$, $N_{\text{Form B}} = 35$, $N_{\text{Form C}} = 35$, $N_{\text{Form D}} = 35$. For gender, 0 = girls, 1 = boys, such that the mean reflects the proportion of boys in each group. *M* = mean; *SD* = standard deviation. Group differences in gender distribution were assessed using a chi-square test for homogeneity of proportions rather than a one-way ANOVA; $\chi^2(3) = 4.85, p = .18$. All other group comparisons were conducted using one-way ANOVAs.

Prior to conducting the one-way ANOVAs, assumptions of normality and homogeneity of variances were evaluated. The Shapiro-Wilk test indicated normality for age, the Visual-Spatial Reasoning task, and the WIAT-II^{UK} Listening Comprehension subtest across all groups.

For other measures, one or more groups showed departures from normality; however, visual inspection of histograms and Q-Q plots indicated that deviations were not substantial. Levene's test supported homogeneity of variances for all measures.

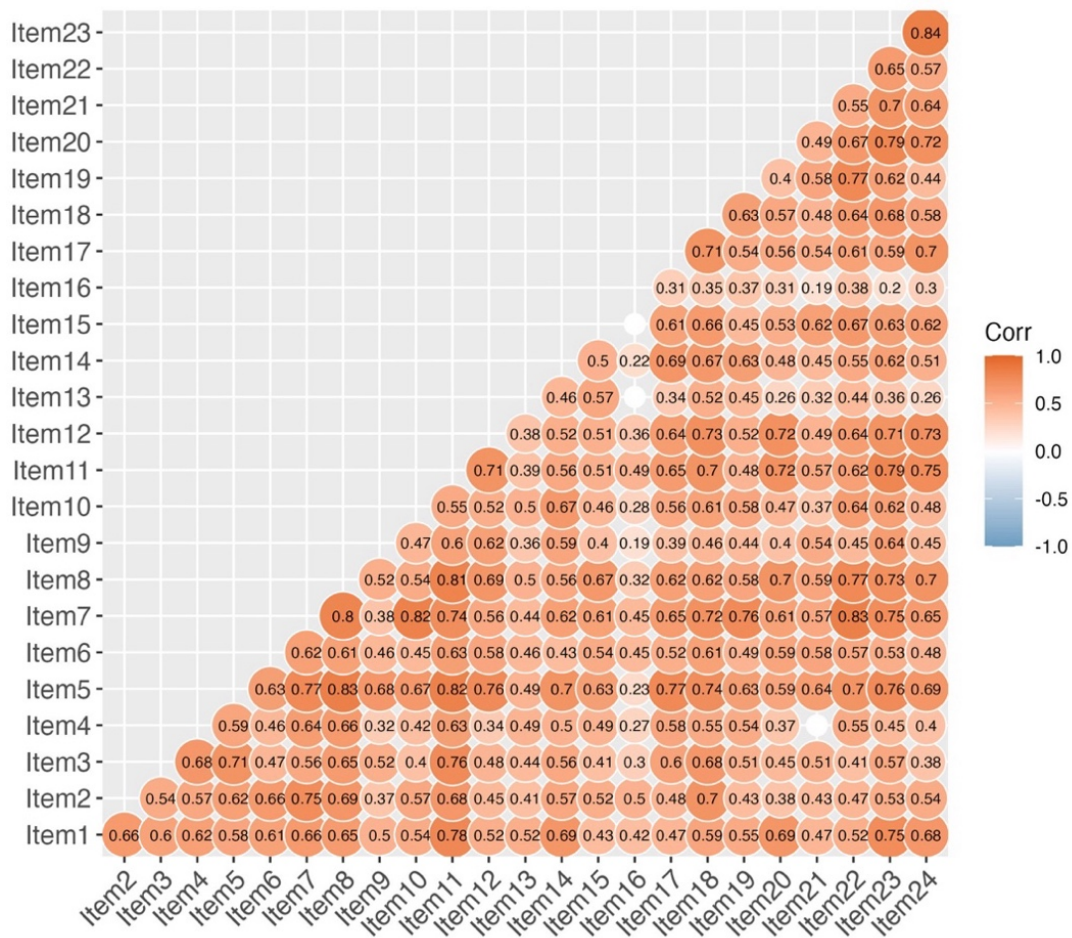
The four groups were compared in terms of gender and age. A one-way ANOVA indicated no significant age differences among the groups ($F(3, 133) = 0.90, p = .44$). Likewise, a chi-square test for homogeneity of proportions indicated no significant differences in gender distribution across forms ($\chi^2(3) = 4.85, p = .18$).

One-way ANOVAs revealed no significant differences among the four groups on the WASI-II Block Design subtest ($F(3, 133) = 1.83, p = .15$), the Visual-Spatial Reasoning task ($F(3, 133) = 1.06, p = .37$), PathSpan ($F(3, 133) = 2.11, p = .10$), the WIAT-II^{UK} Listening Comprehension subtest ($F(3, 133) = 0.36, p = .79$), the WIAT-II^{UK} Numerical Operations subtest ($F(3, 133) = 1.17, p = .33$), the WASI-II Matrix Reasoning subtest ($F(3, 133) = 1.43, p = .24$), and word problems accuracy ($F(3, 133) = 0.31, p = .82$).

Taken together, the similarly high KR-20 reliability coefficients across forms, the absence of statistically significant differences in item difficulty and discrimination indices, and the comparability of the four groups in demographic composition and performance on all measures indicate that the four forms may be treated as equivalent. Furthermore, given that the forms were intentionally designed to be parallel, and considering the small sample sizes and minor departures from the normality assumption, statistical equating methods were deemed neither necessary nor appropriate. Therefore, four forms were treated as equivalent in all subsequent analyses.

2.3.3.5.3 Reliability of the Combined Word Problems Measure

Having established the compatibility of the four alternative test forms, a series of analyses was conducted to assess the overall reliability of the word problems, treating the four forms as equivalent. First, using the *psych* package in R (Revelle, 2026), tetrachoric correlations were computed to examine the interrelationships among the 24 dichotomous items (i.e., correct or incorrect). Tetrachoric correlations were used because standard Pearson correlations underestimate associations between binary variables and because this method assumes an underlying continuous latent ability (Ekström, 2011; Watkins, 2018). As shown in Figure 2-14, most item pairs exhibited statistically significant tetrachoric correlations, ranging from $r_{tet} = .19$ to $.84, p < .05$, indicating weak to very strong associations and no evidence of extreme multicollinearity.

Figure 2-14*Tetrachoric Correlation Coefficients*

Note. Correlations that are not statistically significant are left blank.

Next, the internal consistency of the 24 items—treating the four test forms as equivalent—was re-examined. The KR-20 coefficient was 0.93 (95% CI [0.91, 0.95]), indicating strong internal consistency (Cohen et al., 2018) and suggesting that the items reliably and consistently measure the underlying construct of word-problem-solving ability (Field, 2018). Based on visual inspection of Figure 2-14, the KR-20 coefficient was recalculated after excluding Item 16; however, because this did not yield a meaningful improvement, all items were retained.

The overall item difficulty, averaged across all items and forms, was .48 ($SD = 0.17$), indicating a moderate level of difficulty. The overall item discrimination (point-biserial correlation) was .52 ($SD = 0.13$), suggesting that, on average, the items demonstrated good differentiation between higher- and lower-performing participants.

2.3.3.5.4 Structural Validity of Conceptual Framework

Confirmatory factor analyses (CFAs) were conducted using the *lavaan* package (Rosseel, 2012) to examine the structural validity of the 24-word problems and their categorization according to the proposed 2×2 conceptual framework. The Weighted Least Squares Mean and Variance adjusted (WLSMV) estimator was employed, as there is growing consensus that it is an appropriate method for analyzing dichotomous indicators (Newsom, 2023). Model fit was evaluated using the chi-square statistic ($\chi^2/\text{df} < 2$), the Comparative Fit Index (CFI > 0.95), the Tucker–Lewis Index (TLI > 0.95), the root mean squared error of approximation (RMSEA < 0.05), and the Standardized Root Mean Square Residual (SRMR < 0.08) (Kline, 2016).

Four competing models (M1–M4) were tested: a one-factor model (M1), two alternative two-factor models (M2–M3), and the hypothesized four-factor model (M4). M2 and M3 were two-factor models, each representing one of the two conceptual dimensions: (1) Quantity (yielding discrete-quantity and continuous-quantity factors), and (2) Quantitative Relation (yielding dynamic-relation and static-relation factors), respectively. The final model (M4) was a four-factor model reflecting the interaction of both dimensions and corresponding to the proposed 2×2 conceptual framework (Quantity $\times$ Quantitative Relation).

Table 2-13 summarizes the model fit indices for all models, and [Appendix 2J](#) presents the standardized factor loadings for each model. The first model (M1) was a one-factor model that included all 24 problems. M1 was examined to assess whether the 24 problems could reasonably be represented as a single latent construct and to complement the KR-20 analysis. As shown in Table 2-13, M1 demonstrated good overall fit ($\chi^2/\text{df} = 1.17$, CFI = 0.989, TLI = 0.988, RMSEA = 0.036, SRMR = 0.094), suggesting that the 24 problems largely reflect a common underlying construct—one-step additive word-problem-solving ability. These findings support the interpretation of the KR-20 coefficient as reflecting a largely unidimensional scale.

Although the χ^2 value was statistically significant ($p < .05$), the model was deemed to demonstrate acceptable fit because the χ^2/df ratio was below the commonly accepted threshold of 2. Similarly, although the SRMR value exceeded the conventional cutoff of .08, it remained below .10, which is generally considered indicative of poor fit (Kline, 2016). All standardized factor loadings were statistically significant ($p < .001$), ranging from .40 to .92 (see [Table 2J-1 in Appendix 2J](#)).

Table 2-13*Model Fit Indices of Confirmatory Factor Analyses*

Model	Factors	χ^2	df	χ^2/df	CFI	TLI	RMSEA	SRMR
One-factor (M1)		295.45*	252	1.17	0.989	0.988	0.036	0.094
Two-factor (M2): Quantity	Discrete Quantity, Continuous Quantity	290.00*	251	1.16	0.990	0.989	0.034	0.093
Two-factor (M3): Quantitative Relation	Dynamic Relation, Static Relation	285.45	251	1.14	0.991	0.990	0.032	0.092
Four-factor (M4): Quantity × Quantitative Relation	Discrete Quantity–Dynamic Relation, Discrete Quantity–Static Relation, Continuous Quantity–Dynamic Relation, Continuous Quantity–Static Relation	272.19	246	1.11	0.993	0.992	0.028	0.090

Note. $N = 137$. CFI = comparative fit index; TLI = Tucker–Lewis index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual. Acceptable model fit: $\chi^2/\text{df} < 2$, CFI > 0.95 , TLI > 0.95 , RMSEA < 0.05 , SRMR < 0.08 (Kline, 2016). * $p < .05$.

Models M2–M4 were evaluated with respect to two criteria. First, structural validity of the conceptual framework—namely, the differentiation of word problem types—was assessed using model fit indices. Second, the strength of item–factor relationships within each factor was evaluated by examining standardized factor loadings.

As summarized in Table 2-13, both two-factor models (M2–M3) demonstrated good overall fit, supporting the structural validity of the conceptual framework along each individual dimension. The fit indices for M2 ($\chi^2/\text{df} = 1.16$, CFI = 0.990, TLI = 0.989, RMSEA = 0.034, SRMR = 0.093) support the differentiation of the 24 word problems into discrete-quantity and continuous-quantity categories. Similarly, the fit indices for M3 ($\chi^2/\text{df} = 1.14$, CFI = 0.991, TLI = 0.990, RMSEA = 0.032, SRMR = 0.092) support the categorization of word problems into dynamic-relation and static-relation categories.

All standardized factor loadings in the two-factor models (M2–M3) were statistically significant ($p < .001$). Furthermore, the magnitude of the loadings ranged from moderate to strong, suggesting that the items generally represented their respective factors well (Hair et al., 2010). In M2, standardized factor loadings ranged from .65 to .93 for discrete-quantity and from .41 to .91 for continuous-quantity (see [Tables 2J-2 in Appendix 2J](#)). In M3, standardized factor loadings ranged from .41 to .95 for dynamic-relation and from .65 to .94 for static-relation categories (see [Table 2J-3 in Appendix 2J](#)).

Among the four competing models, the four-factor model (M4) provided the best fit to the data. As shown in Table 2-13, M4 demonstrated the lowest χ^2/df ratio (1.11), highest CFI (0.993) and TLI (0.992) values, and lowest RMSEA (0.028) and SRMR (0.090) values. These fit indices indicate good overall model fit and support the structural validity of the proposed 2×2 conceptual framework (Quantity $\times$ Quantitative Relation), in which problems are categorized along both dimensions simultaneously.

All standardized factor loadings were statistically significant ($p < .001$), and ranged from .72 to .95 for Discrete Quantity–Dynamic Relation, .65 to .93 for Discrete Quantity–Static Relation, .42 to .87 for Continuous Quantity–Dynamic Relation, and .73 to .93 for Continuous Quantity–Static Relation (see [Tables 2J-4 in Appendix 2J](#)).

Several caveats should be considered when interpreting the CFA findings. First, some fit indices did not fully meet conventional thresholds. Similar to M1, the χ^2 statistic for M2 was statistically significant ($p < .05$). In addition, SRMR values for all models slightly exceeded the .08 criterion. However, given that χ^2/df ratios were below 2, SRMR values

remained below .10 (Kline, 2016), and the remaining fit indices (CFI, TLI, and RMSEA) indicated good fit, the models were deemed to demonstrate acceptable structural validity overall. Moreover, although most standardized factor loadings were in the moderate to very strong range, Item 16 exhibited the lowest loading (.42). Given that this loading exceeded the recommended threshold of .40 (Hair et al., 2010) and that the item was theoretically grounded in the 2×2 conceptual framework, it was deemed to have satisfactory relationship with its respective factor and was retained in the model.

Second, the sample size may limit the stability of parameter estimates, as some guidelines recommend a minimum of 200 participants for CFA (Kline, 2016). *A priori* power analyses indicated that 90 participants were required for the two-factor models and 137 for the four-factor model, assuming an anticipated medium effect size of 0.3 (Cohen, 1988), $\alpha = .05$, and power of .80 (Soper, 2025). Therefore, the sample size of 137 was considered adequate for evaluating the four-factor model based on the *a priori* power analysis.

As a final step in evaluating the word problems and the conceptual framework used to categorize them, the correlations among the factors were examined. As anticipated based on the conceptual framework, across Models M2 to M4, factors were highly correlated ($\geq .73$), suggesting the presence of a higher-order latent construct representing one-step additive word-problem-solving ability (Anderson & Gerbing, 1988). Consistent with this interpretation, a second-order CFA of the four-factor model yielded statistically significant higher-order factor loadings ($p < .001$) (see Table 2-14) and acceptable model fit ($\chi^2/df = 1.13$, CFI = 0.992, TLI = 0.991, RMSEA = 0.031, SRMR = 0.091). Second-order analyses were not conducted for the two-factor models because at least three first-order factors are required for identification of a higher-order model (Kline, 2016).

Table 2-14

Factor Loadings of Second-Order Confirmatory Factor Analyses

Factors	β	SE	z	p -value
Discrete Quantity–Dynamic Relation	0.99 [0.91, 1.00]	0.02	40.34	< .001
Discrete Quantity–Static Relation	1.01 [0.98, 1.04]	0.02	60.42	< .001
Continuous Quantity–Dynamic Relation	0.95 [0.88, 1.01]	0.03	27.54	< .001
Continuous Quantity–Static Relation	0.92 [0.86, 0.98]	0.03	30.85	< .001

Note. β = standardized factor loading; SE = standard error.

Overall, the CFA results provided support for the structural validity of the proposed 2×2 conceptual framework (Quantity $\times$ Quantitative Relation) for categorizing the 24 one-step additive word problems into four distinct problem types: Discrete Quantity–Dynamic Relation, Discrete Quantity–Static Relation, Continuous Quantity–Dynamic Relation, and Continuous Quantity–Static Relation. The standardized factor loadings indicated that the items adequately represented their respective problem type factors within the proposed framework. Accordingly, this framework was retained for subsequent analyses addressing RQ2, RQ4, and RQ6 (Studies 1–3).

2.4 Summary of Analytic Approach

As introduced in [Section 1.5](#), this thesis comprises three research themes, each corresponding to Chapters 3 to 5 (Studies 1–3). The three themes are: (1) the relationship between spatial skills and one-step additive word-problem-solving accuracy, (2) the relationship between spatial skills and problem-solving strategies, and (3) the relationship between spatial skills and one-step additive word-problem-solving accuracy, with problem-solving strategies as a mediator.

As summarized in Table 2-15 (Study 1), to investigate whether spatial visualization and visuospatial working memory predict problem-solving accuracy on one-step additive word problems, a cross-classified mixed-effects logistic regression analysis was conducted (RQ1). Problem-solving accuracy for each word-problem item was specified as the binary dependent variable. Item-level logistic regression models were used because they directly model the binary nature of individual responses and avoid assumptions inherent in score aggregation (Boonen et al., 2014). Random intercepts for participants and items were included to account for the cross-classified structure of the data, as each pupil solved all 24 word problems and each problem was completed by all 137 participants (Baayen et al., 2008; Barr et al., 2013).

Table 2-15*Summary of Study 1 Analyses*

Research Theme	Research Questions	Variables and Analyses
Chapter 3: Spatial Skills and One-Step Additive Word- Problem- Solving Accuracy	RQ1) To what extent do spatial visualization and visuospatial working memory predict young primary school pupils' (6–7-year-olds') accuracy on one-step additive word problems?	Dependent Variable: • Word-problem-solving accuracy (binary): • Correct • Incorrect Independent Variables: • Spatial visualization • Visuospatial working memory Covariates: • Language ability • Numerical ability • Fluid reasoning • Age • Gender Analyses: • Cross-Classified mixed-effects logistic regression (RQ1) • Cross-Classified mixed-effects logistic regression with interaction terms (RQ2)
	RQ2) Do the relationships between spatial visualization, visuospatial working memory, and problem-solving accuracy vary across different word problem types?	

Subsequently, a cross-classified mixed-effects logistic regression model with interaction terms (spatial visualization $\times$ word problem type and visuospatial working memory $\times$ word problem type) was estimated to assess whether the relationship between spatial skills and problem-solving accuracy varied across word problem types (RQ2).

As summarized in Table 2-16 (Study 2), to investigate whether spatial visualization and visuospatial working memory predict problem-solving strategy use, a cross-classified mixed-effects multinomial logistic regression model was employed (RQ3). Problem-solving strategy for each word-problem item was specified as the multinomial dependent variable. Random intercepts for participants and items were included to account for the cross-classified data structure. The '*Quantitatively Correct Number Fact*' strategy was specified as the baseline category, based on the *a priori* hypothesis that spatial skills would most strongly predict the likelihood of using this strategy.

Subsequently, a cross-classified mixed-effects multinomial logistic regression model with interaction terms (spatial visualization $\times$ word problem type and visuospatial working memory $\times$ word problem type) was estimated to assess whether the relationship between spatial skills and problem-solving strategy varied across word problem types (RQ4).

Table 2-16*Summary of Study 2 Analyses*

Research Theme	Research Questions	Variables and Analyses
Chapter 4: Spatial Skills and Problem-Solving Strategies	RQ3) To what extent do spatial visualization and visuospatial working memory predict young primary school pupils' (6–7-year-olds') problem-solving strategy use?	Dependent Variable: • Problem-solving strategies (categorical): • Non-Quantitative • Quantitatively Incorrect Counting • Quantitatively Incorrect Number Fact • Quantitatively Correct Counting • Quantitatively Correct Number Fact
	RQ4) Do the relationships between spatial visualization, visuospatial working memory, and problem-solving strategy use vary across different word problem types?	Independent Variables: • Spatial visualization • Visuospatial working memory Covariates: • Language ability • Numerical ability • Fluid reasoning • Age • Gender Analyses: • Cross-classified mixed-effects multinomial logistic regression (RQ3) • Cross-classified mixed-effects multinomial logistic regression with interaction terms (RQ4)

As summarized in Table 2-17 (Study 3), to examine whether problem-solving strategies—specifically the ‘Quantitatively Correct Number Fact’ strategy—mediate the relationship between spatial skills and word-problem-solving accuracy, path analysis was conducted (RQ5). Overall word-problem-solving accuracy (i.e., out of 24 items) was specified as the dependent variable, spatial visualization and visuospatial working memory as the independent variables, and the proportion of ‘Quantitatively Correct Number Fact’ strategy use as the mediator.

To examine whether the mediation effect varied across word problem types (RQ6), subsequent analyses were conducted separately for each word problem type. For these models, accuracy was calculated within each subset of word-problem items corresponding to a given problem type (i.e., out of 6 items). The mediator was defined as the proportion of items solved using the ‘Quantitatively Correct Number Fact’ strategy, computed within the same subset of items as the corresponding accuracy measure.

Table 2-17*Summary of Study 3 Analyses*

Research Theme	Research Questions	Variables and Analyses
Chapter 5: Spatial Skills and One-Step Additive Word- Problem-Solving Accuracy: The Mediating Role of Problem-Solving Strategies	RQ5) To what extent do spatial visualization and visuospatial working memory exert indirect effects on young primary school pupils' (6–7-year-olds') accuracy on one-step additive word problems through 'Quantitatively Correct Number Fact' strategy use? RQ6) Do the indirect effects of spatial visualization and visuospatial working memory on problem-solving accuracy through 'Quantitatively Correct Number Fact' strategy use vary across different word problem types?	Dependent Variable: • Word-problem-solving accuracy (Continuous) Independent Variables: • Spatial visualization • Visuospatial working memory Mediator: • Problem-solving strategies (specifically, proportion of '<i>Quantitatively Correct Number Fact</i>' strategy use) Covariates: • Language ability • Numerical ability • Fluid reasoning • Age • Gender Analyses: • Path-analyses (RQ5 & RQ6)

The level of analysis differed across studies in accordance with the conceptual focus of each research theme and its corresponding research questions. Studies 1 and 2 examined, at the trial level, the association between spatial skills and problem-solving accuracy and strategy use, respectively, requiring models that accounted for the cross-classified structure of the data (responses nested within both participants and problem items). Accordingly, cross-classified mixed-effects models were employed to partition variance attributable to participants and items and to test moderation by problem type (Baayen et al., 2008; Barr et al., 2013).

In contrast, Study 3 investigated whether children's overall tendency to use 'Quantitatively Correct Number Fact' strategies mediated the association between spatial skills and overall problem-solving accuracy (e.g., Foley et al., 2017; Laski et al., 2013). The mediation question was conceptualized at the between-person level, focusing on stable individual differences in strategy use and problem-solving accuracy. Accordingly, person-level aggregated indices were constructed and analyzed using path modeling (MacKinnon, 2008; Hayes, 2018). Although multilevel mediation models are possible (Bauer et al., 2006; Preacher et al., 2010), such approaches are typically used to examine mediation processes at the within-person or item level and would have shifted the focus away from the between-person associations that were the primary interest of the study. Moreover, such analysis require substantially larger numbers of participants and items for stable estimation—particularly when the mediator or outcome is categorical (McNeish & Stapleton, 2016; Raudenbush & Bryk, 2002; Snijders & Bosker, 2012). Given the present design (137 participants and 24 items) and the categorical nature of the mediator and outcome variables, mediation was modeled at the person level to provide reliable parameter estimates for the research question of interest.

2.5 Ethical Considerations

This research was approved by the Central University Research Ethics Committee (CUREC) ([Appendix 2K](#)). It follows the ethical guidelines for educational research recommended by the British Educational Research Association (British Educational Research Association, 2018) and the University of Oxford Policy on Research Involving Human Participants Data Protection (Central University Research Ethics Committee, 2022). As this study did not involve invasive research methods with children, it followed CUREC Approved Procedures AP25: Non-invasive Research Methods with Children (Central University Research Ethics Committee, n.d.). The only personal data collected were those essential for the purposes of the research: age on the day of testing and gender.

Given that this research involved young participants, several ethical considerations were taken into account. First, consent from appropriate gatekeepers, as well as the child's assent ([Appendix 2L](#)), was obtained prior to data collection. Before participation, the school headteacher, teachers, and parents/guardians were provided with relevant information sheets (see [Appendices 2C–2H](#)). Given that this thesis complied with CUREC Approved Procedure 25 (Non-invasive research involving participants aged 3 to 16 years, recruited via an organization), the opt-out recruitment method was approved by CUREC. For instance, the research topic was not expected to be contentious for parents or guardians, and the various tasks used in this thesis resembled typical classroom activities. In addition, the researcher is a certified primary school teacher with experience administering similar tasks to children of this age group. Furthermore, no personally identifiable data were collected, and no video or audio recordings or photographs were taken.

Although the opt-out recruitment method was approved by CUREC, schools were given the option to choose between opt-in and opt-out parental consent procedures. Schools could also select either an extended or simplified parental information sheet/consent. This approach was considered appropriate because schools are better positioned to understand their parent demographic, social context, and institutional policies.

Measures were taken to ensure that children felt comfortable working with an unfamiliar researcher. One-on-one testing took place in a quiet area near the child's classroom, where a familiar adult was nearby. At the beginning of each session, the researcher introduced herself and obtained the child's assent. Children who declined to participate after gentle encouragement (e.g., reassurance about safety and the nature of the activity) were allowed to continue with their classroom activities. Requests to postpone participation (e.g., completing a preferred activity first) were not considered withdrawal of assent; children were approached again at a more suitable time. Throughout testing, children were reminded that they could stop or end the session at any time without negative consequences. The researcher monitored children's comfort through verbal and non-verbal cues. Testing was discontinued if a child expressed a wish to stop, either verbally or through non-verbal indicators (e.g., uncooperative behavior).

Testing sessions were scheduled in consultation with teachers to minimize disruptions to classroom learning and avoid missing essential lessons. To reduce cognitive fatigue, tasks were administered across two sessions of approximately 30 minutes each, typically separated

by at least one hour (but no less than 30 minutes). The activities resembled those typically undertaken during the school day (e.g., puzzles, counting), and most children appeared to find the tasks engaging and appropriately challenging. Previous research indicates that children generally enjoy participating in structured one-on-one tasks with an experienced teacher-researcher (e.g., Hawes et al., 2019b).

All paper records (e.g., consent/opt-out forms, children's work, field notes, and scoring sheets) were scanned and password-protected. Consent-related forms were shredded immediately after scanning. Children's work, field notes, and scoring sheets are stored in a locked filing cabinet accessible only to the researcher. Digital files are encrypted and stored on a password-protected computer and on the University's secure storage system (Oxford's Nexus 365 OneDrive), accessible only to the researcher and supervisors.

Each child was assigned a unique code number to ensure that no personally identifiable information (e.g., name, age, sex) was linked to their data. A separate password-protected and encrypted file linking names to code numbers is stored securely on OneDrive. Following submission of the DPhil thesis, anonymized data will be retained on OneDrive for as long as one of the associated researchers (i.e., researcher and supervisors) remains affiliated with the University of Oxford and has authorized access. Paper records will be permanently destroyed after three years.

2.6 Summary of Methodology Chapter

Using a cross-sectional and correlational design with quantitative methods, this thesis investigates the relationship between spatial skills (spatial visualization and visuospatial working memory) and both problem-solving strategies and accuracy in one-step additive word problems among young primary school pupils (6–7-year-olds) in England. A non-probability convenience sampling method was used to recruit 149 pupils, of whom 137 were included in the final analyses. Participants completed a battery of seven tasks assessing spatial, mathematical, language, and fluid reasoning ability, through one-on-one testing sessions. Most instruments were either standardized tests or tasks that have been widely used in prior research. The one-step additive word problems were developed based on the proposed conceptual framework of this thesis, informed by the literature and refined through an iterative process. The reliability and validity of this novel measure were evaluated using a range of statistical analyses.

3 Study 1: The Role of Spatial Skills in Young Children's Arithmetic Word-Problem-Solving Accuracy

Abstract

Despite extensive evidence of a spatial–mathematical link (Hawes & Ansari, 2020; Lourenco et al., 2018; Mix & Cheng, 2012), the role of spatial ability in arithmetic word problem solving in the early primary years remains underexplored. The present study contributes to addressing this gap in two ways. First, it proposes a 2×2 conceptual framework for categorizing one-step additive word problems by integrating Thompson's (1993) distinction between quantity and quantitative relation, Davydov's (1982) discussion of methods of quantification, and Vergnaud's (1982) classification of dynamic and static relations. Second, using data from 137 Year 2 pupils (6–7-year-olds) in England and cross-classified mixed-effects logistic regression analyses, the study examines the roles of spatial visualization and visuospatial working memory in predicting word-problem-solving accuracy while controlling for language ability, numerical ability, fluid reasoning, age, and gender. Neither spatial skill uniquely predicted overall word-problem-solving accuracy. However, applying the conceptual framework showed that the association between spatial visualization, but not visuospatial working memory, and word-problem-solving accuracy varied across problem types, indicating that the strength of the association between spatial visualization and accuracy may differ according to the quantitative structure of the problems. These findings underscore the importance of distinguishing between spatial skills and accounting for quantitative structure when examining the role of spatial ability in early arithmetic word problem solving.

3.1 Introduction

Arithmetic word problem solving is a fundamental component of early mathematical competence (National Council of Teachers of Mathematics, 2000; OECD, 2023) and serves as a foundation for more advanced mathematical abilities. Recognizing their importance, the national curriculum in England emphasizes working with numerals, words, and the four operations as a key focus of mathematics teaching in Key Stage 1 (ages 5 to 7) (Department for Education, 2013). Moreover, performance on arithmetic word problems predicts later academic achievement (Duque de Blas et al., 2021) and future career success (Murnane et al., 2001).

Arithmetic word problems, also called verbal or story problems, are problems presented in verbal descriptions without number sentences (Wong & Ho, 2017), in which the problem solver must apply mathematical operations (i.e., addition, subtraction, multiplication, and division) to the numerical data provided in the problem statement (Verschaffel et al., 2000). Young children can solve arithmetic word problems using informal methods, such as pretend play and counting, even before receiving formal mathematics instruction (Nunes & Bryant, 1996; Verschaffel et al., 2020). However, solving word problems remains a cognitively demanding process that requires the simultaneous understanding and integration of language and numerical elements (Vessonen et al., 2024). As a result, children often find word problems more challenging than arithmetic computation tasks presented in symbolic or numerical form without context (Jitendra et al., 2007). Furthermore, word problems are generally considered difficult for all pupils, regardless of grade level (Vlahović-Štetić et al., 2004).

Spatial ability—a constellation of related skills concerning “the locations of objects, their shapes, their relations to each other, and the paths they take as they move” (Newcombe, 2017, p. 6)—is one understudied yet promising lens for understanding young children’s word-problem-solving capabilities. Extensive literature indicates that spatial ability plays a fundamental role in the learning and doing of mathematics (Atit et al., 2022; Hawes & Ansari, 2020; Lourenco et al., 2018; Mix & Cheng, 2012). In fact, the spatial–mathematical link is not only correlational (Atit et al., 2022; Xie et al., 2020) but also causational (Hawes & Ansari, 2020; Lourenco et al., 2018; Mix & Cheng, 2012). For instance, a recent meta-analysis of 29 controlled pre-post studies ($N = 3,765$) indicates that spatial training improves mathematics outcomes compared to the control condition, with an average effect size (Hedges’s g) of .28 ($SE = .07$) (Hawes et al., 2022). Furthermore, spatial ability is not only related to mathematical

domains that are inherently spatial, such as geometry (Battista, 1990; Delgado & Prieto, 2004; Goldsmith et al., 2016; Kyttälä & Lehto, 2008) and measurement (Casey et al., 2015; Seah & Horne, 2019), but also to mathematical domains that may not appear to be inherently spatial, such as numeration (e.g., Bates et al., 2021; Casey et al., 2015; Clements, 1999; Gunderson et al., 2012; Hawes et al., 2019b; Kyttälä & Lehto, 2008; Silverman & Ashkenazi, 2022; Zhang et al., 2014), algebra (e.g., Tolar et al., 2009), and data analysis and probability (e.g., Lowrie & Diezmann, 2007).

While research with secondary school pupils suggests that spatial ability may be important for solving certain types of word problems (Gilligan-Lee et al., 2022), few studies have investigated the relationship between spatial ability and arithmetic word problem solving in young children. The present study aims to contribute to addressing this gap in the literature in two ways. First, theoretically, the present study proposes a conceptual framework for categorizing one-step additive word problems, by integrating Thompson's (1993) distinction between quantity and quantitative relation, Davydov's (1982) discussion of methods of quantification, and Vergnaud's (1982) classification of dynamic and static relations. This framework provides a principled basis for examining how spatial ability may differentially relate to distinct forms of quantitative structure, an issue that has been largely overlooked in the spatial–mathematical literature.

Second, empirically, the present study investigates the relationship between spatial visualization, visuospatial working memory, and different types of one-step additive word problems in 6–7-year-old pupils in England. Applying the study's conceptual framework, the present study examines the following two research questions:

RQ1) To what extent do spatial visualization and visuospatial working memory predict young primary school pupils' (6–7-year-olds') accuracy on one-step additive word problems?

RQ2) Do the relationships between spatial visualization, visuospatial working memory, and accuracy on one-step additive word problems vary across different word problem types?

3.2 Literature Review and Conceptual Framework

Prior studies indicate that there are many task features (e.g., number of words, semantic complexity) and individual characteristics (e.g., age, gender, race, socioeconomic status, general intelligence) affecting arithmetic word-problem-solving performance (for a review, see

Verschaffel et al., 2020). In particular, the influence of language and numerical attributes has received considerable attention in the fields of educational psychology and mathematics education (for a review, see Daroczy et al., 2015; Vessonen et al., 2024). Language attributes are important considerations as the problem solver must be able to translate the text of the problem into a problem model (Hegarty et al., 1995; Ng & Lee, 2009). Therefore, the task's language features and the problem solver's language ability can influence the success of arithmetic word problem solving (Jaffe & Bolger, 2023). Likewise, although word problems differ from computational tasks because they require the integration of several additional cognitive processes, solving them still relies on executing mathematical operations, such that both the numerical features of the problem and the problem solver's numerical ability can influence the success of arithmetic word problem solving (Strohmaier et al., 2022).

Furthermore, studies find that non-verbal fluid reasoning is also an important factor in solving word problems (e.g., Fuchs et al., 2006; Fung & Swanson, 2017; Green et al., 2017; Kytälä & Lehto, 2008; Passolunghi et al., 2022; Peng et al., 2020; Wang et al., 2016). Non-verbal fluid reasoning is the ability to identify patterns and relationships, infer and implement rules, and do so independently of verbal and prior knowledge (McGrew, 2009). Although fluid reasoning and spatial skills are highly correlated, they are distinct constructs (Reynolds & Keith, 2017), relying on separable cognitive processes and brain regions (Atit et al., 2022). A recent meta-analysis revealed that fluid reasoning is more strongly related to mathematics than to reading, relates more strongly to more complex mathematical skills than to foundational skills, and that its relationship with mathematics increases with age (Peng et al., 2019).

However, language and numerical abilities and fluid reasoning cannot explain all of the variance in word-problem-solving performance. Even children who show good text comprehension skills and computation skills may fail to solve arithmetic word problems (Spencer et al., 2022). A recent meta-analysis of 45 studies found a significant positive moderate association between spatial and mathematical skills ($r = .36$, robust standard error = 0.035, $\tau^2 = 0.039$), and that a unique, direct relationship between spatial and mathematical skills remained even after including fluid reasoning and verbal skills in the model (Atit et al., 2022).

These findings suggest that successful word problem solving may require more than the ability to decode text; it may also depend on children's ability to construct representations of the quantitative structure described in the problem. Previous research suggests spatial ability may support the representation of quantitative relationships among quantities (e.g., Delgado &

Prieto, 2004). For example, in a recent investigation with first-year university pupils, Reinhold et al. (2020) found that spatial ability, measured using a three-dimensional mental rotation task, made a unique contribution ($\beta = .28$) in complex word problem solving, even after accounting for verbal ability ($\beta = .34$), numerical ability ($\beta = .16$) and fluid reasoning ($\beta = .25$).

Despite studies suggesting a potential role of spatial ability in mathematical word problem solving, this remains an underexplored area of spatial–mathematical link research in the context of early primary education. Two main challenges hinder efforts to delineate the relationship between spatial ability and word-problem-solving accuracy in young children. The first challenge is the diversity of spatial taxonomies and spatial measures (Subsection 3.2.1). The second challenge is the limited use of structurally grounded word problem categorization in spatial–mathematical research (Subsection 3.2.2). These challenges are reviewed in turn.

To address the limited use of structurally grounded approaches to categorizing word problems, Subsection 3.2.3 outlines the theoretical foundations that provide a principled basis for differentiating word problem types. Subsection 3.2.4 then examines how spatial ability may relate to quantitative structure, establishing the relevance of these foundations for understanding the spatial–mathematical link. Finally, building on these theoretical foundations and their connection to spatial ability, Subsection 3.2.5 presents this study’s proposed conceptual framework for categorizing one-step additive word problems.

3.2.1 First Challenge: Diversity of Spatial Taxonomies and Measures

The first challenge in investigating the relationship between spatial ability and word problem solving in early primary education is the diversity of spatial taxonomies and spatial measures. Intuitively, most people have a general understanding of what spatial ability entails. This perhaps owes to the fact that “our world is a world that exists [literally] in space, ... [which means] living in the world requires spatial functioning of some kind” (Newcombe & Shipley, 2015, p. 179). Yet, despite the ubiquity of spatial thinking, decades of research on spatial ability have failed to reach a uniform agreement on its definition or structure (for a review, see Schenck & Nathan, 2024).

While spatial ability is now recognized as distinct from general intelligence, there is still debate over whether it is a unitary or multidimensional construct (Mix & Cheng, 2012). Moreover, when considered as a multidimensional construct, spatial ability is subdivided, defined, and named in multiple ways (Schenck & Nathan, 2024). Beyond the vast differences in how spatial cognition is conceptualized, research on the spatial–mathematical link is further

complicated by the sheer number of spatial tasks available. Given the complex nature of spatial ability, the choice of spatial tasks may partly explain the mixed findings in this area. Given the lack of cohesion and the multiplicity of spatial taxonomies, studies investigating the relationship between word-problem-solving performance and spatial ability in the early primary education context must be intentional about which spatial skills are measured and which spatial tasks are used (Schenck & Nathan, 2024).

Two spatial skills that have been consistently identified as having a particularly strong association with mathematical thinking are spatial visualization (Hawes & Ansari, 2020) and visuospatial working memory (Allen et al., 2019). *Spatial visualization* is “the ability to generate, recall, maintain, and manipulate visual-spatial images and solutions” (Hawes & Ansari, 2020, p. 466). Typical measures of spatial visualization include two-dimensional mental rotation tasks, three-dimensional mental rotation tasks, mental composition tasks, and mental paper-folding tasks (Hawes & Ansari, 2020). An extensive body of literature suggests that spatial visualization skills and mathematical performance share a correlational (Sorby & Panther, 2020), causal (Hawes et al., 2022), and potentially even bi-directional (Geer et al., 2019) relationship. Hawes and Ansari (2020) suggest four non-mutually exclusive mechanistic accounts for the relationship between spatial visualization skills and basic numerical competencies. One of them, the spatial modelling account, may be especially relevant when considering the relationship between spatial visualization and word-problem-solving competencies. According to this account, spatial visualization provides “a ‘mental blackboard’ on which numerical relations and operations can be modelled and visualized” (Hawes & Ansari, 2020, p. 472).

Visuospatial working memory is the ability to temporarily store and manipulate visuospatial information (Baddeley & Hitch, 1974). There are two types of visuospatial working memory: simultaneous and sequential (Allen et al., 2019). The distinction between these two types is based on how stimuli are presented. In simultaneous visuospatial working memory tasks, all information is presented at the same time. In contrast, in sequential visuospatial working memory tasks, stimuli are presented in a sequence. A recent systematic review of 35 independent studies on the relationship between visuospatial working memory and mathematical performance indicates that sequential visuospatial working memory may be particularly promising for studying the spatial–mathematical link (Allen et al., 2019). Furthermore, the results of the review suggest that visuospatial working memory may have a greater influence on solving word problems than on computation tasks. In another meta-

analysis of 110 studies, Peng et al. (2019) found a significant medium correlation between mathematics and working memory ($r = .35$, 95% CI [.32, .37]) and that this relationship was strongest for word problem solving, while domains of working memory did not moderate the effect.

Mix (2019) suggests that spatial visualization and visuospatial working memory may be utilized differently depending on the problem solver's level of experience with the mathematical content. That is, while spatial visualization may be recruited for *spatially modelling novel mathematical content*, visuospatial working memory may support *procedural fluency for familiar mathematical content*. Furthermore, the recruitment of spatial visualization and visuospatial working memory may be cyclical: spatial visualization is recruited whenever new mathematical content is encountered, even at an older age, while visuospatial working memory is recruited whenever familiar mathematical content is encountered, even at a younger age (Mix et al., 2017).

Wang and colleagues (2022) investigated the role of both spatial visualization and visuospatial working memory within word problem solving in the areas of geometry and algebra with eighth-grade pupils, and the findings seem to support Mix's (2019) proposition. This study also varied the difficulty of the word problems, classifying them as either process-open or process-closed. Unlike process-closed problems, which have a well-defined procedure leading to a single correct answer (e.g., finding the average of a set of numbers), process-open problems allow for "a range of different interpretations, methods, and acceptable answers" (Wang et al., 2022, p. 1). Wang et al. (2022) found that both spatial visualization and visuospatial working memory were significantly associated with process-open problems. Furthermore, they found that spatial visualization was more strongly associated with solving difficult geometry and algebra process-open problems, whereas visuospatial working memory was more strongly associated with solving easier geometry and algebra process-open problems. Whether this potentially nuanced relationship between specific spatial skills and specific problem types also applies in the early primary education context remains to be investigated.

3.2.2 Second Challenge: Limited Use of Structurally Grounded Word Problem

Categorization in Spatial–Mathematical Research

In addition to the diversity of spatial taxonomies and measures, another challenge in investigating the spatial–mathematical link in early primary education is the limited use of structurally grounded approaches to categorizing word problems. Consideration of problem

type may be important, as studies with older pupils suggest that certain mathematical word problems require spatial ability more than others. For example, in a study with undergraduate pupils, Strohmaier et al. (2022) found that spatial ability most reliably predicted success on problems with visual representations, whereas other cognitive skills (verbal, arithmetic, and general reasoning skills) better predicted performance on other types of problems (e.g., those involving calculations or solvable using heuristics). Similarly, distinguishing between process-open and process-closed problems (Wang et al., 2022) suggests that spatial ability may relate differently across word problem types. Together, these examples suggest that differences in how mathematical problems are characterized may partly explain why some studies have found spatial ability to be important in solving mathematical word problems (e.g., Boonen et al., 2013; Hegarty & Kozhevnikov, 1999), while others have not (e.g., Nunes et al., 2009a).

One categorization of word problems that has received considerable attention in this area is the distinction between visual–spatial and verbal–analytical word problems (Casey et al., 2015). The primary distinction between these two types is the extent to which visualization is required for their solution (Kilpatrick, 1967; Krutetskii, 1976; Moses, 1977; Tillotson, 1984). *Visual–spatial word problems* are solved “relatively easily by visual-pictorial means if the correlation of the given elements in the problem is expressed visually” (Krutetskii, 1976, p. 156). An example of a visual–spatial problem is as follows:

“A balloon first rose 200 meters from the ground, then moved 100 meters to the east, then dropped 100 meters. It then travelled 50 meters to the east, and finally dropped straight to the ground, how far was the balloon from its original starting point?” (Hegarty & Kozhevnikov, 1999, p. 680).

Verbal–analytical word problems can also be solved visually, but visual representations do not necessarily aid the solution process. Instead, they can be solved using logical and algebraic reasoning without relying on visualization of relations. An example of a verbal–analytical problem is as follows:

“In a store there were 100 kg of berries. An analysis showed that the moisture in the berries was 99%. Determine the total weight of the berries when, by a later analysis, the moisture had decreased 1% and become 98%” (Krutetskii, 1976, p. 160).

It is perhaps not surprising that success in visual–spatial problems may be related to spatial ability given that they involve reasoning about relations that are literally spatial in nature. Indeed, a body of research with older pupils suggests that spatial skills are linked to

performance on visual–spatial word problems (e.g., Boonen et al., 2013, 2014; Casey et al., 2015; Duffy et al., 2020; Hegarty & Kozhevnikov, 1999; Kozhevnikov & Blazhenkova, 2013; Marschark et al., 2013; Oostermeijer et al., 2014; Van Garderen, 2006). It is theorized that spatial ability supports performance on visual–spatial word problems through the generation, manipulation, and transformation of accurate internal representations (e.g., mental imagery) and external representations (e.g., visual-schematic representations) (Boonen et al., 2013, 2014; Hegarty & Kozhevnikov, 1999; Kozhevnikov & Blazhenkova, 2013).

More noteworthy is the finding that success in solving verbal–analytical problems, which do not appear to be inherently spatial, may also be related to spatial ability, as suggested by a growing body of literature (e.g., Casey et al., 2015; Kyttälä & Lehto, 2008; Likhanov et al., 2024; Meyer et al., 2010; Mix et al., 2016, 2017; Tillotson, 1984; Wang et al., 2022). However, the role of spatial ability in verbal–analytical problems remains uncertain, as findings have been mixed (e.g., Kyttälä & Björn, 2014; Pina et al., 2014; Swanson & Fung, 2016), underscoring the need for further research to determine when spatial skills contribute to word problem solving as a function of problem characteristics.

While the distinction between visual–spatial and verbal–analytical word problems has helped delineate the relevance of spatial ability for older secondary school pupils, there is a scarcity of research investigating how spatial ability relates to different word problem types for younger primary school pupils. That is, in the context of early primary education, research on word problems and the spatial–mathematical link remains largely separate. Most studies either examine the relationship between spatial ability and mathematics using composite performance measures or, if they report word-problem performance separately, do not account for differences between word problem types. This absence of a structurally grounded approach may help explain the mixed findings in the literature. For example, some studies have reported positive associations between spatial ability and word-problem-solving performance in primary school pupils (e.g., Chan & Wong, 2019; Fung & Swanson, 2017; Gilligan et al., 2017; Kyttälä et al., 2014; Mix et al., 2017, 2021; Zhang & Lin, 2015, 2017; Zheng et al., 2011), whereas others have not (e.g., Campos et al., 2013; Carr et al., 2008; Ching & Nunes, 2017; Fuchs et al., 2010; Rodán et al., 2019). A theoretically grounded conceptual framework based on quantitative structure may therefore provide clearer insight into when spatial ability relates to word problem solving in the early primary years.

In summary, while the spatial–mathematical link is considered “one of the most robust and well-established findings in cognitive psychology” (Mix & Cheng, 2012, p. 198), there remains a paucity of research delineating the relationship between spatial ability and word-problem-solving performance, particularly in young children (Newcombe et al., 2019). Clarifying this relationship requires a more targeted research approach. In particular, two conditions must be met. First, spatial skills and their corresponding tasks must be selected based on empirical evidence to ensure their suitability for the population and mathematical skills under investigation (Schenck & Nathan, 2024). Second, a theoretically-grounded conceptual framework for categorizing word problems should be applied in research on the spatial–mathematical link.

The present study aims to address these issues by employing two spatial skills—spatial visualization (Hawes & Ansari, 2020) and visuospatial working memory (Allen et al., 2019), both of which have been consistently identified as strongly associated with mathematical thinking—and by proposing and applying a conceptual framework for categorizing one-step additive word problems according to their quantitative structure. To this end, the theoretical foundations informing the proposed conceptual framework are reviewed next.

3.2.3 Theoretical Foundations

The reviewed literature indicates that a theoretically grounded framework for differentiating word problem types is required to empirically examine how spatial ability relates to young pupils’ word problem solving. To address this need, this subsection outlines the theoretical foundation that provide a principled basis for such a framework, drawing on Thompson’s (1993) distinction between quantity and quantitative relation ([Subsection 3.2.3.1](#)), Davydov’s (1982) discussion of methods of quantification ([Subsection 3.2.3.2](#)), and Vergnaud’s (1982) classification of dynamic and static relations ([Subsection 3.2.3.3](#)).

3.2.3.1 Quantity versus Quantitative Relation

The distinction between quantity and quantitative relation (Thompson, 1993) is fundamental to a theoretically grounded framework for categorizing word problems because it underlies quantitative structure. A *quantity* is an attribute of an object that can be measured, such as the set size or the length (Thompson, 1993). In this regard, a quantity is an intrinsic property as it does not depend on anything other than the object itself (Dos Santos, 2022). A *quantitative relation*, in contrast, describes the connection between quantities (Nunes et al.,

2009a). There are two types of quantitative relations: additive (part-whole relation) and multiplicative (one-to-many correspondence or proportions/ratios) (Nunes & Bryant, 2015).

When measured, both quantity and quantitative relation can be represented with numbers. The distinction is that quantitative relation is best represented by “directed numbers” (Vergnaud, 1982, p. 46). Consider the premise, ‘Ann has 8 books. Bob has 3 books. Ann has 5 more books than Bob.’ The first two phrases describe quantities; the quantities of Ann and Bob’s books have been measured and are represented by numbers, 8 and 3, respectively. The last phrase describes the quantitative relation between the two quantities and is best represented by the directed number, +5. Moreover, being a quantitative relation means having an inverse (i.e., Bob has 5 less books than Ann), which can again be represented by the directed number, −5 (Dos Santos, 2022).

Both quantities and quantitative relations can be further subdivided. Quantity may be subdivided based on whether the measurement unit is predetermined or imposed by partitioning a whole into equal parts (Davydov, 1982), whereas quantitative relation may be subdivided based on whether the problem situation describes a dynamic transformation in state or a static relationship between states (Vergnaud, 1982). These subdivisions are presented in turn below.

3.2.3.2 Subdivision of Quantity: Discrete versus Continuous

Building on Thompson’s (1993) notion of quantity, the present study draws on Davydov’s (1982) discussion of measurement to further differentiate quantities. *Measurement* is the process of representing a quantity with a numerical value through the reiteration of a chosen unit (Davydov, 1982; Sophian 2008). In measurement, the unit defines what counts as ‘one,’ and this can either be *predetermined* or *imposed* (Davydov, 1982). In the former case, the whole discrete element itself functions as the unit (Mix et al., 2002). For example, when measuring the numerosity of marbles, what is considered ‘one’ is the whole marble, such that measuring numerosity requires only counting each marble. In contrast, in the latter case, the unit must be imposed and reiterated. For example, when measuring the weight of marbles, one must first determine the unit of measurement (e.g., grams, ounces, or stones) and then reiterate the imposed unit to arrive at the numerical value representing the weight.

From this distinction in measurement, quantities can be broadly categorized as discrete or continuous. *Discrete quantities* are specific and fixed, and measuring them involves only counting the discrete units through a one-to-one mapping between a set of objects and an ordered set of tags (e.g., set size). In contrast, *continuous quantities* describe entities measured

on continuous dimensions (e.g., mass, length, distance, area, density, time, etc.; Newcombe et al., 2015), using an imposed unit, such that measurement first involves dividing them into equal units before enumeration (e.g., position). For example, 'Ann has 8 books' is a statement presenting a discrete quantity, as the number of books Ann has is a specific, fixed quantity. On the other hand, the statement, 'Ann is 8 steps from the wall,' presents a continuous quantity given that Ann's position is determined via an imposed measurement unit (i.e., steps). Both discrete and continuous quantities, when measured, can both be integers (e.g., 6 candies and 8 steps) or rational values (e.g., 1/2 a box of candies and 3.5 meters).

3.2.3.3 Subdivision of Quantitative Relation: Dynamic versus Static

Quantitative relations in word problems can be broadly categorized as dynamic or static (Vergnaud, 1982). In Vergnaud's (1982) conceptualization of additive word problems, these correspond to transformations in state and relationships between states. Similar distinctions have been described by other researchers, such as active and static relationships (Carpenter & Moser, 1982), dynamic and comparison texts (Nesher, 1982), and change and compare situations (Riley et al., 1983). The difference between these two types can be illustrated using representative word problems.

The following is an example of a problem involving a dynamic transformation of state (or, as Vergnaud (1982) describes it, a 'state-transformation-state' problem): 'Ann had 9 candies. She ate 2 candies. How many candies does she have now?' There are two components involved in such problem. One, there is a temporal sequence of events ($T_1 < T_2 < T_3$) (Nesher, 1982). Two, some change in the state of the nature is introduced at T_2 (Carpenter & Moser, 1982), such that the quantity referred to in T_1 is changed (Nesher, 1982).

In contrast, the following is an example of a problem involving static relationship linking states (or, as Vergnaud (1982) describes it, a 'state-relationship-state' problem): 'Ann has 9 candies. She has 2 more than Bob. How many candies does Bob have?' Here, no action is implied (Carpenter & Moser, 1982). Instead, two disjoint quantities are compared on the same attribute (Nesher, 1982).

From this distinction, quantitative relations can be broadly categorized as dynamic or static, corresponding to transformations in state and relations between states (Vergnaud, 1982). A *dynamic relation* describes the transformation in the state of nature over time. For example, 'Ann played a game of marbles and won 10 more marbles,' describes a dynamic relation, or the change in the quantity of Ann's marbles (discrete quantity) over time. Similarly, 'Ann

walked 12 steps away from the wall,’ describes a dynamic relation, or the change in Ann’s position (continuous quantity) over time. In essence, a dynamic relation describes the transformation linking the initial quantity and the final quantity.

A *static relation*, in contrast, describes the comparison of disjoint quantities. Simply put, it represents the difference between the referent quantity and the compared quantity. For example, the statement ‘Bob has 11 more marbles than Ann,’ illustrates a static relation, where the number of marbles Bob possesses is compared to the number of marbles Ann possesses. Similarly, the statement ‘Bob is 8 steps farther from the wall than Ann,’ also describes a static relation, with Bob’s position being compared to Ann’s position.

A summary of the theoretical foundations and their corresponding categories are presented in Table 3-1. Following Thompson’s (1993) distinction, a problem’s quantitative structure can be characterized along the dimensions of quantities and quantitative relations. Quantities may be further subdivided into discrete and continuous categories based on the method of measurement (Davydov, 1982), whereas quantitative relations may be categorized as dynamic or static depending on whether the problem situation involves a transformation or a relationship between states (Vergnaud, 1982).

Table 3-1

Two Dimensions of One-Step Additive Word Problems

Dimension	Category	Definition
Quantity	Discrete Quantity	Problems in which the quantities are specific and fixed, measured through one-to-one counting (e.g., set size)
	Continuous Quantity	Problems in which the quantities are measured using an imposed unit (e.g., position)
Quantitative Relation	Dynamic Relation	Problems that describe the transformation in the quantity over time
	Static Relation	Problems that describe a comparison of disjoint quantities

Together, these distinctions provide a principled basis for characterizing the quantitative structure of word problems in terms of the types of quantities involved and the types of relations linking them. Such distinctions are essential for differentiating word problem types in a theoretically grounded manner. Importantly, they raise the question of how spatial ability may support reasoning about different forms of quantitative structure.

3.2.4 Spatial Ability and Quantitative Structure

Building on the theoretical distinctions outlined above, this subsection examines how spatial ability may support reasoning about different forms of quantitative structure. First, the literature on the association between spatial ability and reasoning about discrete and continuous quantities is reviewed. Next, the literature on the association between spatial ability and reasoning about dynamic and static relations is reviewed.

Existing literature suggests that spatial ability may be associated with reasoning about both discrete and continuous quantities. First, numerical representations themselves may be inherently spatial. For example, Dehaene and colleagues' (1993) findings on the Spatial-Numerical Association of Response Codes (SNARC) provide evidence for the mental number line, “a metaphor used to describe the tendency for individuals from Western cultures to conceive numbers as ordered magnitudes along a left-to-right axis” (Hawes & Ansari, 2020, p. 469). If numerical representations are spatially organized, even problems involving discrete quantities may be reasoned about spatially. This possibility is further supported by meta-analytic evidence indicating a strong association between number line estimation and mathematical performance ($r = .44$; Schneider et al., 2018). This association is stronger than those reported for other foundational numerical skills, such as symbolic ($r = .28$) and non-symbolic ($r = .22$) number comparison (Schneider et al., 2017).

Second, the quantitative structure of arithmetic word problems—regardless of whether they involve discrete or continuous quantities—may be spatially modeled. Solving word problems requires constructing a representation of the problem situation, a process that makes them more cognitively demanding than computation tasks (Jitendra et al., 2007; Wong & Ho, 2017). Research with older pupils suggests that spatial skills may support the construction of schematic representations that capture the “mathematical gist” of the situation (Wong, 2021, p. 20) and accurately reflect its underlying quantitative structure.

However, it may also be argued that spatial ability is more naturally aligned with reasoning about continuous quantities than discrete ones. Because space itself is fundamentally a continuous dimension and “arguably the best candidate for ‘basic’ status, if there is a basic dimension” (Newcombe et al., 2015, p. 495), representing continuous quantities such as length may align more with spatial ability. From this perspective, spatial ability may support the representation of continuous quantitative structure more readily than that of discrete

quantitative structure, thereby facilitating problem solving. Nevertheless, empirical evidence directly comparing these possibilities in young children remains limited.

Taken together, these perspectives suggest that spatial ability may plausibly contribute to reasoning about both discrete and continuous quantitative structures, albeit potentially to different extents and through different mechanisms. However, it remains unclear whether the strength or nature of this contribution differs by quantity category, underscoring the need to account for variation in quantity type in spatial–mathematical investigations. Notably, most prior research with young primary school pupils has focused predominantly on discrete-quantity problems (e.g., Casey et al., 2015; Gilligan et al., 2017; Mix et al., 2017, 2021; Zhang & Lin, 2015, 2017), highlighting the importance of extending investigation to continuous-quantity problems.

A parallel argument applies to quantitative relations. Existing literature suggests that spatial ability may also be associated with reasoning about both dynamic and static relations. First, dynamic-relation problems require mentally tracking transformation in quantity, whether the quantities involved are discrete or continuous. Spatial ability, which supports reasoning about relations between and within objects (Hawes et al., 2022), may facilitate such mental transformations. In particular, spatial visualization—which enables the manipulation and transformation of spatial relations—may be especially relevant, potentially more so than visuospatial working memory, which is primarily involved in recalling and maintaining spatial information (Hawes & Ansari, 2020).

At the same time, spatial ability may also be relevant for reasoning about static relations. Static-relation problems are often more difficult for children than dynamic-relation problems (Cummins et al., 1988; Riley & Greeno, 1988), possibly because of the difficulty of mentally representing and comparing disjoint quantities (Múñez et al., 2013). Unlike dynamic relations, which describe a change that can be simulated (e.g., gaining or losing marbles, moving toward or away from a reference point), static relations involve abstract comparisons such as ‘more than,’ ‘less than,’ ‘father than,’ or ‘closer than.’ These comparisons do not involve a tangible transformation and may therefore be more challenging to represent mentally (Gabler & Ufer, 2024). Given this, it is plausible that children with stronger spatial ability may be better able to structure and coordinate these abstract relations.

Taken together, these considerations suggest that spatial ability may plausibly contribute to reasoning about both dynamic and static relations, albeit potentially through

different cognitive demands. However, it remains unclear whether spatial skills are more strongly associated with one category of quantitative relation than the other, underscoring the need for empirical investigation.

3.2.5 A 2×2 Conceptual Framework for Categorizing One-Step Additive Word Problems

Despite these theoretical and empirical suggestions that spatial ability may relate differently to distinct forms of quantitative structure, the limited use of unified, theoretically grounded approaches to categorizing word problems make it difficult to examine these relationships systematically. To address this limitation, the present study proposes a conceptual framework for categorizing one-step additive word problems based on their quantitative structure, enabling their use in empirical investigation of the spatial–mathematical link.

It is important to note the scope of this conceptual framework. First, it only focuses on *one-step additive* word problems, that is, problems that can be solved by finding either the sum or the difference between two numerical values presented in the problem. Second, it is limited to problems comprising *two quantities and one quantitative relation* linking them (Nunes et al., 2009a). Other categories of one-step additive word problems, such as those comprising three quantities or three quantitative relations (Vergnaud, 1982), are not considered.

By integrating the dimensions of (1) quantities and (2) quantitative relations, a 2×2 conceptual framework (Quantity $\times$ Quantitative Relation) is proposed. As shown in Figure 3-1, the four resulting word problem types, beginning from the top-left quadrant, are: *Discrete Quantity–Dynamic Relation*, *Discrete Quantity–Static Relation*, *Continuous Quantity–Dynamic Relation*, and *Continuous Quantity–Static Relation* problems.

To provide a more nuanced representation of quantitative structures, Vergnaud’s (1982) notation system was adapted. Specifically, a distinction between discrete and continuous quantities was introduced. Whereas Vergnaud uses upright squares ($\square$) to represent quantities in general, the present framework uses upright squares to represent discrete quantities and introduces tilted squares ($\diamond$) to represent continuous quantities. Quantitative relations are represented by circles ($\circ$). Dynamic relations are depicted with horizontal arrows ($\rightarrow$), symbolizing a transformation in quantity, whereas static relations are represented by vertical arrows ($\uparrow$), symbolizing a comparison between disjoint quantities.

Figure 3-1

A 2×2 Conceptual Framework for One-Step Additive Word Problem

		Quantitative Relation					
		Dynamic Relation			Static Relation		
Quantity	Discrete Quantity	“Ann was playing a game of marbles. Before the game she had some marbles. Then she played a game of marbles where she either won more marbles or lost some of her marbles.”			“Ann and Bob are going to play a game of marbles. So, Ann brought some marbles, and Bob also brought some marbles to play with.”		
	Discrete Quantity	Ann had some marbles. Then she played a game of marbles and won 9 more marbles. Now she has 15 marbles. How many marbles did she have to begin with? 	Ann had 9 marbles. Then she played a game of marbles and won 10 more marbles. How many marbles does she have now? 	Ann had 5 marbles. Then she played a game of marbles. Now she has 19 marbles. How many marbles did she win? 	Ann has 19 marbles. She has 8 more marbles than Bob. How many marbles does Bob have? 	Ann has 6 marbles. Bob has 10 more marbles than Ann. How many marbles does Bob have? 	Ann has 13 marbles. Bob has 6 marbles. How many less marbles than Ann does Bob have?
Quantity	Continuous Quantity	“Ann was playing a game. She first stood some distance from the wall. Then she played a game where she either walked towards or walked away from the wall.”			“Ann and Bob were playing Red Light. Green Light. The wall is the finish line.”		
	Continuous Quantity	Ann was some steps from the wall. Then she walked 7 steps away from the wall. Now she is 12 steps from the wall. How far was Ann from the wall to begin with? 	Ann was 5 steps from the wall. Then she walked 9 steps away from the wall. How far is Ann from the wall now? 	Ann was 6 steps from the wall. Then she walked some steps. Now she is 19 steps from the wall. How many steps did Ann walk away from the wall? 	Ann is 19 steps from the wall. She is 7 steps farther from the wall than Bob. How far is Bob from the wall? 	Ann is 5 steps from the wall. Bob is 6 steps farther from the wall than Ann. How far is Bob from the wall? 	Ann is 17 steps from the wall. Bob is 8 steps from the wall. How many steps closer to the wall is Bob than Ann?

Note. Two dimensions (quantity and quantitative relation) form this study’s 2×2 conceptual framework for one-step additive word problems, resulting in four word problem types. Upright squares ($\square$) represent discrete quantities. Tilted squares ($\diamond$) represent continuous quantities. Circles ($\circ$) represent quantitative relations. Dynamic relations are represented horizontally, whereas static relations are represented vertically. When each word problem type (Discrete Quantity–Dynamic Relation (top-left), Discrete Quantity–Static Relation (top-right), Continuous Quantity–Dynamic Relation (bottom-left), and Continuous Quantity–Static Relation (bottom-right)) is framed as a one-step additive word problem, two of the three facts are provided, and the missing third fact becomes the unknown to be solved. The unknown is either a quantity ($\square$ or $\diamond$) or a quantitative relation ($\rightarrow$ or $\uparrow$).

Each word problem comprises two quantities and one quantitative relation. Of these three elements, two are provided in the problem, while the third constitutes the unknown to be solved. As illustrated in each quadrant of Figure 3-1, some problems involve a missing quantity (i.e., indicated by a missing $\square$ or $\diamond$), whereas others involve a missing quantitative relation (i.e., indicated by a missing $\rightarrow$ or $\uparrow$). The proposed 2×2 conceptual framework provides a theoretically grounded basis for differentiating one-step additive word problem types and guides the construction of problem items used to address the research questions.

3.3 Methods

The proposed conceptual framework was used to empirically investigate the relationship between spatial ability and word-problem-solving accuracy in early primary pupils. A correlational, cross-sectional design was used to examine the two substantive research questions: the extent to which spatial visualization and visuospatial working memory predict one-step additive word-problem-solving accuracy in early primary pupils, and whether this relationship varies by word problem type. [Subsection 3.3.1](#) describes the participants, and [Subsection 3.3.2](#) describes the measures, including the development of word-problem items based on the conceptual framework. [Subsection 3.3.3](#) outlines the administration procedure, and [Subsection 3.3.4](#) presents the analytic approach used to address the substantive research questions.

3.3.1 Participants

Data from 137 Year 2 pupils in England (50.4% female, 49.6% male; age range: 6.52–7.74 years; mean age: 7.10 years) were used. This research was approved by the University of Oxford’s Central University Research Ethics Committee and followed the ethical guidelines for educational research as recommended by the British Educational Research Association (2018) and the University of Oxford Policy on Research Involving Human Participants and Data Protection (Central University Research Ethics Committee, 2022). A non-probabilistic convenience sampling method was used to recruit participants. A two-step recruitment process was used whereby four schools in South East England were recruited first, followed by the pupils. Consent from appropriate gatekeepers (i.e., headteachers, teachers, and parents/guardians) and assent from pupils were obtained prior to data collection.

Among the 137 participants, 27 (19.7%) were reported as having Special Education Needs (SEN) and 21 (15.3%) were reported as having English as an Additional Language (EAL). They were included in the analyses either because their SEN support needs were minor,

or because their English proficiency was adequate. Therefore, there was no reason to believe they would be unable to engage with the tasks. As SEN and EAL information was not always available, it is possible that an even higher percentage of SEN and EAL pupils were included in the research.

3.3.2 Measures

As summarized in Table 3-2, each pupil completed a total of seven tasks. [Subsection 3.3.2.1](#) describes the word problem task developed based on the conceptual framework outlined above. [Subsection 3.3.2.2](#) then describes the three spatial ability measures used to assess the independent variables. Finally, [Subsection 3.3.2.3](#) describes the three covariate measures.

3.3.2.1 Word Problems Task

Informed by the 2×2 conceptual framework, 24 one-step additive word problems were developed by adapting problems from the literature (Carpenter et al., 1999; Nunes et al., 2009a). The full problem set is provided in Figure 3-2. The problems were developed through an iterative process, which included two pilot studies and one round of expert review with a psychologist specializing in children's mathematical development and a lecturer in mathematics education. Number facts with sums ranging from 11 to 19 were used in the word problems, as they were considered appropriate based on the National Curriculum for Year 2 pupils in England (Department for Education, 2013) and consultations with primary school teachers. (For a detailed description of the word problem development process, see [Subsection 2.3.3.1](#) of the Methodology chapter.)

Each word problem type was associated with a different scenario. The Discrete Quantity–Dynamic Relation problems (i.e., the top-left quadrant in Figure 3-2; Items 1–6) involved Ann playing a game of marbles by herself. These problems required solving for the number of marbles Ann had either at the end of the game (Items 1–2), at the start of the game (Items 3–4), or the number of marbles she won or lost during the game (Items 5–6). The Continuous Quantity–Dynamic Relation problems (i.e., the bottom-left quadrant in Figure 3-2; Items 13–18) involved Ann playing a game by herself, where she could either walk toward or away from the wall. These problems required solving for Ann's position either at the end of the game (Items 13–14), at the start of the game (Items 15–16), or the number of steps Ann

Table 3-2*Study Measures*

Variable	Task	Description	Scoring
Spatial Visualization	Wechsler Abbreviated Scale of Intelligence–Second Edition (WASI–II) Block Design subtest	While viewing a constructed model or a picture in the Stimulus Book, the participant uses red-and-white cubes to recreate the design within a specified time limit of 30 to 120 seconds per design. There are 11 items for children aged 6 to 8 years; testing is discontinued after two consecutive scores of zero.	• Scores range from 0 to 2 for the first four items. • Scores from 0 to 7 for the final seven items. • The total raw score is the sum of all item scores. • The raw score can range from 0 to 57 for children aged between 6 to 8 years.
	Visual–Spatial Reasoning task	Four subtasks (five items per subtask) that require reasoning about and visualizing solutions to problems involving parts of objects in relation to their whole: missing center, missing border, two-dimension shape composition/decomposition, and paper folding problems. There are 20 items in total and all items are administered.	• Each item is scored as either correct (1) or incorrect (0). • The total raw score is the sum of all item scores. • The raw score can range from 0 to 20.
Visuospatial Working Memory	PathSpan	A digitized version of the Corsi Block-Tapping Task on an iPad, in which the participant replicates a sequence of dots that light up one at a time by tapping them in the same order. The sequence length begins at two, and there are two trials per sequence length. The sequence length increases by one if the participant correctly repeats the pattern on either of the two trials at a given length. Testing is discontinued if the participant fails both trials at a given sequence length. The maximum sequence length is nine.	• The score is the highest sequence length accurately repeated at least once. • The raw score can range from 0 to 9.
Language Ability	Wechsler Individual Achievement Test–Second UK Edition (WIAT–II ^{UK}) Listening Comprehension	Three subtasks that involve selecting the picture that matches a word (Receptive Vocabulary; 16 items) or sentence (Sentence Comprehension; 10 items) and generating a word that matches a picture and oral description (Expressive Vocabulary; 15 items). Each of the three subtasks requires a basal set of three consecutive correct responses, and each subtask is discontinued after six consecutive scores of zero.	• Each item is scored as either correct (1) or incorrect (0). • The total raw score is the sum of all item scores. • The raw score can range from 0 to 41.

Variable	Task	Description	Scoring
Numerical Ability	Wechsler Individual Achievement Test–Second UK Edition (WIAT–II ^{UK}) Numerical Operations	Identification and writing of numbers, counting using one-to-one correspondence, and solving written calculation problems and simple equations involving the basic operations of addition, subtraction, multiplication, and division. A basal set of three consecutive correct responses are required, and testing is discontinued after 6 consecutive scores of zero. There are 54 items in total.	• Each item is scored as either correct (1) or incorrect (0). • The total raw score is the sum of all item scores. • The raw score can range from 0 to 54.
Fluid Reasoning	Wechsler Abbreviated Scale of Intelligence–Second Edition (WASI–II) Matrix Reasoning subtest	While viewing an incomplete matrix or series, the participant selects the response option that would complete the matrix or series from five options. There are 24 items for children aged 6 to 8, but testing is discontinued after three consecutive scores of zero.	• Each item is scored as either correct (1) or incorrect (0). • The total raw score is the sum of all item scores. • The raw score can range from 0 to 24 for children aged between 6 to 8 years.
Word-Problem-Solving Accuracy	Word Problems	Additive word problems that require either the addition or the subtraction of the values given the problem. There are 24 items in total and all items are administered.	• Each item is scored as either correct (1) or incorrect (0).

Note. Adapted from Table 2-2 (Chapter 2).

Figure 3-2*One-Step Additive Word Problems*

		Quantitative Relation					
		Dynamic Relation		Static Relation			
Quantity	Discrete Quantity	“Ann was playing a game of marbles. Before the game she had some marbles. Then she played a game of marbles where she either won more marbles or lost some of her marbles.”		“Ann and Bob are going to play a game of marbles. So, Ann brought some marbles, and Bob also brought some marbles to play with.”			
		1	Ann had x marbles. Then she played a game of marbles and won y more marbles. How many marbles does she have now?	+	7	Ann has x marbles. Bob has y more marbles than Ann. How many marbles does Bob have?	+
		2	Ann had y marbles. Then she played a game of marbles and lost x marbles. How many marbles does she have now?	−	8	Ann has y marbles. Bob has x marbles less than Ann. How many marbles does Bob have?	−
		3	Ann had some marbles. Then she played a game of marbles and won x more marbles. Now she has y marbles. How many marbles did she have to begin with?	−	9	Ann has y marbles. She has x more marbles than Bob. How many marbles does Bob have?	−
		4	Ann had some marbles. Then she played a game of marbles and lost x marbles. Now she has y marbles left. How many marbles did she have to begin with?	+	10	Ann has x marbles. She has y marbles less than Bob. How many marbles does Bob have?	+
		5	Ann had x marbles. Then she played a game of marbles. Now she has y marbles. How many marbles did she win?	−	11	Ann has y marbles. Bob has x marbles. How many less marbles than Ann does Bob have?	−
		6	Ann had y marbles. Then she played a game of marbles. Now she has x marbles left. How many marbles did she lose?	−	12	Ann has x marbles. Bob has y marbles. How many more marbles than Ann does Bob have?	−
	Continuous Quantity	“Ann was playing a game. She first stood some distance from the wall. Then she played a game where she either walked towards or walked away from the wall.”		“Ann and Bob were playing Red Light. Green Light. The wall is the finish line.”			
		13	Ann was x steps from the wall. Then she walked y steps away from the wall. How far is Ann from the wall now?	+	19	Ann is x steps from the wall. Bob is y steps farther from the wall than Ann. How far is Bob from the wall?	+
		14	Ann was y steps from the wall. Then she walked x steps towards the wall. How far is Ann from the wall now?	−	20	Ann is y steps from the wall. Bob is x steps closer to the wall than Ann. How far is Bob from the wall?	−
		15	Ann was some steps from the wall. Then she walked x steps away from the wall. Now she is y steps from the wall. How far was Ann from the wall to begin with?	−	21	Ann is y steps from the wall. She is x steps farther from the wall than Bob. How far is Bob from the wall?	−
		16	Ann was some steps from the wall. Then she walked x steps towards the wall. Now she is y steps from the wall. How far was Ann from the wall to begin with?	+	22	Ann is x steps from the wall. She is y steps closer to the wall than Bob. How far is Bob from the wall?	+
		17	Ann was x steps from the wall. Then she walked some steps. Now she is y steps from the wall. How many steps did Ann walk away from the wall?	−	23	Ann is y steps from the wall. Bob is x steps from the wall. How many steps closer to the wall is Bob than Ann?	−
		18	Ann was y steps from the wall. Then she walked some steps. Now she is x steps from the wall. How many steps did Ann walk towards the wall?	−	24	Ann is x steps from the wall. Bob is y steps from the wall. How many steps is Bob farther from the wall than Ann?	−

Note. Reproduced from Figure 2-9 (Chapter 2).

took during the game (Items 17–18). Items 1–4 and 13–16 were quantity unknown problems, while Items 5–6 and 17–18 were quantitative relation unknown problems.

The Discrete Quantity–Static Relation problems (i.e., the top-right quadrant in Figure 3-2; Items 7–12) involved Ann and Bob playing a game of marbles together. These problems required solving for either the number of marbles Bob had (Items 7–10) or the difference in the number of marbles Ann and Bob had (Items 11–12). The Continuous Quantity–Static Relation problems (i.e., the bottom-right quadrant in Figure 3-2; Items 19–24) involved Ann and Bob playing a game of red light, green light. These problems required solving for either Bob’s position with respect to the finish line (Items 19–22) or the difference in Ann’s and Bob’s positions (Items 23–24). Items 7–10 and 19–22 were quantity unknown problems, while Items 11–12 and 23–24 were quantitative relation unknown problems.

As studies indicate that language and numerical task features affect word-problem-solving success (Verschaffel et al., 2020), best efforts were made to control for them. First, to minimize the effect of language features, the word problems were developed with minimal variation in wording and structure. The continuous-quantity problems were designed to align closely with their discrete-quantity counterparts to ensure consistency. Likewise, the dynamic-relation problems were designed to align closely with their static-relation counterparts. The word problems were condensed to make them as concise as possible without compromising situational understanding (Stern & Lehrndorfer, 1992).

Second, to minimize the effect of numerical features, four alternative parallel versions of the word problem sets (Forms A–D) were designed so that the wording and structure of the items were identical across all forms, with the only difference being the specific numerical values presented in each problem (for a review, see Daroczy et al., 2015; Vessonen et al., 2024). This was done by selecting 24 unique number facts with sums ranging from 11 to 19 (as shown in Figure 3-3) and subjecting them to a systematic randomization process (for details, see [Subsections 2.3.3.1.3](#) and [2.3.3.1.4](#)). The full word problem sets (Forms A–D) are provided in [Subsection 2.3.3.2](#) in the Methodology chapter.

Figure 3-3

24 Number Facts up with Sums Ranging from 11 to 19

5, 14, 19	9, 10, 19	8, 10, 18	8, 9, 17	6, 9, 15	5, 8, 13
6, 13, 19	5, 13, 18	5, 12, 17	5, 11, 16	7, 8, 15	6, 7, 13
7, 12, 19	6, 12, 18	6, 11, 17	6, 10, 16	5, 9, 14	5, 7, 12
8, 11, 19	7, 11, 18	7, 10, 17	7, 9, 16	6, 8, 14	5, 6, 11

Note. Reproduced from Figure 2-8 (Chapter 2).

The measurement quality of the word problem—including their reliability, structural validity, and form equivalence—was evaluated through a series of empirical analyses. Analysis of the items indicated that the four test forms were equivalent in terms of item difficulty and discrimination. The structural validity of the proposed 2×2 conceptual framework was further supported through confirmatory factor analysis; results indicated that the hypothesized four-factor model—differentiating problems by both quantity and quantitative relation dimensions—provided a better fit to the data compared to simpler alternatives. Taken together, these findings provide empirical support for the conceptual framework, confirming that the items reliably and consistently measure word-problem-solving ability according to the specified dimensions (for a detailed description of the various statistical analyses conducted to evaluate the reliability and validity of the word problems and the underlying conceptual framework, see [Subsection 2.3.3.5](#) of the Methodology chapter).

3.3.2.2 Spatial Measures

The two measures of spatial visualization were the WASI–II Block Design subtest (Atit et al., 2022) and the Visual–Spatial Reasoning task (Hawes et al., 2015, 2017, 2019b). The WASI–II Block Design subtest involves recreating a two-dimensional design shown in the Stimulus Book using three-dimensional red-and-white cubes within a specified time limit. The Visual–Spatial Reasoning task consists of four subtasks that require reasoning about and visualizing solutions to problems involving parts of objects in relation to their whole: missing center, missing border, two-dimension shape composition/decomposition, and paper-folding problems.

These two measures have distinct strengths and limitations. The WASI–II Block Design is a standardized test with high reliability and validity (Wechsler, 2011), frequently used to assess spatial visualization since it involves “visualizing and/or mentally manipulating objects or figures” (Atit et al., 2022, p. 708); however, it is occasionally used more broadly as a measure of nonverbal problem-solving ability (e.g., Fuchs et al., 2005). Conversely, while the Visual–Spatial Reasoning task has primarily been used with Canadian pupils, its subtasks align closely with the conceptualization of spatial visualization adopted in this study. To address these individual limitations and create a more parsimonious predictor, both measures were used to derive a single composite score via Principal Component Analysis (PCA) (De Pauw et al., 2009; Schenck & Nathan, 2024; for further detail, see [Subsection 2.3.2.1.1](#) of the Methodology chapter).

To measure visuospatial working memory, PathSpan, a digitized version of the Corsi Block-Tapping task designed for use on an iPad with primary school pupils, was used (e.g., Coolen & Castronovo, 2023; Hawes et al., 2019b; Rittle-Johnson et al., 2019; Valcan et al., 2020). In this task, participants are presented with a sequence of buttons that flash one at a time in a random pattern, which they must then replicate by touching the buttons in the same order they flashed. As such, this is a measure of sequential visuospatial working memory.

3.3.2.3 Covariate Measures

The WIAT-II^{UK} Listening Comprehension subtest, which consists of three subsections—Receptive Vocabulary (16 items), Sentence Comprehension (10 items), and Expressive Vocabulary (15 items)—was used as a measure of language ability. Listening comprehension was deemed a more relevant measure of language ability than reading comprehension, as the one-step additive word problems were presented both in writing and orally (Spencer et al., 2022).

The WIAT-II^{UK} Numerical Operations subtest was used as a measure of numerical ability. This pencil-and-paper task assessed the ability to identify and write numbers, count using one-to-one correspondence, and solve calculation problems and simple equations involving the basic operations of addition, subtraction, multiplication, and division.

The WASI-II Matrix Reasoning subtest, which is modeled after a Raven's Progress Matrix Reasoning (Raven, 1938), was used as a measure of non-verbal fluid reasoning. While fluid reasoning tests may involve some visualization or spatial working memory, they are distinct from measures that primarily assess spatial ability because they require inductive and deductive reasoning (Schneider & McGrew, 2012). Participants were required to select the correct missing piece from five response options to complete the matrix or series (Atit et al., 2022).

3.3.3 Procedure

Pupils were administered a battery of seven tasks one-on-one over two testing sessions, each lasting approximately 30 minutes and typically spaced at least an hour apart (but no less than 30 minutes) to minimize fatigue effects and ensure consistent performance across tasks. The order of the seven tasks was randomized. The 24 word problems were divided evenly across two sessions (i.e., 12 items per session).

Pupils were randomly assigned to one of the four forms of word problems (Forms A–D). The word problems were presented on A4-size papers, with four problems arranged in each

quadrant of a page. Each problem item was read aloud to minimize the effect of reading fluency (Nunes et al., 2015). Problems were read slowly and repeated if needed to reduce memory load (Riley et al., 1983). Pupils were provided with 30 MathLink cubes, and a pencil. At the beginning of the task, they were informed that they could use these materials if they found them helpful and were reminded of the tools if they were stuck on a problem for more than a minute ('Think about how you would solve this problem. You can draw on the paper and use these counters, if you would like.').

Within each form, the order of the four problem types (i.e., scenarios) was randomized subject to predefined sequencing constraints. Specifically, within each session, the two discrete-quantity problems (top quadrants in Figure 3-2) were not administered consecutively, nor were the two continuous-quantity problems (bottom quadrants in Figure 3-2), in order to reduce boredom due to repetition.

All six problems belonging to the same word problem type were administered consecutively. Before the first problem from each quadrant was introduced, the scenario associated with that quadrant (shown in the blue shaded cells in Figure 3-2) was read aloud to provide situational context (Stern & Lehrndorfer, 1992). Pupils' comprehension of the situational context was then checked by asking clarifying questions (e.g., 'What does it mean to walk towards something?').

The order of the six problems within each problem type was randomized to control for potential practice, fatigue, and carryover effects (Punch & Oancea, 2014). However, the easier problem pair (i.e., Items 1–2, 7–8, 13–14, and 19–20) was always administered first to increase pupil motivation by increasing the likelihood of early problem-solving success (Hattie, 2009). The internal order of the two items was randomized (Hegarty et al., 1995; Langhorst et al., 2012; Mayer & Hegarty, 1996; Rosenthal & Resnick, 1974). After the easier problem pair was administered, the remaining problem items within the same quadrant were presented in randomized order; however, the two items within each problem pair (i.e., Items 3–4, 5–6, 9–10, 11–12, 15–16, 17–18, 21–22, and 23–24) were always administered consecutively, with their internal order randomized.

The interview-like procedure for administering word problems was adapted from Hegarty and Kozhevnikov (1999) and Boonen et al. (2013, 2014). After solving each problem, pupils were asked to report how they solved it; however, these strategy reports were not used

in the present study. (The detailed administration procedure for the word problems is outlined in [Subsection 2.3.3.3](#) of the Methodology chapter.)

Ethical procedures were implemented to ensure pupils' comfort and voluntary participation. Testing was conducted in the presence of a familiar adult, and pupils provided assent at each session. Participants were reminded that they could discontinue at any time without consequence. Testing was stopped immediately if a pupil indicated discomfort or withdrawal of assent.

3.3.4 Analytic Approach

Statistical analyses were conducted in R version 4.4.0 (R Core Team, 2024), using RStudio version 2023.12.1.402 (Posit Team, 2024) and *lme4* package (Bates et al., 2015). The first substantive research question—examining the relationship between spatial ability (i.e., spatial visualization and visuospatial working memory) and one-step additive word-problem-solving accuracy in 6–7-year-olds—was addressed using cross-classified mixed-effects logistic regression. The second research question—whether the relationship between spatial ability and word-problem-solving accuracy varies by word problem type—was examined by adding interaction terms between spatial skills and word problem type.

Each word problem was scored as either correct (1) or incorrect (0), and the binary outcome served as the dependent variable. Mixed-effects logistic regression models were estimated using a logit link and maximum likelihood estimation. The *bobyqa* optimizer was used with an increased iteration limit (100,000 function evaluations) to facilitate convergence (Powell, 2009).

Although test-level analyses provide an accessible summary, aggregating dichotomous item responses into total scores obscures item-level information, as pupils with identical scores may exhibit different problem-solving patterns. Moreover, when outcomes are derived from dichotomous items, applying ordinary least squares (OLS) regression can be inappropriate, as it may violate assumptions such as homoscedasticity and yield predicted probabilities outside the [0, 1] range (Cohen et al., 2003). In contrast, item-level logistic regression appropriately handles binary outcomes and avoids these limitations of OLS (Boonen et al., 2014).

To account for the cross-classified structure of the data, in which responses were nested within both participants and items, the models included random intercepts for participants (capturing between-participant variability in ability) and items (capturing between-item variability in difficulty) (Baayen et al., 2008; Barr et al., 2013). In addition, a more complex

random-effects structure was evaluated in which the effects of spatial visualization and visuospatial working memory were allowed to vary across items through random slopes. This specification was considered because the association between spatial skills and problem-solving accuracy may differ across individual word problems. Model comparisons were subsequently conducted to determine whether the additional random-effects complexity was supported by the data.

Continuous variables were mean centered prior to analysis to facilitate interpretation of main effects and reduce non-essential multicollinearity in models including interaction terms (Aiken & West, 1991). The composite spatial visualization score was derived from a principal component analysis of the Visual–Spatial Reasoning task and the WASI–II Block Design subtest; the first principal component was retained as the spatial visualization index.

Model assumptions and robustness were evaluated using the *performance* (Lüdtke et al., 2021) and *DHARMA* (Hartig, 2024) packages, including checks for multicollinearity, dispersion, residual patterns, influential observations, and convergence. Linearity of the logit was assessed by comparing linear and restricted cubic spline specifications (2–5 degrees of freedom) using AIC and BIC (Gauthier et al., 2020; Harrell, 2015), supplemented by visual inspection of predicted probability curves (see [Appendix 3A](#)).

Because the outcome variable was binary, intraclass correlation coefficient (ICC) and R^2 statistics were calculated using the latent-variable approach commonly applied in multilevel logistic regression, which assumes a level-1 residual variance of $\pi^2/3 \approx 3.29$ on the logit scale (Merlo et al., 2006). Accordingly, the marginal and conditional R^2 values reported should be interpreted as pseudo- R^2 measures rather than as directly comparable to R^2 from linear regression models.

3.4 Findings

This section presents the findings from the mixed-effects logistic regression analyses. [Subsection 3.4.1](#) presents the preliminary analyses prior to examining the research questions. [Subsection 3.4.2](#) addresses RQ1 by examining whether spatial visualization and visuospatial working memory predict one-step additive word-problem-solving accuracy in young children. [Subsection 3.4.3](#) addresses RQ2 by examining whether this relationship varies across word problem types.

3.4.1 Preliminary Analyses

Table 3-3 presents descriptive statistics and bivariate correlations among variables. Pearson correlations were used for continuous variables, while correlations involving the binary variable (gender), which was dummy-coded, correspond to point-biserial correlations. Total problem-solving accuracy was operationalized as the sum of correctly solved items (out of 24) and treated as continuous for the purposes of these correlations. Spatial visualization and visuospatial working memory were positively associated with word-problem-solving accuracy ($r = .44, p < .001$; $r = .35, p < .001$). All other predictors were also significantly associated with word-problem-solving accuracy, with the exception of gender.

Across the 24 word problems, the overall proportion of correct responses was .41. Item-level accuracy ranged from .18 to .80, indicating substantial variability in item difficulty. This variability in item-level accuracy supported the use of a mixed-effects modeling framework with random intercepts for items.

Mean problem-solving accuracy differed across problem types. Discrete Quantity–Dynamic Relation problems showed the highest accuracy ($M = .54, SD = .50$), followed by Discrete Quantity–Static Relation problems ($M = .39, SD = .49$) and Continuous Quantity–Dynamic Relation problems ($M = .34, SD = .47$). Continuous Quantity–Static Relation problems showed the lowest accuracy ($M = .29, SD = .45$). This variability in accuracy across word problem types suggests potential differences in problem structure that may moderate the association between predictors and word-problem-solving accuracy.

Table 3-3*Descriptive Statistics and Bivariate Correlations*

	Mean (<i>SD</i>)	Min	Max	1	2	3	4	5	6	7	8
1. Spatial Visualization	0.00 (1.00)	-1.90	3.53	-	.42***	.26**	.55***	.53***	.44***	.02	.11
2. Visuospatial Working Memory	4.39 (1.21)	2	8	-	-	.16	.39***	.29***	.35***	.06	.07
3. Language Ability	17.78 (3.89)	8	27	-	-	-	.36***	.42***	.46***	-.01	-.17*
4. Numerical Ability	12.71 (3.26)	5	23	-	-	-	-	.57***	.68***	.19*	.16
5. Fluid Reasoning	9.03 (4.22)	2	20	-	-	-	-	-	.62***	.23**	.02
6. Word Problem Solving	9.77 (7.12)	0	23	-	-	-	-	-	-	.24**	.14
7. Age (years)	7.10 (0.30)	6.52	7.74	-	-	-	-	-	-	-	.09
8. Gender	0.50 (0.50)	0	1	-	-	-	-	-	-	-	-

Note. $N = 137$. *SD* = standard deviation. For gender, 0 = girls, and 1 = boys. The spatial visualization score was derived from PCA of Visual–Spatial Reasoning task and the WASI–II Block Design subtest. *** $p < .001$, ** $p < .01$, * $p < .05$.

3.4.2 RQ1: The Relationship between Spatial Skills and Word-Problem-Solving Accuracy

To address RQ1, an item-level mixed-effects logistic regression model was fitted with random intercepts for participants and items. Diagnostic checks indicated low multicollinearity (all VIFs < 2), no influential observations, and no evidence of overdispersion. Spline specifications yielded limited or inconsistent improvements in model fit across predictors; accordingly, linear specifications were retained for all continuous predictors. Table 3-4 summarizes the fixed effects (odds ratio and 95% confidence intervals), random effects, and model fit indices for the main-effects model.

Table 3-4

Mixed-Effects Logistic Regression Predicting Word-Problem-Solving Accuracy (Main-Effects Model)

	β (SE)	Odds Ratios [95% CI]	z
Fixed Effects			
(Intercept)	-1.02 (0.28)		-3.66***
Spatial Visualization	-0.02 (0.16)	0.98 [0.72, 1.35]	-0.10
VSWM	0.15 (0.11)	1.16 [0.93, 1.45]	1.30
Language Ability	0.11 (0.04)	1.12 [1.04, 1.20]	3.11**
Numerical Ability	0.27 (0.05)	1.31 [1.18, 1.45]	5.31***
Fluid Reasoning	0.10 (0.04)	1.11 [1.02, 1.19]	2.60**
Age	0.92 (0.42)	2.51 [1.09, 5.92]	2.18*
Gender	0.39 (0.25)	1.47 [0.90, 2.43]	1.56
Random Effects			
$\sigma_{\text{Participant}}$		1.25	
σ_{Item}		1.07	
ICC		.32	
Marginal R^2 / Conditional R^2		.29/.61	
AIC		3039	
BIC		3100	
log L		-1509	

Note. $n_{\text{participants}} = 137$; $n_{\text{items}} = 24$; $n_{\text{observations}} = 3288$. β = log-odds coefficient; SE = standard error; σ = standard deviation; ICC = intraclass correlation coefficient; AIC = Akaike information criterion; BIC = Bayesian information criterion; log L = log-likelihood. *** $p < .001$, ** $p < .01$, * $p < .05$.

Random intercept standard deviation was 1.25 for participants and 1.07 for items, indicating variability in baseline problem-solving accuracy across both levels, with greater

variation attributable to participants than items. The unadjusted intraclass correlation coefficient ($ICC = .32$) suggested that approximately 32% of the variance in problem-solving accuracy was attributable to clustering across participants and items. The fixed effects accounted for 29% of the variance (marginal pseudo- $R^2 = .29$), while the full model including random effects accounted for 61% of the variance (conditional pseudo- $R^2 = .61$).

Spatial visualization ($OR = 0.98$, 95% CI [0.72, 1.35]) and visuospatial working memory ($OR = 1.16$, 95% CI [0.93, 1.45]) were not significant predictors of word-problem-solving accuracy, controlling for covariates. In contrast, language ability was positively associated with problem-solving accuracy ($OR = 1.12$, 95% CI [1.04, 1.20]), corresponding to a 12% increase in the odds of solving a problem correctly for each one-unit increase in language ability. Numerical ability was also positively associated with accuracy ($OR = 1.31$, 95% CI [1.18, 1.45]), as was fluid reasoning ($OR = 1.11$, 95% CI [1.02, 1.19]). Age was positively associated with accuracy ($OR = 2.51$, 95% CI [1.09, 5.92]), whereas gender ($OR = 1.47$, 95% CI [0.90, 2.43]) was not a significant predictor.

3.4.3 RQ2: Interaction Effects of Spatial Skills and Problem Type on Word-Problem-Solving Accuracy

To examine whether the association between spatial skills and problem-solving accuracy varied across word problem types, interaction terms were added to the main-effects model described above. Discrete Quantity–Dynamic Relation problems were specified as the reference category because they exhibited the highest average accuracy and thus served as a meaningful benchmark.

A model allowing the effects of spatial visualization and visuospatial working memory to vary across items was compared with a simpler random-intercept model. The likelihood-ratio test indicated that the random slopes did not significantly improve model fit, $\chi^2(5) = 0.64$, $p = .99$. Consistent with this findings, the random-slope model exhibited a higher AIC (3046) than the simple random-intercept model (3036). Accordingly, the more parsimonious random-intercept model was retained. Table 3-5 summarizes the fixed effects (odds ratio and 95% confidence intervals), random effects, and model fit indices for the interaction-effects model.

The interaction model provided a statistically significant improvement in fit relative to the main-effects model, $\chi^2(9) = 20.58$, $p = .015$. The AIC decreased slightly ($\Delta AIC = -2.58$), whereas BIC increased due to the additional parameters. Given the significant likelihood ratio test and the theoretical motivation for examining moderation by word problem type, the

Table 3-5

Mixed-Effects Logistic Regression Predicting Word-Problem-Solving Accuracy (Interaction-Effects Model)

	β (SE)	Odds Ratios [95% CI]	z
Main Effects			
(Intercept)	0.10 (0.38)		0.25
Spatial Visualization	0.19 (0.19)	1.21 [0.84, 1.80]	1.03
Visuospatial Working Memory	0.07 (0.13)	1.07 [0.82, 1.40]	0.50
Language Ability	0.11 (0.04)	1.12 [1.04, 1.20]	3.11**
Numerical Ability	0.27 (0.05)	1.30 [1.18, 1.45]	5.30***
Fluid Reasoning	0.10 (0.04)	1.11 [1.03, 1.20]	2.66**
Age	0.92 (0.42)	2.51 [1.09, 5.92]	2.18*
Gender	0.39 (0.25)	1.47 [0.90, 2.43]	1.56
Discrete Quantity–Static Relation	−1.09 (0.49)	0.33 [0.12, 0.91]	−2.22*
Continuous Quantity–Dynamic Relation	−1.48 (0.49)	0.23 [0.08, 0.62]	−2.99**
Continuous Quantity–Static Relation	−1.86 (0.50)	0.16 [0.06, 0.43]	−3.75***
Interaction Effects			
Spatial Visualization × Discrete Quantity–Static Relation	−0.16 (0.16)	0.85 [0.62, 1.16]	−1.03
Spatial Visualization × Continuous Quantity–Dynamic Relation	−0.28 (0.16)	0.76 [0.55, 1.03]	−1.78†
Spatial Visualization × Continuous Quantity–Static Relation	−0.40 (0.16)	0.67 [0.49, 0.91]	−2.58**
Visuospatial Working Memory × Discrete Quantity–Static Relation	0.03 (0.12)	1.03 [0.81, 1.32]	0.25
Visuospatial Working Memory × Continuous Quantity–Dynamic Relation	0.08 (0.13)	1.09 [0.84, 1.40]	0.67
Visuospatial Working Memory × Continuous Quantity–Static Relation	0.22 (0.13)	1.25 [0.97, 1.61]	1.74†
Random Effects			
$\sigma_{\text{Participant}}$		1.55	
σ_{Item}		0.67	
Intraclass Correlation Coefficient (ICC)		.26	
Marginal R^2 / Conditional R^2		.35/.61	
Akaike Information Criterion (AIC)		3036	
Bayesian Information Criterion (BIC)		3152	
Log-Likelihood (log L)		−1499	

Note. $n_{\text{participants}} = 137$; $n_{\text{items}} = 24$; $n_{\text{observations}} = 3288$. Reference category: Discrete Quantity–Dynamic Relation. *** $p < .001$, ** $p < .01$, * $p < .05$, † $p < .10$.

interaction model was retained. The addition of interaction terms led to a modest increase in the variance explained by the fixed effects (marginal pseudo- $R^2 = .35$, up from $.29$), suggesting that the interactions may capture additional systematic variance. The total variance explained by the model, including random effects, remained stable (conditional pseudo- $R^2 = .61$).

Baseline accuracy differed across word problem types. Relative to Discrete Quantity–Dynamic Relation problems, accuracy was significantly lower for Discrete Quantity–Static Relation (OR = 0.33, 95% CI [0.12, 0.91]), Continuous Quantity–Dynamic Relation (OR = 0.23, 95% CI [0.08, 0.62]), and Continuous Quantity–Static Relation (OR = 0.16, 95% CI [0.06, 0.43]) problems. Although the CIs are relatively wide, indicating some potential imprecision in the estimates, they are consistently below 1, suggesting a stable directional effect.

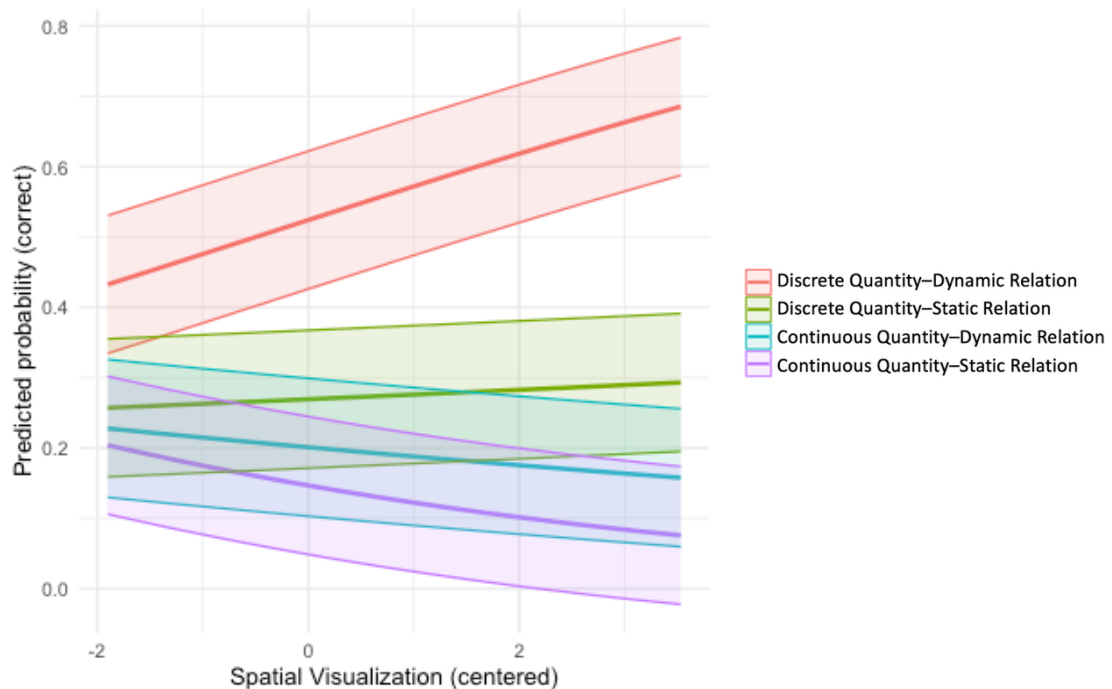
There was a significant interaction between spatial visualization and word problem type. Specifically, the interaction between spatial visualization and Continuous Quantity–Static Relation problems (relative to the reference Discrete Quantity–Dynamic Relation problems) was statistically significant (interaction OR = 0.67, 95% CI [0.49, 0.91]), indicating that the association between spatial visualization and problem-solving accuracy differed between these two problem types.

Simple slope analyses showed that, in the reference Discrete Quantity–Dynamic Relation problems, spatial visualization was positively—but not statistically significantly—associated with problem-solving accuracy ($\beta = 0.19$, 95% CI [−0.17, 0.56], SE = 0.19). In contrast, in Continuous Quantity–Static Relation problems, the association with problem-solving accuracy was non-significantly negative ($\beta = -0.21$, 95% CI [−0.56, 0.14], SE = 0.18). Thus, the significant interaction reflects a statistically reliable difference between conditional effects rather than a significant predictive effect within any individual problem type.

Figure 3-4 illustrates the predicted probabilities of problem-solving accuracy across levels of spatial visualization for each word problem type. Descriptively, the association between spatial visualization and problem-solving accuracy was most positive for Discrete Quantity–Dynamic Relation problems, near zero for Discrete Quantity–Static Relation problems, and negative for Continuous Quantity–Dynamic Relation and Continuous Quantity–Static Relation problems; however, these within-condition slopes were not statistically significant.

Figure 3-4

Predicted Probability of Problem-Solving Accuracy as a Function of Spatial Visualization across Word Problem Types



In contrast, none of the interaction terms involving visuospatial working memory were statistically significant, indicating that its association with problem-solving accuracy did not differ significantly across problem types relative to the reference Discrete Quantity–Dynamic Relation problems.

Taken together, these findings indicate that the role of spatial visualization, but not visuospatial working memory, in predicting problem-solving accuracy varied across problem types as a function of their quantitative structure. Although the difference in slopes between problem types was statistically significant, spatial visualization did not independently predict accuracy within any single problem type. The interaction therefore reflects a relative difference between problem types rather than a robust within-condition effect.

3.5 Discussion

This study aimed to contribute to the literature on delineating the relationship between early primary pupils' spatial visualization, visuospatial working memory, and accuracy in solving one-step additive word problems in two ways. First, the study aimed to make a theoretical contribution by proposing a conceptual framework for categorizing one-step additive word problems comprising two quantities and one quantitative relation (Thompson,

1993). Two dimensions—quantity (discrete vs. continuous; Davydov, 1982), and quantitative relation (dynamic vs. static; Vergnaud, 1982)—were adapted and integrated to form a 2×2 conceptual framework, resulting in four word problem types: Discrete Quantity–Dynamic Relation, Discrete Quantity–Static Relation, Continuous Quantity–Dynamic Relation, and Continuous Quantity–Static Relation. 24 corresponding one-step additive word problems were developed.

Second, the study aimed to make an empirical contribution by applying the proposed conceptual framework to investigate two substantive research questions, using data from Year 2 pupils in England (6–7-year-olds). The first substantive research question investigated the extent to which spatial visualization and visuospatial working memory predict accuracy in solving one-step additive word problems. The second substantive research question examined whether this relationship is more nuanced such that it varies across different word problem types. To the best of our knowledge, few studies—if any—have examined how multiple spatial skills operate across different quantitative structures in arithmetic word problem solving during the early primary years. This study’s design therefore provides a more nuanced examination of spatial–mathematical link than approaches that aggregate across word problem types. In addition, the use of cross-classified mixed-effects logistic regression appropriately models random effects for both participants and items, thereby accounting for the non-independence of observations arising from each pupil responding to multiple word problems and each problem being completed by multiple pupils.

The discussion proceeds as follows. First, in relation to RQ1 ([Subsection 3.5.1](#)), the associations between spatial visualization, visuospatial working memory, and accuracy in solving word problems are examined. Second, in relation to RQ2 ([Subsection 3.5.2](#)), the moderating role of word problem type in the association between spatial skills and problem-solving accuracy is discussed. Limitations of this study ([Subsection 3.5.3](#)) and directions for future research ([Subsection 3.5.4](#)) are considered.

3.5.1 RQ1: Interpreting the Relationship between Spatial Skills and Word-Problem-Solving Accuracy

Neither spatial visualization nor visuospatial working memory independently predicted word-problem-solving accuracy after controlling for language ability, numerical ability, fluid reasoning, and demographic characteristics (i.e., age and gender). This pattern suggests that the relationship between spatial skills and word-problem-solving accuracy may overlap

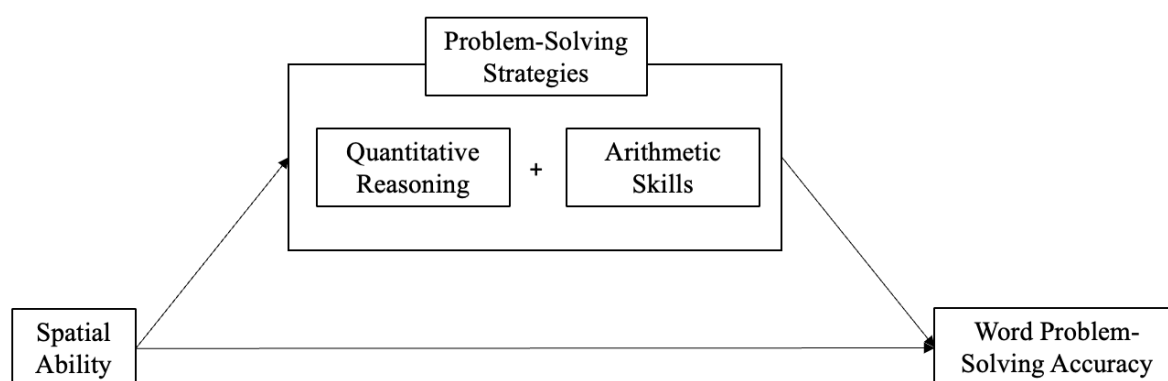
substantially with broader cognitive abilities rather than reflecting a substantial unique contribution of spatial skills to problem-solving accuracy.

Although the present study focused on the association between spatial skills and problem-solving accuracy, prior research suggests that spatial skills may influence mathematical performance indirectly through other cognitive abilities. For example, numerical ability (Fischer & Shaki, 2014; Zhang et al., 2014), fluid reasoning, and language ability (Atit et al., 2022; Kytälä & Lehto, 2008) have all been proposed as potential mediators of the association between spatial ability and mathematical achievement. In the current sample, spatial visualization was weakly to strongly correlated with fluid reasoning, numerical ability, and language ability, whereas visuospatial working memory was moderately correlated with fluid reasoning and numerical ability. These associations suggest that indirect pathways linking spatial skills and word-problem-solving accuracy are plausible. However, although such cognitive abilities may statistically mediate the association between spatial skills and mathematical performance, they provide limited insight into the specific cognitive processes involved in solving arithmetic word problems.

A complementary, more process-oriented perspective involves examining children's problem-solving strategies as potential mediators. Both quantitative reasoning and arithmetic skills are critical for successful word problem solving (Nunes et al., 2009a). Consequently, as illustrated in Figure 3-5, problem-solving strategies that integrate quantitative reasoning and arithmetic processes may help clarify the role of spatial ability in arithmetic word problem solving while accounting for broader cognitive abilities (see [Study 2](#)).

Figure 3-5

Conceptual Model of Indirect Pathways from Spatial Ability to Word-Problem-Solving Accuracy via Problem-Solving Strategies



This perspective may be particularly relevant in young children, for whom the problem-solving process does not necessarily unfold in clearly separated phases of ‘representation’ and ‘computation’ (Carpenter & Moser, 1984; Mayer, 1985; Nunes & Bryant, 1996). Future research should therefore examine children’s problem-solving strategies as a potential mechanism through which spatial ability influences word-problem-solving accuracy, thereby further elucidating whether spatial skills contribute to performance indirectly, beyond language ability, numerical ability, and fluid reasoning (see [Study 3](#)).

In addition to the possibility that spatial skills influence problem-solving accuracy indirectly through other cognitive processes, another potential explanation for the absence of a significant association between spatial skills and problem-solving accuracy is the inclusion of fluid reasoning as a covariate. Fluid reasoning was included because it is typically conceptualized as distinct from spatial ability, despite their strong empirical association. However, some research suggests that commonly used instruments, such as Matrix Reasoning and Block Design, may not always clearly differentiate fluid reasoning from spatial ability (Fuchs et al., 2005; Reynolds & Keith, 2017). Indeed, some scholars argued that Matrix Reasoning partially taps spatial ability (Keith et al., 2006; Sattler, 2001; Sattler & Dumont, 2004). To mitigate this concern, an additional measure, the Visual–Spatial Reasoning task, was included alongside the Block Design subtest to derive a composite measure of spatial visualization (De Pauw et al., 2009; Schenck & Nathan, 2024).

Nevertheless, overlap between spatial visualization and fluid reasoning may persist. An exploratory factor analysis of the Visual–Spatial Reasoning task, the WASI–II Block Design subtest, and the WASI–II Matrix Reasoning subtest yielded a single-factor solution, indicating substantial shared variance among the three measures. However, although Matrix Reasoning loaded on the factor (.60), its loading was lower than those of the two spatial visualization measures (.77 and .84). This pattern suggests that Matrix Reasoning shares variance with spatial visualization while also reflecting broader fluid reasoning processes (Schneider & McGrew, 2012). Future research should further refine the measurement of spatial visualization and fluid reasoning to better delineate their theoretically distinct yet empirically related contributions to mathematical performance.

Taken together, the present findings suggest that spatial visualization and visuospatial working memory do not uniquely predict overall word-problem-solving accuracy once broader cognitive abilities are taken into account. Future studies should examine whether the association between spatial skills and problem-solving accuracy may operate indirectly

through intermediate processes (see [Studies 2](#) and [3](#)). At the same time, it is possible that aggregating performance across word problems with distinct quantitative structure may have obscured more nuanced patterns of association. Accordingly, the next section examines whether the relationship between spatial skills and word-problem-solving accuracy varies as a function of word problem type.

3.5.2 RQ2: Interpreting Interaction Effects of Spatial Skills and Problem Type on Word-Problem-Solving Accuracy

The findings for RQ2 indicated that the relationship between spatial visualization and word-problem-solving accuracy depended on the quantitative structure of the problems. Specifically, the slope of spatial visualization differed between Discrete Quantity–Dynamic Relation and Continuous Quantity–Static Relation problems. This finding provides preliminary support for the proposed 2×2 conceptual framework, suggesting that distinctions in quantity category and quantitative-relation category may meaningfully shape the cognitive demands placed on spatial processing in young children. Consistent with this interpretation, allowing the association between spatial skills and accuracy to vary across problem types significantly improved model fit and reduced unexplained item-level variance. This suggests that some between-item differences reflect systematic differences in the underlying quantitative structure of the problems rather than merely random variation in difficulty. These findings indicate that the relationship between spatial skills and problem-solving accuracy is not uniform but contingent upon quantitative structure.

In contrast to spatial visualization, the association between visuospatial working memory and word-problem-solving accuracy did not vary significantly across word problem types. None of the interaction terms reached statistical significance, suggesting that the contribution of visuospatial working memory may be relatively uniform across structural forms of word problems. The interaction between spatial visualization and problem type in predicting word-problem accuracy is discussed first ([Subsection 3.5.2.1](#)), followed by the corresponding interaction for visuospatial working memory ([Subsection 3.5.2.2](#)).

3.5.2.1 Interaction of Spatial Visualization and Problem Type in Predicting Word-Problem-Solving Accuracy

While the data do not permit conclusions about underlying mechanisms, the significant interaction suggests that the association between spatial visualization and problem-solving accuracy was not uniform across problem types. One possible explanation is that different

quantitative structures place different demands on the representation of quantitative relationships. Spatial visualization has been described as a “mental blackboard” (Hawes & Ansari, 2020, p. 472) on which quantitative structure can be internally represented and manipulated. Previous research with older secondary school pupils has shown that spatial visualization supports the construction and use of visual representations when reasoning about quantitative relations (e.g., Boonen et al., 2013, 2014; Casey et al., 2015; Duffy et al., 2020; Hegarty & Kozhevnikov, 1999; Kozhevnikov & Blazhenkova, 2013; Marschark et al., 2013; Oostermeijer et al., 2014; Van Garderen, 2006). Consequently, the extent to which spatial visualization is associated with word problem solving may depend on the demands that a problem’s quantitative structure places on representation.

In addition to supporting the representation of quantitative structure, spatial visualization may contribute to word problem solving by providing an internal workspace for organizing numerical relations. Missing-term studies (Cheng & Mix, 2014; Gilligan et al., 2020) have shown that even brief spatial visualization training can improve performance on noncanonical arithmetic problems such as $3 + _ = 7$, suggesting that spatial processes may facilitate the internal organization of numerical relationships. If so, spatial visualization may support word problem solving—even when the situation is not explicitly spatial—by enabling children to represent and manipulate numerical relationships spatially, which is consistent with findings with studies with older pupils (e.g., Mix et al., 2016). This aligns with earlier arguments that even discrete-quantities may be spatially structured due to the spatial organization of numerical representations (Dehaene et al., 1993).

An additional speculative possibility concerns the nature of the entities described in the problems. The discrete-quantity problems involved permanent entities (e.g., marbles), whereas the continuous problems described transient entities (e.g., steps taken). It is conceivable that spatial visualization may be more readily engaged when reasoning about stable, manipulable entities with permanence (vs. transience). However, this interpretation remains tentative and warrants further empirical investigation, particularly given that dynamic-relation problems, which describe transformations over time, can also be considered more transient relative to static comparison problems. Thus, Discrete Quantity–Dynamic Relation problems may simultaneously involve permanent entities (i.e., discrete quantity) while conveying transient relations (i.e., dynamic relation).

Importantly, the conditional effects within individual problem types were not statistically significant. Therefore, the present findings do not support strong conclusions

regarding which specific problem types rely most heavily on spatial visualization. Rather, they suggest that the association between spatial visualization and problem-solving accuracy may vary according to the representational demands imposed by different quantitative structures. This interpretation is broadly consistent with the possibility that spatial visualization contributes to word-problem solving through representational processes, a possibility examined more directly in Studies 2 and 3 through analyses of children's problem-solving strategies.

Taken together, the present findings indicate that the association between spatial visualization and problem-solving accuracy is not uniform across problem types. Rather, the role of spatial visualization appears to vary according to the quantitative structure of the problem, consistent with the proposed 2×2 conceptual framework. Although the conditional effects within individual problem types were not statistically significant, the reliable difference between slopes suggests that different quantitative structures may place different demands on the representation and organization of quantitative relationships. These findings underscore the importance of considering quantitative structure when examining the role of spatial skills in arithmetic word problem solving.

3.5.2.2 Interaction of Visuospatial Working Memory and Problem Type in Predicting Word-Problem-Solving Accuracy

The association between visuospatial working memory and word-problem-solving accuracy did not vary significantly across word problem types. This differing pattern between spatial visualization and visuospatial working memory is consistent with theoretical accounts proposing that these spatial skills may serve distinct roles in mathematical thinking (Hawes et al., 2019b; Mix, 2019). One account proposes that visuospatial working memory is primarily engaged in supporting the procedural fluency for *easier, familiar* content, whereas spatial visualization is more strongly recruited for modeling more *difficult, novel* content (Mix, 2019). However, the present findings complicate a simple difficulty- or familiarity-based explanation. Although static-relation problems are generally considered more difficult than dynamic-relation problems (Cummins et al., 1988; Riley & Greeno, 1988), and continuous-quantity problems may be less commonly encountered than discrete-quantity problems in early mathematics instruction, visuospatial working memory did not show differential associations across these problem types with distinct quantitative structures. This suggests that problem difficulty or familiarity alone may not fully explain how visuospatial working memory is engaged in mathematical problem solving.

An alternative interpretation concerns the distinct cognitive functions of the two spatial skills. Spatial visualization may be particularly important for *coordinating* and *transforming* internal representations of quantitative structure, whereas visuospatial working memory may play a more general role in *maintaining* intermediate information during problem solving (Hawes & Ansari, 2020). All word problem types require children to hold quantitative information in working memory, which may account for the relatively stable association observed across word problem types. This distinction between representation coordination and representation maintenance aligns with prior findings linking visuospatial working memory to numerical reasoning more broadly (e.g., Hawes et al., 2019b; Mix et al., 2016, 2017).

Although the interaction between visuospatial working memory and Continuous Quantity–Static Relation problems (relative to Discrete Quantity–Dynamic Relation problems) approached statistical significance, its confidence interval included 1 and should therefore be interpreted with caution. The positive interaction coefficient suggests that the association between visuospatial working memory and problem-solving accuracy may have been somewhat stronger in Continuous Quantity–Static Relation problems, which require maintaining multiple magnitudes relative to the reference point. However, this difference did not reach conventional levels of statistical significance, and further research with greater statistical power is needed to determine whether it reflects a reliable effect.

Taken together, the present findings suggest that, unlike spatial visualization, visuospatial working memory does not appear to vary systematically in its association with word-problem-solving accuracy across quantitative structures. Rather, its contribution may be relatively stable across problem types, consistent with the view that visuospatial working memory primarily supports the maintenance and coordination of quantitative information during problem solving. This pattern reinforces the conceptual distinction between spatial visualization and visuospatial working memory in mathematical word problem solving, while highlighting the need for further research to clarify how these spatial skills interact with quantitative structure.

3.5.3 Limitations

Several limitations should be considered when interpreting the findings and designing future research. The first limitation concerns the interpretation of the interaction effects. Although the relationships between spatial visualization and problem-solving accuracy varied reliably across word problem types, the simple slopes within each problem type were not

statistically significant. Accordingly, these interaction effects should be interpreted with caution.

Although spatial visualization was operationalized through tasks selected to align with its theoretical definition, some degree of overlap with fluid reasoning may remain despite efforts to differentiate these constructs. Such overlap could attenuate the conditional effects observed within specific problem types. Accordingly, the present findings should be interpreted as preliminary pending further refinement of measurement distinctions between spatial visualization and related cognitive abilities.

It is also possible that the conditional effects within individual problem types were subtle. Detecting such nuanced differences often requires larger samples than were available in the present study. Thus, the absence of statistically significant simple slopes should not be interpreted as evidence that spatial visualization plays no role within specific problem types, but rather as an indication that its effects may be modest. Future studies with larger samples are needed to clarify this issue.

The second limitation concerns the word problems. Although they were developed through an iterative process (i.e., two pilot studies and consultation with experts) and rigorously tested (e.g., confirmatory factor analyses), they would benefit from further validation with a larger sample size. This would strengthen confidence in the reliability and validity of both the word problems and the underlying 2×2 conceptual framework used to categorize them. In particular, although continuous-quantity problems were developed based on previous literature and pilot-tested with Year 2 children from England (who were not part of the final dataset), and despite efforts to match them to their discrete quantity counterparts, the findings indicate that continuous-quantity problems were more difficult than discrete-quantity problems. While it is possible that pupils find continuous-quantity problems inherently more difficult than discrete-quantity problems due to limited exposure or differences in cognitive processing (e.g., counting bias), it is also plausible that the matching process was not entirely successful in equalizing the problem types. Testing the problems with a larger sample would help clarify the source of this disparity between problem types.

The third limitation concerns the generalizability of the findings. The study's reliance on a specific geographical area limits the extent to which the results can be generalized to broader populations. On the other hand, because this study included pupils with diverse learning profiles (including SEN and EAL pupils who were able to meaningfully engage with

the tasks), it may have greater ecological validity than studies that are more selective (e.g., those focusing only on typically developing children). Nonetheless, conducting the study with a larger and more geographically diverse sample would help increase the external validity of future research in this domain (Cohen et al., 2018; Punch & Oancea, 2014).

3.5.4 Conclusion

This study contributes valuable insights into the relationship between young children's spatial ability and accuracy in solving one-step additive word problems by proposing and empirically examining a 2×2 conceptual framework that captures structural variation in quantity and quantitative relation. By examining spatial visualization and visuospatial working memory simultaneously, controlling for key cognitive and demographic characteristics (Peng et al., 2020; Verschaffel et al., 2020; Vessonen et al., 2024), and accounting for both between-participant and between-item variability through cross-classified mixed-effects logistic modeling, the study provides a more granular analysis of the spatial–mathematical link.

Although neither spatial visualization nor visuospatial working memory uniquely predicted overall word-problem-solving accuracy, the findings reveal a more nuanced pattern. The association between spatial visualization and accuracy varied as a function of problem structure, consistent with the proposed framework. In contrast, visuospatial working memory did not demonstrate systematic variation across problem types, suggesting a more general role in maintaining quantitative information regardless of structural form.

Together, these findings reinforce the conceptual distinction between spatial visualization and visuospatial working memory and underscore the importance of considering quantitative structure when examining the relationship between spatial ability and early arithmetic word problem solving. Future research should investigate children's problem-solving strategies, including arithmetic strategies and representations of quantitative structure (Carpenter & Moser, 1982; Fuson, 1982; Nunes & Bryant, 2015; Steffe et al., 1988), to better clarify how these spatial skills relate to problem-solving processes. Moreover, examining problem-solving strategy use may help illuminate potential indirect pathways linking spatial skills and word-problem-solving accuracy.

4 Study 2: The Role of Spatial Skills in Young Children's Arithmetic Word-Problem-Solving Strategies

Abstract

A spatial–mathematical link has been widely documented (Hawes & Ansari, 2020; Lourenco et al., 2018; Mix & Cheng, 2012), yet less is known about the association between spatial ability and arithmetic word problem solving in the early primary years. The present study offers an initial contribution to addressing this gap in two ways. Theoretically, it proposes a holistic perspective on problem-solving strategies along two dimensions—arithmetic strategy (counting-based or number-fact) and representation of quantitative structure (correct, incorrect, or none). Empirically, it examines whether spatial visualization and visuospatial working memory are associated with problem-solving strategy use and whether these associations vary by the quantitative structure of word problems. Using data from 137 Year 2 pupils (6–7-year-olds) in England and controlling for language ability, numerical ability, fluid reasoning, age, and gender, cross-classified mixed-effects multinomial models indicated that, across word problem types, higher spatial visualization was consistently associated with greater relative use of '*Quantitatively Correct Number Fact*' strategies compared with '*Quantitatively Incorrect Counting*' and '*Quantitatively Correct Counting*' strategies. In contrast, higher visuospatial working memory was associated with reduced use of '*Non-Quantitative*' strategies across most problem types; for comparison problems involving disjoint quantities measured using an imposed unit, it was instead more consistently associated with reduced relative use of '*Quantitatively Incorrect Counting*' strategies. These findings suggest that the two spatial skills may play distinct yet complementary roles in arithmetic word problem solving and highlight problem-solving strategies as a potential intermediary target for future investigation of the relationship between spatial skills and arithmetic word problem solving.

4.1 Introduction

Solving mathematical word problems, which are problems presented in verbal descriptions and not as number sentences, is an important mathematical skill (OECD, 2023). In particular, arithmetic word problems are a prime tool for teaching both the skills needed in everyday situations, and the general heuristic and metacognitive skills needed in solving difficult mathematical problems (Vicente et al., 2022). Given their importance, arithmetic word problems are a key focus of early mathematics teaching, as reflected in key policy documents such as the national curriculum in England (Department for Education, 2013) and the Common Core State Standards for Mathematics in the United States (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010).

A potential way to understand children’s arithmetic word problem solving is through the lens of spatial ability, given the strong spatial–mathematical link, which is “one of the most robust and well-established findings in cognitive psychology” (Mix & Cheng, 2012, p. 198). *Spatial ability* is a constellation of related skills concerning “the locations of objects, their shapes, their relations to each other, and the paths they take as they move” (Newcombe, 2017, p. 6). Indeed, studies suggest that spatial ability may be one of the problem solver’s characteristics—along with age, gender, language and numerical abilities, and fluid reasoning (Verschaffel et al., 2020)—that make a unique contribution to arithmetic word-problem-solving performance (e.g., Casey et al., 2015; Kozhevnikov & Blazhenkova, 2013; Reinhold et al., 2020; Strohmaier et al., 2022). Furthermore, studies with secondary school pupils suggest that producing visual representations that accurately capture the quantitative structure of the solution-relevant elements is one way in which spatial ability facilitates arithmetic word problem solving (e.g., Boonen et al., 2013, 2014; Hegarty & Kozhevnikov, 1999).

Despite growing evidence linking spatial ability to arithmetic word-problem-solving performance in older pupils, there is limited research of this kind in the lower primary school context (e.g., [Study 1](#); Wong & Ho, 2017). Crucially, there is limited research on the cognitive processes that provide an explanation for the association between spatial ability and problem-solving performance in young pupils. One such process may be the use of problem-solving strategies. This study characterizes *problem-solving strategies* as goal-directed procedures (Lemaire & Siegler, 1995; Schoenfeld, 1985; Tzur et al., 2021) employed by problem solvers to produce a numerical answer to *one-step additive word problems*—problems that describe a part–whole relation and can be solved with a single addition or subtraction calculation (Nunes & Bryant, 2015). These strategies range from formal mathematical heuristics to less structured

response approaches that may not involve engagement with the problem's quantitative structure.

The present study takes a first step toward identifying problem-solving strategies as a potential target for future investigation into the association between lower primary school pupils' (6–7-year-olds) spatial skills and arithmetic problem solving. It makes both theoretical and empirical contributions. Theoretically, it proposes a framework for characterizing young pupils' problem-solving strategies. Empirically, it applies this framework to address two substantive research questions. Specifically, the study examines the distinct contributions of spatial visualization and visuospatial working memory to young pupils' problem-solving strategies in one-step additive word problems (RQ3). It also investigates whether the associations between these spatial skills and problem-solving strategies vary as a function of the quantitative structure of the problems (RQ4).

This study examines spatial visualization and visuospatial working memory because these two spatial skills have consistently been identified as having a particularly strong association with mathematical thinking (for a review, see Allen et al., 2019; Hawes & Ansari, 2020). *Spatial visualization* is “the ability to generate, recall, maintain, and manipulate visual-spatial images and solutions” (Hawes & Ansari, 2020, p. 466), whereas *visuospatial working memory* is the ability to temporarily maintain and process visuospatial information (Baddeley & Hitch, 1974). Prior research indicates that these two spatial skills may play distinct roles in mathematical development and thinking (e.g., Hawes et al., 2019b; Mix et al., 2016, 2017, 2019), although the findings have not been consistent (e.g., Hegarty & Waller, 2005; Mix et al., 2021; Nath & Szűcs, 2014; Reuhkala, 2001).

4.2 Literature Review

Despite the breadth of literature on the spatial–mathematical link (for a review, see Hawes & Ansari, 2020; Lourenco et al., 2018), little research has examined the relationship between lower primary school pupils' (6–7-year-olds') spatial ability and problem-solving strategies for arithmetic word problems. Most existing studies have focused on secondary school pupils and predominantly on one form of problem-solving strategy: visual representation. The first part of this literature review outlines the types of visual representations produced by secondary school pupils when solving word problems ([Subsection 4.2.1](#)). It then discusses how different types of visual representations can either facilitate or hinder problem

solving and how spatial ability may be associated with the production of supportive visual representations.

Given that the production of visual representations is uncommon among lower primary school pupils, the second part reviews literature to identify the problem-solving strategies used by children in this age group when solving one-step additive word problems, with the aim of developing a holistic framework for categorizing them ([Subsection 4.2.2](#)). The final part of the literature review considers how the relationship between spatial skills and problem-solving strategies may vary by word problem type ([Subsection 4.2.3](#)). It therefore adopts the 2×2 conceptual framework from Study 1 for categorizing one-step additive word problems, providing a basis for examining the potential moderating role of the problems' quantitative structure. Finally, the study's research questions are presented ([Subsection 4.2.4](#)).

4.2.1 Spatial Ability and Visual Representations in Secondary School Pupils

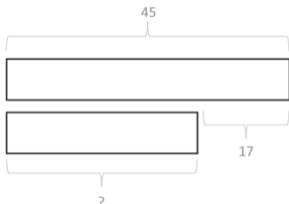
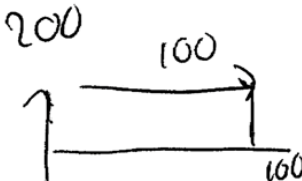
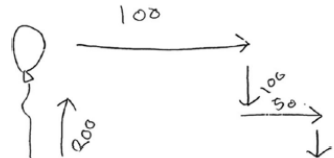
One of the suggested ways in which spatial ability facilitates arithmetic word problem solving in secondary school pupils is through the production of visual representations (e.g., Boonen et al., 2013, 2014; de Koning et al., 2022; Hegarty et al., 1995; Ng & Lee, 2009; Van Garderen, 2006). Studies have specifically focused on visual representations taking the form of *external* expressions such as gestures, drawings, and the use of manipulatives. In particular, the link between spatial ability and the production of visual representations has been explored specifically in the context of solving *visual-spatial* word problems, which are problems that can be solved “relatively easily by visual-pictorial means if the correlation of the given elements in the problem is expressed visually” (Krutetskii, 1976, p. 156). These differ from *verbal-analytical* word problems that can be solved using logical and algebraic reasoning without relying on the visualization of relations. (Further details regarding the distinction between these two word problem types are provided in [Subsection 3.2.2](#) of Study 1.)

As summarized in Table 4-1, the visual representations have been categorized into two types: iconic representation and schematic representation. *Iconic representations* focus on the vivid and detailed visual appearance of the objects and persons (e.g., shape, color, and texture) described in the problem (Aleman et al., 2005; Blazhenkova & Pechenkova, 2019; Farah et al., 1988; Guarnera et al., 2019; Hegarty & Kozhevnikov, 1999; Kozhevnikov et al., 2005; Levine et al., 1985). Studies found that the production of iconic representations is negatively related to word-problem-solving performance (Ahmad et al., 2010; Hegarty & Kozhevnikov, 1999;

Kozhevnikov et al., 2002; Krawec, 2014; Van Garderen, 2006; Van Garderen & Montague, 2003).

Table 4-1

Examples of Visual Representations

Iconic Representation	Schematic Representation	
	Inaccurate	Accurate
Problem 1: “At the Bijenkorf, jeans cost 45 euros. That is 17 euro less than at C&A. How much will you pay at C&A?” (de Koning et al., 2022, p. 31) ^a		
		
Problem 2: “A balloon first rose 200 meters from the ground, then moved 100 meters to the east, then dropped 100 meters. It then travelled 50 meters to the east, and finally dropped straight to the ground, how far was the balloon from its original starting point?” (Hegarty & Kozhevnikov, 1999, p. 680) ^b		
		

Note. ^aAdapted from “Model method drawing acts as a double-edged sword for solving inconsistent word problems,” by B. B. de Koning, A. J. H. Boonen, J. Jongerling, F. van Wesel, & M. van der Schoot, 2022, *Educational Studies in Mathematics*, 111(1), p. 32 & p. 34 (<https://doi.org/10.1007/s10649-022-10150-8>). Copyright 2022 by the authors. *Reprinted with permission.*

^bAdapted from “The role of visual representation type, spatial ability, and reading comprehension in word problem solving: An item-level analysis in elementary school children,” by A. J. H. Boonen, F. van Wesel, J. Jolles, & M. van der Schoot, 2014, *International Journal of Educational Research*, 68, p. 32 & p. 18-19 (<https://doi.org/10.1016/j.ijer.2014.08.001>). Copyright 2014 by Elsevier Ltd. *Reprinted with permission.*

In contrast, *schematic representations* focus on the quantitative relations between the solution-relevant elements (Boonen et al., 2014). They can either be an accurate or an inaccurate portrayal of the quantitative relations. As can be seen in the examples in Table 4-1, accurate schematic representations may facilitate problem solving by correctly reducing the situation in the problem to its “abstract mathematical gist” (Wong, 2021, p. 20). On the other hand, inaccurate schematic representations may be detrimental for solving problems, particularly so for more difficult word problems, such that when an inaccurate schematic representation is used, word-problem-solving success is close to zero (2.60%) (de Koning et

al., 2022). In essence, accurate schematic representations facilitate problem solving whereas iconic representations and inaccurate schematic representations hinder problem solving.

Spatial ability assists with the production of accurate schematic representations. For example, Boonen et al. (2014) found in a study with 11-year-olds that spatial ability, as measured by the Paper Folding task and the Picture Rotation task, predicted word-problem-solving performance ($r = .59$, $p < .001$), and the production of accurate schematic representations ($r = .31$, $p < .001$). Moreover, they found that spatial ability was the most powerful predictor of problem-solving success when an accurate schematic representation was produced (Odds Ratio = 2.22) but was the lowest predictor when iconic representation was produced (Odds Ratio = 1.13). Through a path analysis, Boonen et al. (2013) found that 21% of the relationship between spatial ability and word-problem-solving success was explained by the production of schematic representations.

4.2.2 Spatial Ability and Additive Word-Problem-Solving Strategies in Lower Primary School Pupils

Investigating the relationship between spatial ability and one-step additive word problem solving in younger pupils requires a different focus from that adopted for older pupils. First, young pupils rarely produce visual representations spontaneously when solving word problems (Jitendra & Woodward, 2019). Second, even before formal mathematics instruction, young pupils can solve word problems using informal methods, such that the problem-solving process does not necessarily involve sequential phases of ‘representation’ and ‘computation,’ (Carpenter & Moser, 1984; Nunes & Bryant, 1996) as is often assumed for older pupils and adults (Mayer, 1985). Therefore, rather than focusing narrowly on visual representation, research on the role of spatial ability in young pupils’ problem solving should adopt a more holistic perspective that considers the range of strategies used at this age. This can be achieved by drawing on research indicating that both arithmetic skills ([Subsection 4.2.2.1](#)) and quantitative reasoning ([Subsection 4.2.2.2](#)) are important for solving mathematical word problems (Nunes et al., 2012). The literature on these two components of problem solving is reviewed in turn. This review informs a proposed holistic framework for problem-solving strategies that integrates the dimensions of (1) arithmetic strategy and (2) representation of quantitative structure ([Subsection 4.2.2.3](#)).

4.2.2.1 Arithmetic Strategies: Counting-Based versus Number-Fact Strategies

Younger children can solve arithmetic word problems using informal methods of direct modeling (e.g., pretend play) and counting (Carpenter & Moser, 1984; Nunes & Bryant, 1996), well before formal instruction in the corresponding mathematical knowledge and skills (Verschaffel et al., 2020). In other words, young children can begin to solve additive word problems without necessarily understanding the arithmetic operation (i.e., either addition or subtraction) associated with the problem (Carpenter & Moser, 1982). With development, children tend to move from externalized, typically less efficient, more error-prone, and laborious *counting-based* strategies that are often “direct reflection or simulation of the problem’s semantics” (Verschaffel et al., 2020, p. 4) through more economic (Carpenter & Moser, 1982) and non-perceptual (Steffe et al., 1982) counting-based strategies, and eventually to more efficient, flexible, and abbreviated *number-fact* strategies (Verschaffel et al., 2020).

As summarized in Table 4-2, there are various counting-based strategies that young pupils can use to solve one-step additive word problems (Carpenter & Moser, 1982; Fuson, 1982). Count-all and count-on are strategies used to find the sum of two entities. Count-up, separate-from, count-down, and matching are strategies used to find the difference of two entities. These counting strategies can be executed using representational tools such as concrete objects (e.g., fingers, counters, manipulatives), figural objects (e.g., nodding or pointing in the air, eye-tracking, and gesturing), marks or drawing on paper, or mentally through vocal or subvocal counting (Carpenter & Moser, 1982; Steffe et al., 1982).

Counting-based strategies can be used to either model the problem situation directly or to keep track of the counting steps (Carpenter & Moser, 1984). For example, if a problem presents the initial quantity and the change to the quantity, and asks for the final quantity, children can directly enact this problem situation (e.g., either gaining or losing objects) to solve the problem. In contrast, if a problem presents the initial and the final quantities, and asks for the change in the quantity, the representational tools can be used to keep track of the counting steps between the two quantities.

One of the major mathematical accomplishments during early primary school is gradually moving away from inefficient and laborious counting-based strategies toward more efficient and less effortful counting-based strategies (Carr & Alexeev, 2011). For example, children move away from a count-all strategy, which requires them to count both quantities and is more time-consuming and error-prone, toward a count-on strategy, which requires

counting forward from one of the quantities in the problem. Even within count-on strategies, children can move away from a count-on-from-first strategy to a count-on-from-larger strategy. In the former, the child counts forward from the first quantity presented in the problem, whereas in the latter, the child begins counting forward from the larger of the two quantities, thereby requiring fewer counting steps (Carpenter & Moser, 1982).

Table 4-2*Counting-Based Arithmetic Strategies*

	Description	Example
Count-All	Counting from one to determine the total of both numbers	What is two plus six? ' <u>One, two, three, four, five, six, seven, eight.</u> The answer is eight.'
Count-On-From-First	Counting on from the first number and continuing for as many steps as indicated by the second number	What is two plus six? ' <i>Two</i> , <u>three, four, five, six, seven, eight.</u> The answer is eight.'
Count-On-From-Larger	Counting on from the larger number and continuing for as many steps as indicated by the smaller number	What is two plus six? ' <i>Six</i> , <u>seven, eight.</u> The answer is eight.'
Count-Up	Counting up from the smaller number until reaching the larger number	What is six minus two? ' <i>Two</i> , <u>three, four, five, six.</u> The answer is four.'
Separate-From	Counting the remaining units after taking away units corresponding to the smaller number from the larger number	What is six minus two? (Constructs a set of six. Then takes two away, and counts the remaining units.) ' <u>One, two, three, four.</u> The answer is four.'
Count-Down	Counting backward from the larger number, producing as many counting number words as indicated by the smaller number	What is six minus two? ' <i>Six</i> , <u>five, four.</u> The answer is four.'
Matching	Pairing objects from two groups one-to-one to determine the difference between the two numbers	What is six minus two? (Constructs a set of six and a set of two. Pairs the two sets. Counts the unmatched counters.) ' <u>One, two, three, four.</u> The answer is four.'

Note. Adapted from Table 2-10 (Chapter 2). Italics indicate the starting number; underlining indicates the counting steps.

Moreover, with experience and practice, children will progress away from counting-based strategies altogether toward number-fact strategies that are more advanced, internalized, faster, accurate, and flexible (Verschaffel et al., 2020). As summarized in Table 4-3, there are two types of number-fact strategies: recalled and derived (Carpenter & Moser, 1982; Fuson,

1982). A child can recall a number fact if they have it committed to their memory (for an alternative account for those with small operands (≤ 4), see Barrouillet & Thevenot, 2013). If a number fact cannot be recalled directly (e.g., $6 + 8 = ?$), the child may instead decompose the problem into number facts they do have committed to memory (e.g., $6 + 6 = 12$) and use these to derive the final answer (e.g., two more than 12 is 14). Such a derived number-fact strategy is more common than recalled number-fact strategies for complex problems, such as mixed-digit (e.g., $13 + 8$) and double-digit problems (e.g., $12 + 23$) (Casey et al., 2017).

Table 4-3*Number-Fact Arithmetic Strategies*

	Description	Example
Recalled Number Fact	Retrieve the answer directly from memorized number facts	What is six plus eight? 'Fourteen. I know by heart.'
Derived Number Fact	Derive the answer by decomposing the problem into known number facts	What is six plus eight? 'Six plus six is twelve. So, six add eight is just two more than twelve. The answer is fourteen.'

The number-fact strategies are generally considered more advanced and sophisticated than counting-based strategies for several reasons. First, as stated above, number-fact strategies are more efficient because they require less time to arrive at the numerical solution than do counting strategies (Ashcraft & Fierman, 1982). Second, number-fact strategies draw on memory-based mental representations of numerical information and require applying prior knowledge of mathematical facts (Geary, 2011). In addition, derived number-fact strategies require active problem solving and a more complex series of reasoning processes, as they involve more than just rote procedures (Casey & Fell, 2018).

It is important to note that children who have access to multiple arithmetic strategies may adaptively use them (Siegler, 2005). That is, within the same session, they may use a mix of counting-based and number-fact strategies when solving arithmetic tasks (Foley et al., 2017). For example, depending on the characteristics of the problems, a child might use fingers to carry out a count-on strategy for one problem (e.g., $8 + 3$, as it does not require many counting steps), derive a number fact for the next (e.g., $8 + 9$, which can be solved using $8 + 8 + 1$), and then use counters to carry out a separate-from strategy for another (e.g., $13 - 5$, as it can be readily represented using counters but may be more difficult to recall) (Schiffman & Laski, 2018).

Although children can use a mix of counting-based and number-fact strategies, the relative frequency of these strategies has many implications. First, number-fact strategies are associated with greater computational success than counting-based strategies (Geary, 2011), especially after approximately age 6–8 years (Jordan et al., 2008; Krenger & Thevenot, 2025). Second, cross-sectionally, pupils who frequently use number-fact strategies tend to have higher overall mathematical achievement, beyond arithmetic fluency, than those who rely on counting-based strategies (Carr et al., 2008). Finally, longitudinally, the use of number-fact strategies in lower primary school years predicts not only later performance in arithmetic word problem solving but overall mathematical achievement in later school years (Carr & Alexeev, 2011).

While research on young pupils' spatial ability and strategy use in arithmetic word problem solving is limited, an emerging body of literature has examined links between spatial ability and *arithmetic strategies* in addition and subtraction computation tasks (for a review, see Casey & Fell, 2018; Laski et al., 2013; Casey et al., 2017; Möhring et al., 2024; Vasilyeva et al., 2025). This literature suggests that spatial ability positively predicts the frequency of number-fact strategy use and negatively predicts the frequency of counting-based strategy use. Spatial visualization and visuospatial working memory have individually received attention in this area of research, although they have not been studied together in a single study.

For example, in a study with 127 first grade girls (6–7-year-olds), Laski et al. (2013) found that spatial visualization—as measured by the WISC–IV Block Design subtest, 2-D mental rotation and 3-D mental rotation tasks—in addition to predicting the overall accuracy, also positively related to the frequency of derived and recall number-fact strategies, negatively related to the count-all strategies, and unrelated to the count-on strategies when solving additive computation tasks. Furthermore, while both spatial and verbal skills predicted the frequency with which children used the derived number-fact strategies, only spatial visualization predicted the frequency of the recalled number-fact strategies.

In another study, Foley et al. (2017) found that in second and fourth grade children (7–8-year-olds, and 9–10-year-olds), *simultaneous* visuospatial working memory (i.e., remembering spatial patterns presented at the same time rather than in sequence; Allen et al., 2019), as measured by a visual pattern task, predicted the frequency of derived number-fact strategy use when solving additive computation tasks. They also found that the frequency of derived number-fact strategy use mediated the relationship between visuospatial working

memory and arithmetic performance. Interestingly, while verbal working memory also predicted the arithmetic performance, it did not predict the arithmetic strategy use.

The relationship between spatial ability and number-fact strategies in young pupils may stem from humans' predisposition to associate numbers with space (for a review, see Cipora et al., 2018). Spatial skills contribute critically to the early development of numerical reasoning by supporting the acquisition of a meaningful mental representation of numbers—specifically, the linear number line (Gunderson et al., 2012). Consequently, young pupils with higher spatial skills may be better able to map numbers onto a mental number line and recognize their relative spatial positions, which in turn enables them to use number-fact strategies. In contrast, pupils with weaker spatial skills may struggle to mentally manipulate numerical magnitudes and therefore rely more heavily on counting-based strategies (Laski et al., 2013). Moreover, spatial skills may not only support the visualization of numerical magnitudes but also enable pupils to flexibly segment and transform large magnitudes into smaller parts (Möhring et al., 2024) and maintain intermediate solutions and procedures in mind (Foley et al., 2017), both of which facilitate number-fact strategies.

4.2.2.2 Quantitative Reasoning: Representation of Quantitative Structure

In addition to the distinction between counting-based and number-fact arithmetic strategies, another important distinction in young pupils' word-problem-solving strategies is their representation of the problem's quantitative structure, which is reflective of their quantitative reasoning. Unlike computation tasks in which the problems are presented in number sentences, when solving word problems, children need to “represent quantitative information and act on the representations to come to conclusions not previously known” (Nunes et al., 2015, p. 178). In other words, solving one-step additive word problems involves representing the quantitative structure. Quantitative structure can be defined as a network of *quantities* (measurable intrinsic properties, e.g., set size, position) and *quantitative relations* (connections between quantities, e.g., changes or comparisons of set size/position) (Dos Santos, 2022; Nunes et al., 2009a; Thompson, 1993). (For more details on quantitative structure, see [Subsection 1.3](#).) Correct representation of the quantitative structure increases the likelihood of obtaining the correct solution, whereas incorrect representation, or failing to consider the quantitative structure, decreases that likelihood (e.g., Boonen et al., 2014).

Akin to the categorization of visual representations presented in [Subsection 4.2.1](#), representations of quantitative structure can be classified into three types (see Table 4-4). When

problem solvers use a ‘*Non-Quantitative*’ strategy, engagement with the quantitative structure is not evident. Nevertheless, these strategies remain goal-directed, involving superficial tactics such as restating a number from the problem, random guessing, or producing an irrelevant or iconic representation. In contrast, when problem solvers use either *quantitatively-incorrect* or *quantitatively-correct* strategies, they do engage with the quantitative structure of the problem. However, the difference between the two is that the former is an inaccurate representation of the quantitative structure (akin to inaccurate visual representation), whereas the latter is an accurate representation of the quantitative structure (akin to accurate visual representation).

Table 4-4*Representations of Quantitative Structure*

Description	
Non-Quantitative	Does not represent the quantitative structure
Quantitatively-Incorrect	Inaccurately represents the quantitative structure
Quantitatively-Correct	Accurately represents the quantitative structure

Importantly, the correctness of the quantitative structure representation is independent of the accuracy of the final numerical solution. Consider this additive problem: ‘Ann had five marbles and won three more marbles. How many marbles does Ann have now?’ A child using a quantitatively-correct strategy would identify that the final quantity, the unknown of interest, must be greater than the initial quantity. They might carry out a counting strategy using their fingers (e.g., counting on from five). However, while applying the quantitatively-correct strategy, they may make an execution error—for example, folding the first finger while saying ‘five,’ then stopping at the third finger and concluding the answer is seven—thus arriving at an erroneous numerical conclusion. Then in the next problem, ‘Ann had five marbles and won four more marbles. How many marbles does Ann have now?’ the child might use a derived number-fact strategy to say that since three more than five was seven in the previous question, four more than five must be one more than seven, which is eight. In both problems, the child accurately represented the quantitative structure, though they did not arrive at the correct numerical solution.

Although it is improbable that the use of a non-quantitative or quantitatively-incorrect representation will result in the correct numerical solution, it is not impossible. However, what is important to note is that arriving at the correct numerical solution following a quantitatively-correct representation is qualitatively different than arriving at the correct numerical solution despite using a non-quantitative or quantitatively-incorrect representation. Thus, the

investigation of the correctness of the numerical solution is not sufficient to determine the qualitative differences, since it alone cannot indicate whether the problem solver used a non-quantitative, quantitatively-incorrect, or quantitatively-correct representation. In other words, the assessment of the representation of quantitative structure cannot be conflated with the problem-solving accuracy.

While there is limited research on the association between young pupils' spatial skills and their representation of quantitative structure, a potential link can be inferred from studies with older pupils, as outlined above in [Subsection 4.2.1](#). In older pupils, a key aspect of spatial skill in solving word problems is the production of accurate schematic representations, which support reasoning about the problem's quantitative structure (Casey & Fell, 2018). By extension, when young pupils solve one-step additive word problems, their spatial skills may facilitate the representation of the quantitative structure (Thevenot, 2010). More specifically, spatial visualization may facilitate the representation by providing a "mental blackboard" on which quantitative structure can be coordinated and transformed, whereas visuospatial working memory may support the temporary maintenance of this representation during problem solving (Hawes & Ansari, 2020, p. 472).

4.2.2.3 Problem-Solving Strategies: A Holistic Perspective

Based on reviewed literature on young pupils' (1) arithmetic strategies (counting-based versus number-fact approaches) and (2) representations of quantitative structure (i.e., correctly or incorrectly representing the quantitative structure, if at all), their problem-solving strategies for solving one-step additive word problems can be synthesized as shown in Table 4-5. This proposed problem-solving framework offers a more holistic perspective by integrating arithmetic strategy use with the representation of quantitative structure, two dimensions that have largely been examined separately in prior spatial–mathematical research. Moreover, it reflects the idea that young pupils' problem-solving processes do not necessarily unfold in sequential phases of 'representation' and 'computation,' (Carpenter & Moser, 1984; Nunes & Bryant, 1996) as is often assumed for older pupils and adults (Mayer, 1985).

As discussed above, prior research indicates that spatial ability is associated with the use of number-fact strategies, although this work has primarily focused on computation tasks (e.g., Casey et al., 2017; Foley et al., 2017; Laski et al., 2013; Möhring et al., 2024; Vasilyeva et al., 2025). It is also associated with the accurate representation of quantitative structure, although the research has primarily focused on secondary school pupils (e.g., Boonen et al.,

2013, 2014). Taken together, the reviewed literature suggests that young primary school pupils' spatial ability may be associated with a greater likelihood of using '*Quantitatively Correct Number Fact*' strategies, which involve correctly representing the quantitative structure of the problem while using a number-fact strategy. However, such research has yet to be conducted.

Table 4-5

Problem-Solving Strategies for One-Step Additive Word Problems

Strategy	Definition
Non-Quantitative	Does not appear to actively engage with the problem's quantitative structure (e.g., guessing without attempting; producing an iconic visual representation)
Quantitatively Incorrect Counting	Represents the problem's quantitative structure incorrectly while using a counting-based strategy (e.g., counting-all for a subtraction problem; using a counting strategy based on an inaccurate schematic representation)
Quantitatively Incorrect Number Fact	Represents the problem's quantitative structure incorrectly while using a number-fact strategy (e.g., mentally subtracting for an addition problem; using a number-fact strategy based on an inaccurate schematic representation)
Quantitatively Correct Counting	Represents the problem's quantitative structure correctly while using a counting-based strategy (e.g., separate-from for a subtraction problem; using a counting strategy based on an accurate schematic representation)
Quantitatively Correct Number Fact	Represents the problem's quantitative structure correctly while using a number-fact strategy (e.g., mentally adding for an addition problem; using a number-fact strategy based on an accurate schematic representation)

Note. Reproduced from Table 1-3 (Chapter 1).

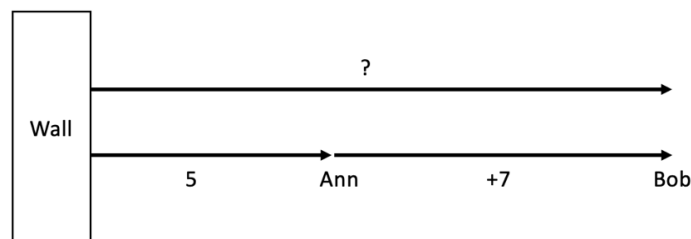
4.2.3 Word Problem Type as a Moderator of the Spatial–Strategy Relationship

The relationship between young pupils' spatial skills and their use of problem-solving strategies may depend on the quantitative structure of the word problem. For example, consider the following problem: 'Ann is 5 steps away from the wall. Bob is 7 steps farther from the wall than Ann. How far is Bob from the wall?' As depicted in Figure 4-1, although the defining feature of this problem type is the use of continuous quantities measured through imposed units, the present problem is also embedded in a spatial context involving positions relative to a reference point (akin to visual–spatial word problems described in [Subsection 4.2.1](#)). Consequently, spatial skills may facilitate the representation of the problem's quantitative structure. Moreover, because the quantitative structure maps directly onto the problem's

situation, children may be able to model the problem using concrete representational tools (e.g., fingers or counters). As a result, they may be able to rely on counting-based strategies to generate the numerical solution even if they do not yet have the sum of 5 and 7 committed to memory (Siegler, 2005).

Figure 4-1

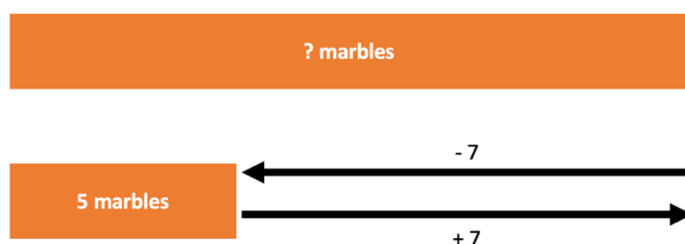
Example of a Quantitative Structure Amenable to Direct Modeling



In contrast, consider the following problem: ‘Ann had some marbles. She played a game of marbles and lost 7 marbles. Now she has 5 marbles. How many marbles did she have to begin with?’ As shown in Figure 4-2, the quantitative structure of this problem comprises two quantities (represented as bars) and one quantitative relation linking them (represented as a leftward-pointing arrow). Although the defining feature of this problem type is the use of discrete quantities, a crucial aspect of solving it involves constructing a representation that accounts for the *inverse* of the given quantitative relation (Dos Santos, 2022)—that is, the fact that Ann has seven fewer marbles now can be represented as her having seven more marbles at the beginning (a rightward-pointing arrow). Spatial visualization may support such coordination and transformation of representations of a problem’s quantitative structure; visuospatial working memory may, in turn, facilitate maintaining this representation during the quantitative reasoning process (Hawes & Ansari, 2020). Unlike the previous example, this problem may be difficult to model directly using concrete representations because the initial quantity is the unknown (Carpenter & Moser, 1984; Riley et al., 1983).

Figure 4-2

Example of a Quantitative Structure involving the Inverse of a Quantitative Relation

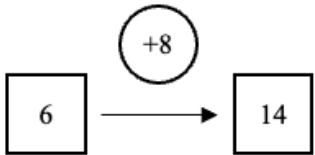
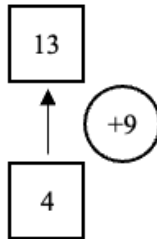
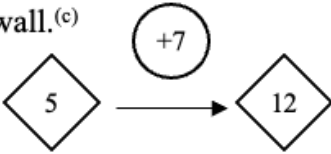
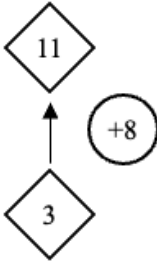


As the two examples above illustrate, different word problems may require the same operation (i.e., both can be solved by finding the sum of 5 and 7), yet differ in their quantitative structure and thus in the demands they place on its representation. Therefore, to gain a more nuanced understanding of the association between spatial skills and problem-solving strategies, it is important to examine whether this association varies across word problem types that differ in quantitative structure. Such investigation may be facilitated by the 2×2 conceptual framework (Quantity $\times$ Quantitative Relation) for categorizing one-step additive word problems proposed in Study 1 (see Figure 4-3).

This conceptual framework builds on Thompson's (1993) distinction between quantity and quantitative relation by integrating methods of quantification (Davydov, 1982) and classification of quantitative relations (Vergnaud, 1982). The resulting framework categorizes word problem types along two binary dimensions: (1) *Quantity*, and (2) *Quantitative Relation*. First, if quantities are specific and fixed, measured through one-to-one counting (e.g., set size), they are *discrete*; if they are measured using an imposed unit (e.g., position), they are *continuous*. Second, if quantitative relations describe a transformation in quantity over time, they are *dynamic*; if they describe a comparison between disjoint quantities, they are *static*. As summarized in Figure 4-3, this 2×2 conceptual framework yields four word problem types: *Discrete Quantity–Dynamic Relation*, *Discrete Quantity–Static Relation*, *Continuous Quantity–Dynamic Relation*, and *Continuous Quantity–Static Relation*.

Consider first the potential moderating effect of the dimension of quantity, drawing on the well-established link between spatial skills and problem solving in geometry and measurement (Casey & Fell, 2018; Vasilyeva et al., 2009). Developing competence in geometry and measurement requires understanding the spatial nature of these domains (e.g., measuring area and volume involves enumerating arrays of squares and cubes; Battista, 2003). By extension, one hypothesis is that the relationship between spatial skills and 'Quantitatively Correct Number Fact' strategy use may be stronger for continuous-quantity problems, as these resemble geometry and measurement problems in that regard. Alternatively, the relationship may also be strong for discrete-quantity problems if successful performance depends on representing the quantitative structure, thereby drawing more heavily on spatial skills.

Figure 4-3*Conceptual Framework for One-Step Additive Word Problems*

		Quantitative Relation	
		Dynamic Relation	Static Relation
Quantity	Discrete Quantity	Ann had 6 marbles.^(a) Ann won 8 more marbles.^(b) Ann now has 14 marbles.^(c) 	Ann has 4 marbles.^(a) Bob has 9 more marbles than Ann.^(b) Bob has 13 marbles.^(c) 
	Continuous Quantity	Ann was 5 steps from the wall.^(a) Ann walked 7 more steps away from the wall.^(b) Ann is now 12 steps from the wall.^(c) 	Ann is 3 steps from the wall.^(a) Bob is 8 steps farther from the wall than Ann.^(b) Bob is 11 steps from the wall.^(c) 

Note. When each word problem type (Discrete Quantity–Dynamic Relation, Discrete Quantity–Static Relation, Continuous Quantity–Dynamic Relation, and Continuous Quantity–Static Relation) is framed as a one-step additive word problem, two of the three facts are provided, and the missing third fact becomes the unknown to be solved. The unknown is either a quantity (a or c) or a quantitative relation (b). Upright squares ($\square$) represent discrete quantities. Tilted squares ($\diamond$) represent continuous quantities. Circles ($\circ$) represent quantitative relations. Dynamic relations are represented horizontally, whereas static relations are represented vertically.

Similarly, the relationship between spatial skills and ‘Quantitatively Correct Number Fact’ strategy use may differ across the dimension of quantitative relation. On the one hand, the association may be stronger for dynamic-relation problems, as solving these problems often requires mentally transforming representations of quantitative structure—a process that spatial skills are likely to facilitate (Mix & Cheng, 2012). On the other hand, the association may be stronger for static-relation problems, as these problems may be more difficult to represent (Riley et al., 1983), thereby increasing reliance on spatial skills that support the representation of quantitative structure (Mix, 2019).

Taken together, the reviewed literature suggests that problem-solving strategies along the dimensions of arithmetic strategy and representation of quantitative structure may provide a more holistic perspective on how young pupils solve one-step additive word problems. There is a need to examine the relationship between spatial ability and problem-solving strategies, and whether this relationship is moderated by the quantitative structure of the problem, defined by the dimensions of quantity and quantitative relation. The present study aims to examine this relationship empirically.

4.2.4 Present Study

The present study aims to investigate whether young pupils’ spatial visualization and visuospatial working memory are associated with their use of problem-solving strategies in one-step additive word problems. It further examines whether these associations vary across word problem types, thereby clarifying how the relationship between spatial ability and problem-solving strategies depends on quantitative structure. To this end, the study addresses the following research questions:

RQ3) To what extent do spatial visualization and visuospatial working memory predict young primary school pupils’ (6–7-year-olds’) problem-solving strategy use?

RQ4) Do the relationships between spatial visualization, visuospatial working memory, and problem-solving strategy use vary across different word problem types?

The present study extends previous research on the relationship between spatial ability and problem-solving strategies in several ways. First, it focuses on lower primary school pupils rather than secondary school pupils. Second, it examines strategy use in one-step additive word problems rather than arithmetic computation tasks. Third, it investigates the unique associations of both spatial visualization and visuospatial working memory while controlling for other factors known to contribute to arithmetic problem solving, such as language ability,

numerical ability, fluid reasoning, age, and gender (e.g., Daroczy et al., 2015; Fung & Swanson, 2017; Verschaffel et al., 2020; Vessonen et al., 2024).

4.3 Methods

This study employed a correlational, cross-sectional design to investigate the relationship between spatial visualization, visuospatial working memory, and the use of problem-solving strategies in one-step additive word problems. The following section describes the participants ([Subsection 4.3.1](#)) and measures ([Subsection 4.3.2](#)). Next, it outlines the data collection process, including task administration procedure ([Subsection 4.3.3](#)). Finally, it details the coding of problem-solving strategies ([Subsection 4.3.4](#)) and the analytic approach ([Subsection 4.3.5](#)) used to address the research questions.

4.3.1 Participants

A total of 149 Year 2 pupils in South East England were recruited using a non-probabilistic convenience sampling method. The testing of 12 participants discontinued either because they could not understand the tasks or because they expressed a desire to stop. The analysis uses the data of 137 pupils (50.4% female, 49.6% male; age range: 6.52–7.74 years; mean age: 7.10 years) who completed all seven tasks. Although Special Education Needs (SEN) and English as an Additional Language (EAL) data were not available for all participants, 27 (19.7%) of the retained sample were reported as having SEN and 21 (15.3%) as having EAL, meaning the true proportions may be higher.

Schools were recruited first, then the pupils were recruited through either an opt-out or opt-in parental consent approach, as decided by the school. This research has received university ethics approval and followed the ethical guidelines for educational research as recommended by the British Educational Research Association (2018) and the University of Oxford Policy on Research Involving Human Participants and Data Protection (Central University Research Ethics Committee, 2022).

4.3.2 Measures

The seven measures used in the present study are summarized in Table 4-6. Spatial visualization was assessed using two measures: the Wechsler Abbreviated Scale of Intelligence–Second Edition (WASI–II) Block Design subtest (Atit et al., 2022) and the Visual–Spatial Reasoning task (Hawes et al., 2015, 2017, 2019). In the Block Design subtest, participants used red-and-white cubes to recreate target designs shown in the stimulus book within a specified time limit. The Visual–Spatial Reasoning task comprised four subtasks that

Table 4-6*Study Measures*

Variable	Task	Description	Scoring
Spatial Visualization	Wechsler Abbreviated Scale of Intelligence–Second Edition (WASI–II) Block Design subtest	• Within a specified time limit, recreate a design presented in the Stimulus Book using two, four, or nine red-and-white cubes • A total of 11 items for children aged 6 to 8, but testing discontinued after two consecutive scores of zero	• Scores depend on both accuracy and speed of construction, and can range from 0 to 2 for the first four items, and 0 to 7 for the final seven items • The total raw score is the sum of all item scores, and can range from 0 to 57 for children aged between 6 to 8
	Visual–Spatial Reasoning task	• Reason and visualize solutions to problems involving parts of objects in relation to their whole (i.e., missing center, missing border, shape composition/decomposition, and paper folding) • A total of 20 items	• Each item is scored as either correct (1) or incorrect (0) • The total raw score is the sum of all item scores, and can range from 0 to 20
Visuospatial Working Memory	PathSpan	• Replicate a sequence of dots that light up one at a time by tapping them in the same order • The sequence length begins at two and increases by one if successful on either of the two trials in the previous length, up to the maximum length of nine • Testing is discontinued if unsuccessful on both trials at a given sequence length	• The total is the highest sequence length correctly repeated at least once, and it can range from two to nine

Variable	Task	Description	Scoring
Language Ability	Wechsler Individual Achievement Test–Second UK Edition (WIAT-II ^{UK}) Listening Comprehension	- Select the picture that matches either a word (16 Receptive Vocabulary items) or sentence (10 Sentence Comprehension items), and generate a word that matches a picture and oral description (15 Expressive Vocabulary items) - Each of the three subtasks requires a basal set of three consecutive correct responses, and each subtask is discontinued after six consecutive scores of zero	- Each item is scored as either correct (1) or incorrect (0) - The total raw score is the sum of all item scores and can range from 0 to 41
Numerical Ability	Wechsler Individual Achievement Test–Second UK Edition (WIAT-II ^{UK}) Numerical Operations	Identify and write numbers, count using one-to-one correspondence, and solve written calculation problems and simple equations involving the basic operations of addition, subtraction, multiplication, and division.	- Each item is scored as either correct (1) or incorrect (0) - The total raw score is the sum of all item scores and can range from 0 to 54
Fluid Reasoning	Wechsler Abbreviated Scale of Intelligence–Second Edition (WASI-II) Matrix Reasoning subtest	- Complete the matrix or series by choosing from five options - A total of 24 items for children aged 6 to 8, but testing discontinued after three consecutive scores of zero	- Each item is scored as either correct (1) or incorrect (0) - The total raw score is the sum of all item scores, and can range from 0 to 24 for children aged 6 to 8
Word-Problem-Solving Strategies	Additive Word Problems	- Solve word problems that require either the addition or subtraction of the values given in the problem - Can use various representational tools such as counters (MathLink cubes) and paper and pencil - A total of 24 items	- The problem-solving strategy used for each item is coded as: - Non-Quantitative - Quantitatively Incorrect Counting - Quantitatively Incorrect Number Fact - Quantitatively Correct Counting - Quantitatively Correct Number Fact

Note. Adapted from Table 2-2 (Chapter 2).

required reasoning and visualization of parts–whole relationships: missing center, missing border, two-dimensional shape composition/decomposition, and paper folding. The two measures of spatial visualization were combined into a single composite score using Principal Component Analysis (PCA) (De Pauw et al., 2009). (The details of this spatial visualization composite score, including its justification, development process, and findings, are presented in [Subsection 2.3.2.1.1](#) of the Methodology Chapter.)

Sequential visuospatial working memory was assessed using PathSpan, a digitized version of Corsi Block-Tapping task on iPad (Hawes et al., 2019b). In this measure, participants were presented with a sequence of dots that light up one at a time in a random pattern, which they must then replicate by touching the dots in the same order they were lit.

Language ability was assessed using the Wechsler Individual Achievement Test–Second UK Edition (WIAT–II^{UK}) Listening Comprehension subtest, which involves selecting a picture that matches a word or sentence and generating a word that matches a picture or oral description. In this study, listening comprehension was deemed a more relevant measure of language ability than reading comprehension because the one-step additive word problems are read aloud to the participants (Spencer et al., 2022). Numerical ability was assessed using the WIAT–II^{UK} Numerical Operations subtest, which involves identifying and writing numbers, counting, and solving arithmetic computation tasks. Fluid reasoning—the ability to identify patterns and relationships, infer and implement rules, and do so independently of verbal and prior knowledge (McGrew, 2009)—was assessed using the WASI–II Matrix subtest, in which the participants were required to select the response option that would complete the matrix or series from five options (Atit et al., 2022).

Finally, the word-problem-solving strategies were assessed using the 24 one-step additive word problem items adapted from Study 1. The problems used number facts with sums ranging from 11 to 19 as the curriculum for Year 2 pupils in England focus on mastery of addition and subtraction up to 20 (Department for Education, 2013). Of the 24 items, six belonged to each of the four word problem types outlined in [Subsection 4.2.3](#). (See [Subsection 2.3.3.2](#) for the full problem set.)

4.3.3 Procedure

Pupils met individually with the primary author and completed a series of seven measures. The measures were administered in a randomized order across two sessions, each lasting approximately 30 minutes. Sessions were typically scheduled at least one hour apart,

with a minimum interval of 30 minutes. All sessions took place in a quiet space, either in the hallway or in an open area near the pupil's classroom. Prior to testing, parental consent (either opt-in or opt-out, based on the school's discretion) and child assent were obtained.

Pupils were randomly assigned to one of the four forms of word problems (Forms A–D). Within each form, half of the word problems (i.e., 12 items) were administered in the first testing session, and the remaining half were administered in the second session. The problems were presented on A4-size papers, with four problems arranged in each quadrant of a page. Each problem was read aloud slowly and repeated if necessary. Pupils were provided with 30 MathLink cubes and a pencil and were informed that they could use them if they found them helpful; they were also reminded that they could use these materials if they were stuck on a problem for more than a minute.

While each pupil was solving a problem, overt signs of representational tool use and arithmetic approaches (e.g., using counters to count-all) were documented. Additional observations of behavior (e.g., the pattern and order in which counters were placed on the desk) were noted on the paper, as no photographic or video recordings were made. After each problem was solved and a numerical answer provided, pupils were asked to explain how they arrived at their answer: 'What were you thinking in your mind?' The retrospective reports were noted verbatim. (The detailed administration procedure for the word problems, including the randomization of the item order, is outlined in [Subsection 2.3.3.3](#) of the Methodology chapter.)

4.3.4 Coding of Problem-Solving Strategies

Pupils' strategies for solving word problems were coded in three stages. The first stage of coding was during testing. For each word problem item, the primary author coded the representational tools and arithmetic approaches based on pupils' observable behavior and retrospective reports. This method of complementing observations with retrospective reporting of the strategy was deemed most appropriate based on prior research, which indicates that such an approach is a valid indicator of capturing the strategies children use (Carr et al., 2008; Cho et al., 2011). (The final codes of representational tools and arithmetic approaches, as well as the coding procedure are presented in [Subsection 2.3.3.4](#) of the Methodology Chapter.)

The second stage of coding was done based on field notes—which included codes for representational tools and arithmetic approaches, additional observations of behavior, and verbatim retrospective reports from pupils. Using the field notes, the primary author identified the problem-solving strategy used by each pupil for each problem item. The five mutually

exclusive problem-solving strategy codes are summarized in Table 4-7. Only one final problem-solving strategy code was assigned per item per pupil.

Importantly, problem-solving strategies were coded independently of problem-solving accuracy (i.e., the correctness of the numerical solution). That is, even if the final numerical solution was incorrect, as long as the representation of the quantitative structure was accurate, the problem-solving strategy was coded as either a ‘Quantitatively Correct Counting’ or ‘Quantitatively Correct Number Fact’ strategy, depending on the arithmetic approach.

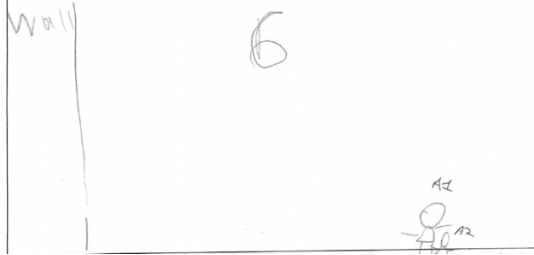
Furthermore, if there were multiple arithmetic approach codes, the final problem-solving strategy code was determined by considering the relative sophistication of the multiple codes. Consider this example: ‘Ann had 14 marbles. Then she played a game of marbles and lost 6 marbles. How many marbles does she have now?’ A pupil can solve this by first recalling that six plus six is 12, then counting-on using their fingers to reach 14, and finally determining that, because 14 is two more than 12, the answer is two more than six, which is eight. In this case, the final problem-solving code was deemed to be ‘Quantitatively Correct Number Fact’ rather than ‘Quantitatively Correct Counting’ because decomposing the problem was the critical step that enabled the pupil to reach the answer (Laski et al., 2013).

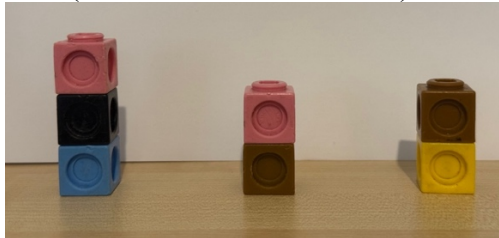
The third, and final, stage of coding was conducted to establish inter-rater reliability. A second rater independently coded the problem-solving strategies for 20.4% of the data (28 participants $\times$ 24 items) using the primary author’s field notes of pupils’ observable behavior and retrospective reports. From the total sample of 137 participants, 28 were selected at random. Interrater reliability was high. The two raters showed 97.2% agreement, and Cohen’s κ was .96, 95% CI [.95, .98], $p < .001$, indicating excellent reliability (Siegel & Castellan, 1988).

Table 4-7

Problem-Solving Strategies and Examples

Strategy	Example	
	Discrete Quantity	Continuous Quantity
‘Non-Quantitative’: Does not appear to actively engage with the problem’s quantitative structure e.g., producing an irrelevant or iconic representation, or guessing without attempting to solve	11) Ann had 6 marbles. Then she played a game of marbles. Now she has 18 marbles. How many marbles did she win?  Drew a person. Retrospective reporting: [Points to the number 18 on the problem]	14) Ann is 19 steps from the wall. Bob is 6 steps closer to the wall than Ann. How far is Bob from the wall?  Retrospective reporting: ‘Because it is.’

Strategy	Example	
	Discrete Quantity	Continuous Quantity
Quantitatively Incorrect Counting: Represents the problem's quantitative structure incorrectly while using a counting-based strategy e.g., using representational tools (concrete or figural objects, marks or drawings, fingers) and/or vocal or subvocal counting to carry out a counting strategy based on flawed quantitative reasoning	10) Ann had some marbles. Then she played a game of marbles and lost 5 marbles. Now she has 13 marbles left. How many marbles did she have to begin with? 8  Drew 13 circles and crossed out five Retrospective reporting: 'It says some, so I don't know what to draw. So, I start with 13 and take away 5.'	22) Ann was some steps from the wall. Then she walked 5 steps towards the wall. Now she is 14 steps from the wall. How far was Ann from the wall to begin with? ⁹ Used fingers to count-down from 14. Retrospective reporting: 'I minused 5 from 14.'
Quantitatively Incorrect Number-Fact: Represents the problem's quantitative structure incorrectly while using a number-fact strategy e.g., using derived or recalled number facts based on flawed quantitative reasoning	4) Ann has 5 marbles. She has 9 marbles less than Bob. How many marbles does Bob have? ⁴ Retrospective reporting: 'Ann had 5. But 9 take away 5 is 4, and that's how many Bob has.'	22) Ann was some steps from the wall. Then she walked 6 steps towards the wall. Now she is 12 steps from the wall. How far was Ann from the wall to begin with?  Drew the wall. Then pointed with their finger and glided it from left to right, starting at the wall and moving once for each counted number while counting aloud to six, then drew A1 (labeled by the researcher). Repeated the gesture, counting to twelve from the wall, and drew A2 (labeled by the researcher). Finally, stated that the answer was six. Retrospective reporting: 'I know 12 minus 6 is 6.'

Strategy	Example	
	Discrete Quantity	Continuous Quantity
Quantitatively Correct Counting: Represents the problem's quantitative structure correctly while using a counting-based strategy e.g., using representational tools (concrete or figural objects, marks or drawings, fingers) and/or vocal or subvocal counting to carry out a counting strategy based on correct quantitative reasoning	13) Ann is 5 steps from the wall. Bob is 13 steps farther from the wall than Ann. How far is Bob from the wall? 18 Used left hand to count-on from 13. Retrospective reporting: 'Because she has 5 steps. He's farther 13 steps. So, I added it up.'	16) Ann is 5 steps from the wall. She is 7 steps closer to the wall than Bob. How far is Bob from the wall. 13 Stacked three MathLink cubes to represent the wall, two to represent Ann, and two to represent Bob. Placed the MathLinks representing the wall down first, then positioned Ann's cubes arbitrarily to the right of the wall, and Bob's cubes arbitrarily right of Ann (i.e., farther from the wall). Recreation:  Finally, used fingers to count on from five.
Quantitatively Correct Number-Fact: Represents the problem's quantitative structure correctly while using a number-fact strategy e.g., Using derived or recalled number facts based on correct quantitative reasoning	7) Ann had 9 marbles. Then she played a game of marbles and won 10 more marbles. How many marbles does she have now? 16p Retrospective reporting: 'Because 10 plus 10 is 20. So, minus 1 is 19.'	24) Ann was 16 steps from the wall. Then she walked some steps. Now she is 5 steps from the wall. How many steps did Ann walk towards the wall? 16 Retrospective reporting: 'Cause if you have 16, take away 5 is 11.'

4.3.5 Analytic Approach

Statistical analyses were conducted in R version 4.4.0 (R Core Team, 2024) using RStudio version 2023.12.1.402 (Posit Team, 2024). Both research questions—examining (1) the relationship between spatial skills and problem-solving strategy use, and (2) the moderating effect of word problem type—were investigated using mixed-effects multinomial logistic regression models (Snijders & Bosker, 2012) implemented in the *mclogit* package (Elff, 2026). Random intercepts for participants and items were included to account for the cross-classified structure of the data, as each pupil solved all 24 word problems and each problem was completed by all 137 participants (Baayen et al., 2008; Barr et al., 2013). Consistent with findings from Study 1 indicating that item-level random slopes did not improve model fit, the multinomial models retained a parsimonious random-intercept structure for participants and items. The ‘Quantitatively Correct Number Fact’ strategy was used as the baseline category based on the *a priori* hypothesis that spatial skills would most strongly predict the likelihood of using this strategy.

In most previous studies, the relationship between spatial ability and problem-solving strategies has been investigated using aggregated participant-level data and separate regression analyses for each strategy (e.g., Casey et al., 2017; Laski et al., 2013; Möhring et al., 2024). However, aggregating item-level responses to the participant level may obscure meaningful variation, as individuals with identical scores may differ in how they use strategies across items, potentially masking important differences in problem-solving patterns. This discards informative within-participant and item-level variation and obscures the binomial response structure of the data (Jaeger, 2008).

Furthermore, the use of proportions at the participant level introduces additional statistical considerations, as strategy proportions are not independent but are constrained to sum to one, resulting in compositional data (Aitchison, 1982). As a result, modeling each strategy separately may not fully account for the dependency among outcomes, given that selecting one strategy is inherently related to not selecting others. This, in turn, may lead to biased estimates and inflated Type I error rates.

To address these limitations, the present study employed an item-level modeling approach in which individual item responses served as the unit of analysis while accounting for clustering across both participants and items through cross-classified mixed-effects models. This approach allows the relationship between spatial skills and strategy use to be examined at

the level at which strategy selection occurs (Boonen et al., 2014). Beyond addressing the limitations associated with aggregation and model specification, this approach offers two additional advantages. First, it can increase statistical power by retaining information contained in individual item responses rather than aggregating performance across items. Second, by modeling variability at both the participant and item levels, the analysis supports generalization beyond the specific sample of pupils and set of word problems included in the study (DeBruine & Barr, 2021).

RQ3 was investigated by fitting a cross-classified mixed-effects multinomial logistic regression model with spatial skills (i.e., spatial visualization and visuospatial working memory) as independent variables; language ability, numerical ability, fluid reasoning, age, and gender as covariates; problem-solving strategy as the dependent variable; and participants and problem items as crossed random effects. RQ4, examining the moderating effect of word problem type, was addressed by adding interaction terms between word problem type and the two spatial skills. Discrete Quantity–Dynamic Relation problems were used as the reference category, as they provide a meaningful and interpretable baseline for comparing variation in strategy use across problem types within the proposed framework (see Study 1). To facilitate interpretation, conditional odds ratios were calculated by multiplying the main-effect odds ratio by the interaction-term odds ratio; these should be interpreted as estimates of the conditional effect for a given problem type relative to the reference word problem type.

All continuous predictors were standardized to place them on a common metric and facilitate comparison of effect estimates across predictors. Age was mean centered rather than standardized because its original metric (i.e., year) was considered substantively meaningful and readily interpretable. Mean centering additionally improved the interpretability of model intercepts and reduced potential multicollinearity (Enders & Tofighi, 2007). Model assumptions were evaluated for the multinomial models, with full details provided in [Appendix 4A](#). Diagnostics were conducted in R using relevant packages, including *splines* (R Core Team, 2024), *car* (Fox & Weisberg, 2019) and *caret* (Kuhn, 2024).

4.4 Findings

The findings from cross-classified mixed-effects multinomial logistic regression analyses are reported below. [Subsection 4.4.1](#) reports the association between pupils' spatial skills and their use of problem-solving strategies (RQ3). [Subsection 4.4.2](#) reports the

moderating effect of word problem type on these associations (RQ4) by presenting the final model, which includes interaction terms between word problem type and spatial skills.

4.4.1 RQ3: The Relationship between Spatial Skills and Problem-Solving Strategy Use

RQ3) To what extent do spatial visualization and visuospatial working memory predict young primary school pupils' (6–7-year-olds') problem-solving strategy use?

4.4.1.1 Preliminary Analyses

Prior to presenting the main-effects model, preliminary analyses are reported, including the distribution of problem-solving strategies ([Subsection 4.4.1.1.1](#)), between-pupil ([Subsection 4.4.1.1.2](#)) and between-item ([Subsection 4.4.1.1.3](#)) variation in strategy use, and descriptive statistics ([Subsection 4.4.1.1.4](#)).

4.4.1.1.1 Problem-Solving Strategy Distribution

Table 4-8 summarizes the overall distribution of problem-solving strategies across all item-level responses. The largest proportion of problem items was solved using a ‘Non-Quantitative’ strategy (34.2%), followed by ‘Quantitatively Correct Counting’ (29.1%) and ‘Quantitatively Correct Number Fact’ (20.2%) strategies. ‘Quantitatively Incorrect Counting’ strategies accounted for about 11.5% of responses, while ‘Quantitatively Incorrect Number Fact’ strategies were used least frequently (5.1%).

Table 4-8

Distribution of Problem-Solving Strategies Across all Item-Level Responses

	Frequency (<i>n</i>)	Percent (%)
Non-Quantitative	1124	34.2%
Quantitatively Incorrect Counting	378	11.5%
Quantitatively Incorrect Number Fact	167	5.1%
Quantitatively Correct Counting	955	29.1%
Quantitatively Correct Number Fact	664	20.2%

Note. $N = 3288$ item-level responses (137 pupils $\times$ 24 problems).

Taken together, counting-based strategies (40.6%)—whether quantitatively correct or incorrect—were used more frequently than number-fact strategies (25.3%). Moreover, pupils relied on ‘Non-Quantitative’ strategies (34.2%) more often than on quantitatively-incorrect strategies (counting-based and number-fact combined; 16.6%), suggesting that not representing the quantitative structure was more common than representing it incorrectly.

4.4.1.1.2 Between-Pupil Variation in Problem-Solving Strategy Distribution

Table 4-9 presents between-pupil variation in problem-solving strategy use. As indicated by the coefficient of variation (CV), there was substantial between-pupil variability across all five strategies. CVs were highest for ‘Quantitatively Incorrect Number Fact’ (1.51) and ‘Quantitatively Incorrect Counting’ (1.24), followed by ‘Quantitatively Correct Number Fact’ (1.23), ‘Non-Quantitative’ (1.05), and ‘Quantitatively Correct Counting’ (0.74).

Table 4-9

Between-Pupil Variation in Problem-Solving Strategy Use

	Mean (%)	SD (%)	CV
Non-Quantitative	34.2	35.8	1.05
Quantitatively Incorrect Counting	11.5	14.3	1.24
Quantitatively Incorrect Number Fact	5.1	7.7	1.51
Quantitatively Correct Counting	29.1	21.5	0.74
Quantitatively Correct Number Fact	20.2	24.9	1.23

Note. $N = 137$ pupils. SD = Standard Deviation; CV = Coefficient of Variation ($SD/$ Mean).

4.4.1.1.3 Between-Item Variation in Problem-Solving Strategy Distribution

Table 4-10 presents between-item variation in problem-solving strategy use, indicating the extent to which problem-solving strategies varied by item. ‘Quantitatively Incorrect Number Fact’ strategies exhibited the highest between-item variability ($CV = 0.89$), followed by ‘Quantitatively Incorrect Counting’ ($CV = 0.56$), suggesting that certain items prompted these strategies more frequently than others. ‘Quantitatively Correct Counting’ ($CV = 0.35$) and ‘Quantitatively Correct Number Fact’ ($CV = 0.35$) strategies demonstrated moderate variability across items. In contrast, ‘Non-Quantitative’ strategies showed relatively low between-item variability ($CV = 0.28$), indicating that this strategy was used at more similar rates across items.

Table 4-10

Between-Item Variation in Problem-Solving Strategy Use

	Mean (%)	SD (%)	CV
Non-Quantitative	34.2	9.6	0.28
Quantitatively Incorrect Counting	11.5	6.5	0.56
Quantitatively Incorrect Number Fact	5.1	4.5	0.89
Quantitatively Correct Counting	29.1	10.1	0.35
Quantitatively Correct Number Fact	20.2	7.1	0.35

Note. Based on 24 items. SD = Standard Deviation; CV = Coefficient of Variation ($SD/$ Mean).

4.4.1.1.4 Descriptive Statistics and Correlations

Table 4-11 presents descriptive statistics and zero-order and partial correlations among key variables. Zero-order correlations (above diagonal) indicated associations between spatial skills and problem-solving strategies. Both spatial visualization and visuospatial working memory were positively correlated with number-fact strategies, regardless of the correctness of the representation of quantitative structure: ‘Quantitatively Incorrect Number Fact’ ($r = .49, p < .001$; $r = .26, p < .01$, respectively) and ‘Quantitatively Correct Number Fact’ ($r = .55, p < .001$; $r = .37, p < .001$, respectively). Both spatial skills were negatively correlated with ‘Non-Quantitative’ strategies ($r = -.37, p < .001$; $r = -.31, p < .001$, respectively). In addition, spatial visualization was negatively correlated with ‘Quantitatively Incorrect Counting’ strategies ($r = -.18, p < .05$).

After controlling for language ability, numerical ability, fluid reasoning, age, gender, and the other spatial skill (partial correlations; below diagonal) only spatial visualization remained significantly associated with strategy use. Specifically, spatial visualization was positively associated with ‘Quantitatively Incorrect Number Fact’ ($r = .28, p < .01$) and ‘Quantitatively Correct Number Fact’ ($r = .22, p < .05$). Although the zero-order correlation between spatial visualization and ‘Quantitatively Correct Counting’ strategy use was not significant, the partial correlation controlling for covariates and visuospatial working memory was significant ($r = -.21, p < .05$).

Examination of zero-order correlations between covariates and problem-solving strategies indicated that numerical ability and fluid reasoning were positively correlated with number-fact strategies, regardless of the correctness of the representation of the problem’s quantitative structure: ‘Quantitatively Incorrect Number Fact’ ($r = .44, p < .001$; $r = .38, p < .001$, respectively) and ‘Quantitatively Correct Number Fact’ ($r = .64, p < .001$; $r = .54, p < .001$, respectively). In contrast, language ability was positively correlated with both quantitatively-correct strategies, regardless of the arithmetic strategy: ‘Quantitatively Correct Counting’ ($r = .26, p < .01$) and ‘Quantitatively Correct Number Fact’ ($r = .32, p < .001$). All three covariates were negatively correlated with ‘Non-Quantitative’ strategies ($r = -.60, p < .001$; $r = -.48, p < .001$; $r = -.39, p < .001$, respectively). Moreover, age was negatively associated with ‘Non-Quantitative’ strategy ($r = -.21, p < .05$). Gender was positively associated with ‘Quantitatively Correct Number Fact’ strategy ($r = .20, p < .05$) strategy, indicating that boys tended to use this strategy more frequently than girls.

Table 4-11*Descriptive Statistics and Bivariate Correlations Among Variables*

Variable	Mean (SD)	Min	Max	1	2	3	4	5	6	7	8	9	10	11	12
1. Composite Spatial Visualization	0.00 (1.00)	-1.90	3.53	-	.42***	-.37***	-.18*	.49***	-.08	.55***	.26**	.55***	.53***	.02	.11
2. Visuospatial Working Memory	4.39 (1.21)	2	8		-	-.31***	-.06	.26**	.03	.37***	.16	.39***	.29***	.06	.07
3. 'Non-Quantitative' Strategies	8.20 (8.60)	0	24	0.02	-0.09	-	-.37***	-.36***	-.61***	-.59***	-.39***	-.60***	-.48***	-.21*	-.10
4. 'Quantitatively Incorrect Counting' Strategies	2.76 (3.43)	0	16	-0.13	0.00		-	-.27**	.45***	-.34***	-.06	-.03	-.16	.03	-.08
5. 'Quantitatively Incorrect Number Fact' Strategies	1.22 (1.84)	0	11	0.28**	0.02			-	-.26**	.58***	.14	.44***	.38***	.06	.14
6. 'Quantitatively Correct Counting' Strategies	6.97 (5.16)	0	21	-0.21*	0.03				-	-.16	.26**	.13	.15	.13	-.06
7. 'Quantitatively Correct Number Fact' Strategies	4.85 (5.98)	0	23	0.22*	0.10					-	.32***	.64***	.54***	.15	.20*
8. Language Ability	17.78 (3.89)	8	27								-	.36***	.42***	-.01	-.17*
9. Numerical Ability	12.71 (3.26)	5	23									-	.57***	.19*	.16
10. Fluid Reasoning	9.03 (4.22)	2	20										-	.23**	.02
11. Age (years)	7.10 (0.30)	6.52	7.74											-	.09
12. Gender (0 = girls; 1 = boys)	0.50 (0.50)	0	1												-

Note. $N = 137$. SD = standard deviation. The descriptive statistics reported for strategy variables (3–7) are the mean number of problems (out of 24) on which each strategy type was used per pupil. The composite spatial visualization score was derived from the principal component analysis of the Visual–Spatial Reasoning task and the WASI–II Block Design subtest. Above diagonal: zero-order correlations; below diagonal: partial correlations, controlled for the other spatial skill, age, gender, language and numerical abilities, and fluid reasoning. Strategy variables are expressed as raw frequencies in the descriptive statistics and as participant-level proportions in the correlations. *** $p < .001$, ** $p < .01$, * $p < .05$.

4.4.1.2 Inferential Analyses (Main-Effects Model)

A cross-classified mixed-effects multinomial logistic regression model was fitted to examine whether young pupils' spatial skills predicted problem-solving strategy use on individual one-step additive word problems. The model converged without singularity. The deviance-based pseudo- R^2 of .39 indicated a substantial improvement in fit relative to an intercept-only model.

Random intercept standard deviations were nonzero at both the pupil ($\sigma_{\text{Participant}} = 0.98\text{--}2.00$) and item levels ($\sigma_{\text{Item}} = 0.24\text{--}0.65$), indicating that variability in strategy use was present at both levels, with greater variability attributable to between-pupil differences than between-item differences. This pattern is consistent with the descriptive analyses of between-pupil and between-item variation. Model assumptions were largely satisfied (see [Appendix 4A](#)). Because numerical ability exhibited evidence of nonlinearity, it was modeled using a natural spline term ($df = 2$), which improved model fit relative to a linear specification ($\Delta AIC = -28.6$; $\Delta BIC = -4.22$).

A pseudo-likelihood-ratio comparison between the main-effects model and a reduced model excluding spatial visualization was significant ($\chi^2(4) = 10.75, p = .03$), indicating that spatial visualization significantly improved overall model fit. The corresponding comparison excluding visuospatial working memory was not significant ($\chi^2(4) = 7.16, p = .13$), suggesting that visuospatial working memory did not explain additional variance in strategy use beyond the other predictors. These findings were consistent with the partial correlation analyses, in which spatial visualization showed stronger associations with strategy use than visuospatial working memory.

Table 4-12 presents odds ratios (OR) and 95% confidence intervals for all predictors. Examination of individual model coefficients indicated that, holding all other continuous predictors at their mean values (i.e., zero after standardization) and gender at its reference category (i.e., girls), higher spatial visualization was associated with reduced odds of using 'Quantitatively Incorrect Counting' (OR = 0.61, 95% CI [0.39, 0.95], $p = .03$) and 'Quantitatively Correct Counting' (OR = 0.59, 95% CI [0.40, 0.87], $p = .01$) strategies relative to the baseline 'Quantitatively Correct Number Fact' strategy. Spatial visualization was not significantly associated with the odds of using 'Non-Quantitative' (OR = 0.90, 95% CI [0.52, 1.57], $p = .71$) or 'Quantitatively Incorrect Number Fact' (OR = 1.20, 95% CI [0.85, 1.67], $p = .30$) strategies relative to the baseline.

Table 4-12*RQ3: Cross-Classified Mixed-Effects Multinomial Model Odds Ratios and 95% Confidence Interval*

Predictors	Non-Quantitative		Quantitatively Incorrect Counting		Quantitatively Incorrect Number Fact		Quantitatively Correct Counting	
	Odds Ratios [95% CI]	<i>p</i>	Odds Ratios [95% CI]	<i>p</i>	Odds Ratios [95% CI]	<i>p</i>	Odds Ratios [95% CI]	<i>p</i>
Spatial Visualization	0.90 [0.52, 1.57]	.71	0.61 [0.39, 0.95]	.03	1.20 [0.85, 1.67]	.30	0.59 [0.40, 0.87]	.007
Visuospatial Working Memory	0.75 [0.47, 1.18]	.22	0.81 [0.57, 1.15]	.24	0.87 [0.63, 1.18]	.36	0.88 [0.64, 1.21]	.42
Language Ability	0.48 [0.30, 0.76]	.002	0.82 [0.57, 1.18]	.30	0.73 [0.54, 1.00]	.047	0.99 [0.72, 1.37]	.97
Numerical Ability (df = 2)1	9.73×10^{-5} [1.03×10^{-6} , 0.01]	< .001	92.21 [1.45, 5869.09]	.03	0.25 [0.01, 6.03]	.39	8.11 [0.27, 239.73]	.23
Numerical Ability (df = 2)2	4.95×10^{-4} [1.22×10^{-5} , 0.02]	< .001	0.01 [0.00, 0.10]	< .001	0.41 [0.08, 2.13]	.29	0.03 [0.00, 0.24]	.001
Fluid Reasoning	0.43 [0.24, 0.76]	.004	0.62 [0.40, 0.95]	.03	0.82 [0.58, 1.16]	.26	0.81 [0.55, 1.17]	.26
Age	0.33 [0.08, 1.37]	.13	1.07 [0.34, 3.33]	.91	0.91 [0.33, 2.49]	.85	1.39 [0.50, 3.87]	.53
Gender	0.37 [0.16, 0.84]	.02	0.48 [0.25, 0.92]	.03	0.68 [0.38, 1.21]	.19	0.55 [0.31, 1.00]	.049
$\sigma_{\text{Participant}}$	2.00		1.46		0.98		1.36	
σ_{Item}	0.65		0.46		0.24		0.65	

Note. *N* = 137 pupils; 24 items; 3,288 item-level responses. ‘Quantitatively Correct Number Fact’ strategy is the baseline outcome category. Because spline terms represent basis functions rather than the original predictor, their odds ratios are not directly interpretable; interpretation focuses on predicted probabilities. Random effects: AIC = 6539, BIC = 6881.

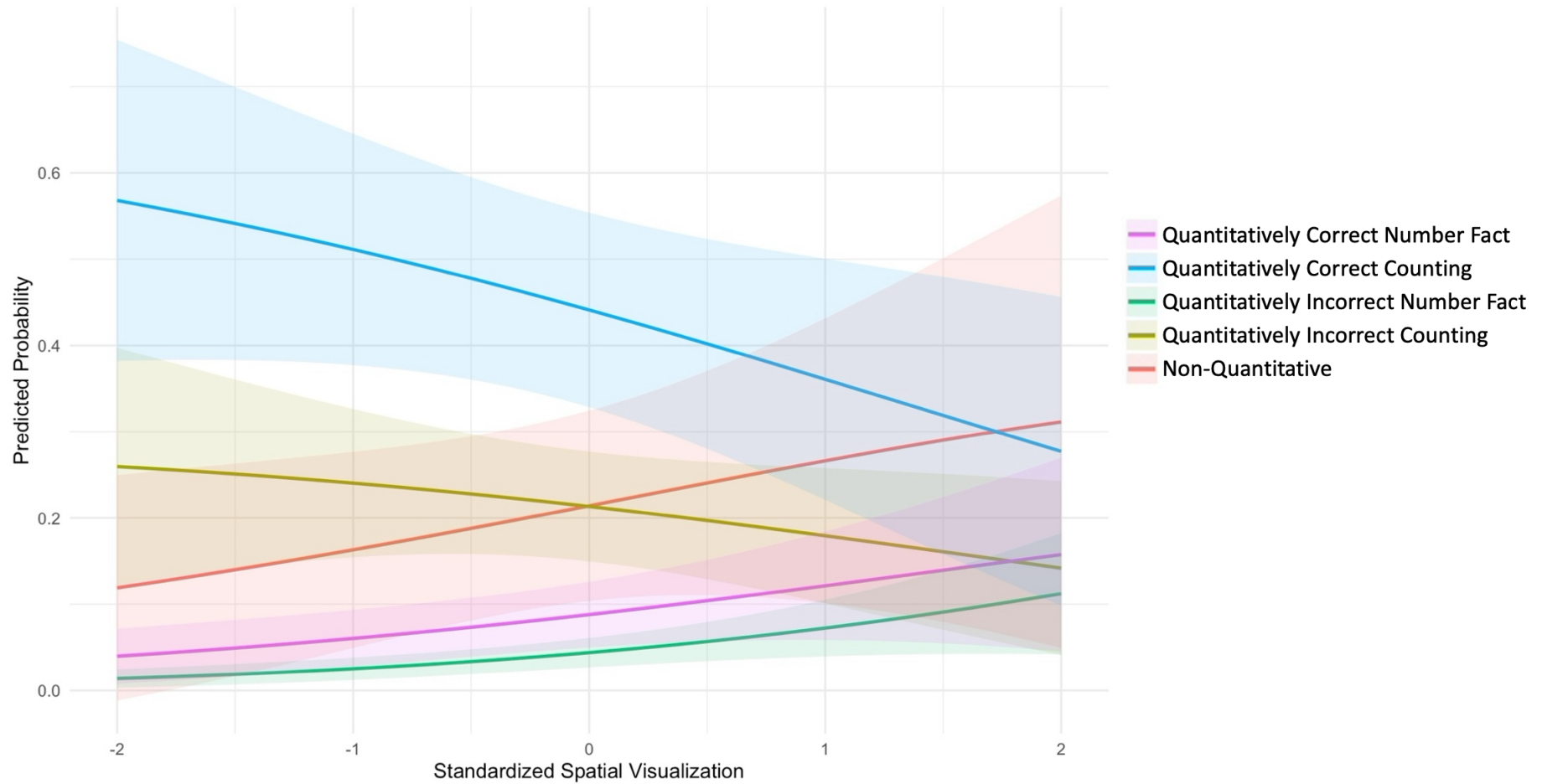
In contrast, visuospatial working memory did not significantly predict the likelihood of using any problem-solving strategies relative to the baseline category. Figure 4-4 and Figure 4-5 illustrate the predicted probabilities of each problem-solving strategy across the observed ranges of spatial visualization and visuospatial working memory scores, respectively, for pupils in the reference gender category (girls), with other continuous covariates held at their mean values.

Among the demographic variables, gender, but not age, significantly predicted pupils' problem-solving strategy use. On individual one-step additive word problems, boys were less likely than girls to use counting-based ('Quantitatively Incorrect Counting': OR = 0.48, 95% CI [0.25, 0.92], $p = .03$; 'Quantitatively Correct Counting': OR = 0.55, 95% CI [0.31, 1.00], $p = .049$) or 'Non-Quantitative' (OR = 0.37, 95% CI [0.16, 0.84], $p = .02$) strategies relative to the baseline category.

All cognitive control variables (i.e., language ability, numerical ability, and fluid reasoning) were also significantly associated with pupils' problem-solving strategy use. Higher language ability was associated with lower odds of using 'Non-Quantitative' (OR = 0.48, 95% CI [0.30, 0.76], $p = .002$) and 'Quantitatively Incorrect Number Fact' (OR = 0.73, 95% CI [0.54, 1.00], $p = .047$) strategies, relative to the baseline category. Numerical ability, modeled with a natural spline term ($df = 2$), significantly predicted strategy use overall; predicted probabilities across its observed range are presented in [Appendix 4B](#) for descriptive interpretation. Finally, higher fluid reasoning was associated with lower odds of using a 'Non-Quantitative' (OR = 0.43, 95% CI [0.24, 0.76], $p = .004$) and 'Quantitatively Incorrect Counting' (OR = 0.62, 95% CI [0.40, 0.95], $p = .03$) strategies relative to the baseline category.

Figure 4-4

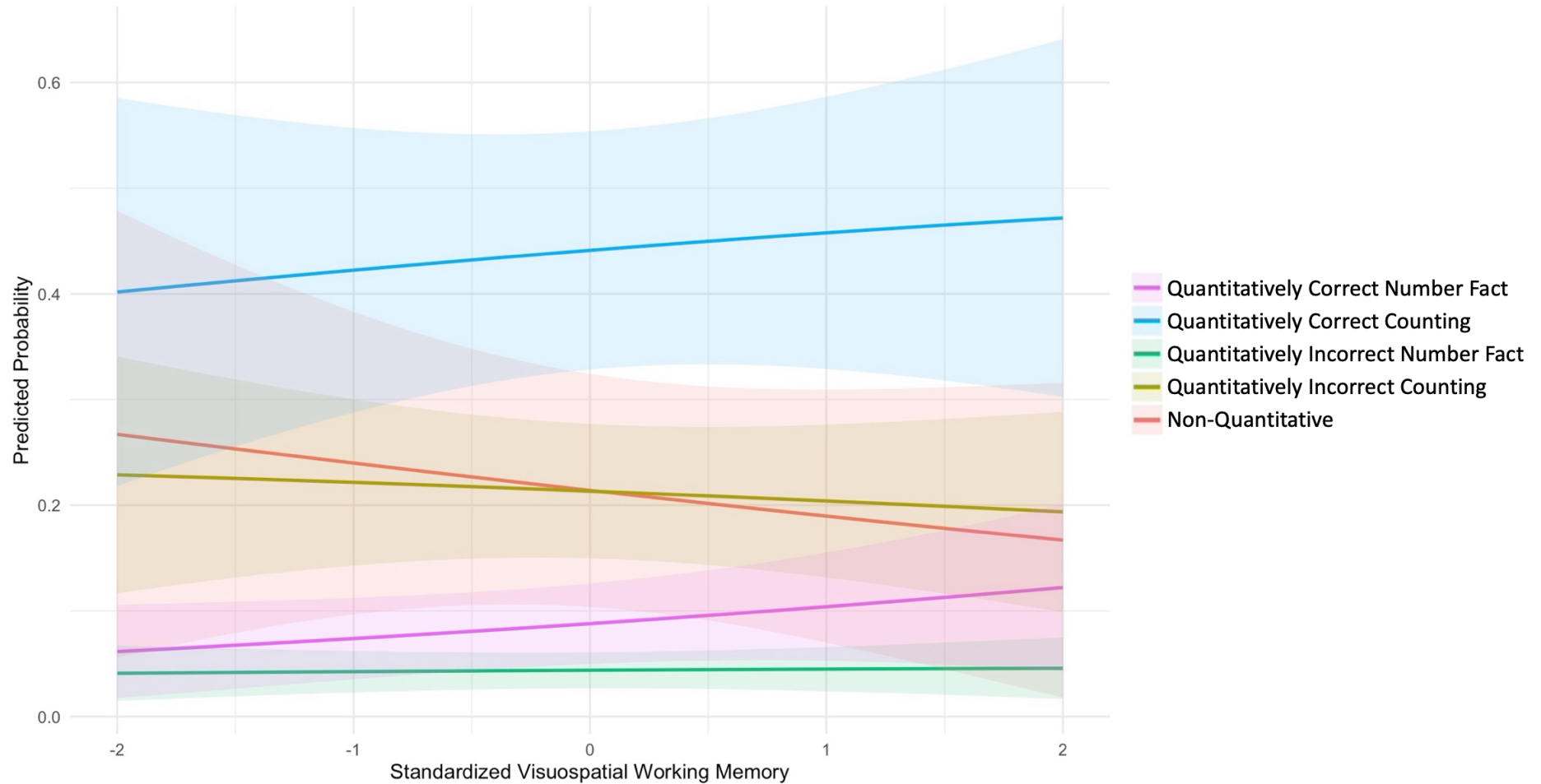
Predicted Probabilities of Problem-Solving Strategy Use across Spatial Visualization



Note. Predicted probabilities were estimated from the multinomial mixed-effects logit model while holding continuous covariates at their mean values and gender at the reference level (Female). Shaded bands indicate 95% confidence intervals.

Figure 4-5

Predicted Probabilities of Problem-Solving Strategy Use across Visuospatial Working Memory



Note. Predicted probabilities were estimated from the multinomial mixed-effects logit model while holding continuous covariates at their mean values and gender at the reference level (Female). Shaded bands indicate 95% confidence intervals.

4.4.2 RQ4: Interaction Effects of Spatial Skills and Problem Type on Problem-Solving Strategy Use

RQ4) Do the relationships between spatial visualization, visuospatial working memory, and problem-solving strategy use vary across different word problem types?

4.4.2.1 Descriptive Statistics

Prior to examining the potential moderating effect of word problem type on the relationship between spatial skills and problem-solving strategy use, descriptive distributions of strategies across the four word problem types were examined. As summarized in Table 4-13, strategy use varied across the four word problem types. Descriptively, Discrete Quantity–Dynamic Relation problems exhibited a distinctive strategy profile compared to the other problem types. Whereas ‘Non-Quantitative’ was the most common strategy for solving the other three problem types, this problem type was solved primarily using ‘Quantitatively Correct Counting’ (38.1%) and ‘Quantitatively Correct Number Fact’ (27.5%) strategies. It also showed the lowest proportions of ‘Non-Quantitative’ (23.4%), ‘Quantitatively Incorrect Counting’ (9.1%), and ‘Quantitatively Incorrect Number Fact’ (2.0%) strategies.

Table 4-13

Distribution of Problem-Solving Strategies across Problem Types

	Discrete Quantity –Dynamic Relation	Discrete Quantity –Static Relation	Continuous Quantity –Dynamic Relation	Continuous Quantity –Static Relation
	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)
Non-Quantitative	192 (23.4%)	313 (38.1%)	261 (31.8%)	358 (43.6%)
Quantitatively Incorrect Counting	75 (9.1%)	87 (10.6%)	121 (14.7%)	95 (11.6%)
Quantitatively Incorrect Number Fact	16 (2.0%)	40 (4.9%)	63 (7.7%)	48 (5.8%)
Quantitatively Correct Counting	313 (38.1%)	225 (27.4%)	228 (27.7%)	189 (23.0%)
Quantitatively Correct Number Fact	226 (27.5%)	157 (19.1%)	149 (18.1%)	132 (16.1%)

Note. Values represent frequencies (*n*) and percentages of each problem-solving strategy within each problem type. Each column is based on 822 item-level responses (137 pupils × 6 problems per type).

In contrast, Continuous Quantity–Static Relation problems were characterized by the highest use of ‘Non-Quantitative’ strategies (43.6%), and the lower use of ‘Quantitatively Correct Counting’ (23.0%) and ‘Quantitatively Correct Number Fact’ (16.1%) strategies. Across all problem types, quantitatively-incorrect strategies were generally used less

frequently, with ‘Quantitatively Incorrect Number Fact’ being particularly rare (2.0%–7.7%). Both quantitatively-incorrect strategies were most common in Continuous Quantity–Dynamic Relation problems (14.7% and 7.7%, respectively).

Descriptively, strategy use differed across both the quantity and quantitative relation dimensions. First, both discrete-quantity problems were associated with lower use of quantitatively-incorrect strategies and higher use of quantitatively-correct strategies than continuous-quantity problems. Second, both dynamic-relation problems showed lower use of ‘Non-Quantitative’ strategies and higher use of ‘Quantitatively Correct Counting’ strategies than static-relation problems.

4.4.2.2 Inferential Analyses (Interaction Model)

To examine whether word problem type moderates the association between pupils’ spatial skills and problem-solving strategy use, the previous cross-classified mixed-effects multinomial logit model was extended to include interaction terms for spatial visualization $\times$ problem type and visuospatial working memory $\times$ problem type. Discrete Quantity–Dynamic Relation problems served as the reference category.

The interaction model converged without singularity. Model fit improved modestly relative to the main-effects model (deviance-based pseudo- $R^2 = .41$ from $.39$), indicating that the inclusion of interaction terms provided additional explanatory value. Random intercept standard deviations changed only slightly following the inclusion of interaction terms ($\sigma_{\text{Participant}} = 1.00\text{--}2.07$; $\sigma_{\text{Item}} = 0.37\text{--}0.66$), with between-pupil variability remaining substantially larger than between-item variability. Table 4-14 summarizes the odds ratios, and 95% confidence intervals for all predictors.

4.4.2.2.1 Spatial Visualization and Problem-Solving Strategy by Problem Type

A pseudo-likelihood-ratio comparison between the full interaction model and a reduced model excluding the spatial visualization $\times$ problem type interaction term was significant, $\chi^2(12) = 25.17, p = .01$, indicating that the association between spatial visualization and strategy use differed across problem types at the model level. However, none of the individual interaction coefficients reached conventional levels of statistical significance. This pattern suggests that the moderating effect of problem type is relatively modest and distributed across interaction terms rather than concentrated in a single, large interaction effect.

Table 4-14*RQ4: Cross-Classified Mixed-Effects Multinomial Model with Interaction Terms Odds Ratios and 95% Confidence Interval*

Predictors	Non-Quantitative		Quantitatively Incorrect Counting		Quantitatively Incorrect Number Fact		Quantitatively Correct Counting	
	Odds Ratios [95% CI]	<i>p</i>	Odds Ratios [95% CI]	<i>p</i>	Odds Ratios [95% CI]	<i>p</i>	Odds Ratios [95% CI]	<i>p</i>
Main Effects								
Spatial Visualization	0.85 [0.43, 1.68]	.63	0.51 [0.29, 0.90]	.02	0.93 [0.50, 1.73]	.83	0.63 [0.41, 0.96]	.03
Visuospatial Working Memory	0.50 [0.28, 0.87]	.01	1.09 [0.69, 1.72]	.70	1.11 [0.59, 2.09]	.75	0.87 [0.60, 1.25]	.45
Language Ability	0.45 [0.28, 0.74]	.001	0.81 [0.57, 1.17]	.26	0.71 [0.52, 0.97]	.03	0.99 [0.72, 1.36]	.94
Numerical Ability (df = 2)1	6.83×10^{-5} [6.20×10^{-7} , 0.01]	< .001	76.13 [1.23, 4720.60]	.04	0.21 [0.01, 5.26]	.34	6.42 [0.22, 185.49]	.28
Numerical Ability (df = 2)2	2.49×10^{-4} [5.15×10^{-6} , 0.01]	< .001	0.01 [0.00, 0.09]	< .001	0.35 [0.06, 1.85]	.21	0.03 [0.00, 0.24]	.001
Fluid Reasoning	0.42 [0.23, 0.76]	.004	0.61 [0.40, 0.93]	.02	0.80 [0.56, 1.13]	.20	0.80 [0.55, 1.16]	.23
Age	0.35 [0.08, 1.50]	.16	1.05 [0.34, 3.24]	.94	0.88 [0.32, 2.43]	.80	1.34 [0.49, 3.71]	.57
Gender	0.35 [0.15, 0.82]	.02	0.47 [0.24, 0.90]	.02	0.66 [0.37, 1.18]	.16	0.55 [0.31, 0.98]	.04
Discrete Quantity–Static Relation	9.85 [4.33, 22.41]	< .001	3.10 [1.44, 6.67]	.004	4.63 [1.99, 10.77]	< .001	1.52 [0.66, 3.49]	.33
Continuous Quantity–Dynamic Relation	4.98 [2.20, 11.27]	< .001	4.28 [2.03, 9.04]	< .001	7.71 [3.49, 17.03]	< .001	1.37 [0.60, 3.13]	.46
Continuous Quantity–Static Relation	20.80 [9.06, 47.71]	< .001	4.40 [2.05, 9.47]	< .001	6.58 [2.85, 15.16]	< .001	1.49 [0.64, 3.44]	.35

Predictors	Non-Quantitative		Quantitatively Incorrect Counting		Quantitatively Incorrect Number Fact		Quantitatively Correct Counting	
	Odds Ratios [95% CI]	<i>p</i>	Odds Ratios [95% CI]	<i>p</i>	Odds Ratios [95% CI]	<i>p</i>	Odds Ratios [95% CI]	<i>p</i>
Interaction Effects								
Spatial Visualization × Discrete Quantity–Static Relation	0.66 [0.37, 1.17]	.15	0.84 [0.46, 1.53]	.57	1.32 [0.68, 2.58]	.41	0.70 [0.47, 1.03]	.07
Spatial Visualization × Continuous Quantity–Dynamic Relation	0.78 [0.44, 1.37]	.38	1.26 [0.73, 2.18]	.40	0.99 [0.52, 1.86]	.97	0.90 [0.62, 1.31]	.58
Spatial Visualization × Continuous Quantity–Static Relation	1.67 [0.95, 2.94]	.07	1.52 [0.86, 2.68]	.15	1.90 [0.99, 3.65]	.055	1.15 [0.78, 1.69]	.49
Visuospatial Working Memory × Discrete Quantity–Static Relation	1.78 [1.10, 2.88]	.02	0.67 [0.41, 1.08]	.10	0.72 [0.35, 1.48]	.37	1.01 [0.71, 1.45]	.95
Visuospatial Working Memory × Continuous Quantity–Dynamic Relation	1.61 [1.00, 2.60]	.05	0.89 [0.56, 1.41]	.62	1.06 [0.53, 2.11]	.87	1.31 [0.91, 1.87]	.15
Visuospatial Working Memory × Continuous Quantity–Static Relation	1.41 [0.86, 2.30]	.18	0.52 [0.32, 0.84]	.007	0.51 [0.25, 1.04]	.065	0.72 [0.50, 1.05]	.09
Random Effects								
$\sigma_{\text{Participant}}$	2.07		1.45		1.00		1.35	
σ_{Item}	0.59		0.52		0.37		0.66	

Note. $N = 137$ pupils; 24 items; 3,288 item-level responses. ‘Quantitatively Correct Number Fact’ strategy is the baseline outcome category. Discrete Quantity–Dynamic Relation is the reference problem type. AIC = 6443, BIC = 7004. Because spline terms represent basis functions rather than the original predictor, their odds ratios are not directly interpretable; interpretation focuses on predicted probabilities.

For the reference Discrete Quantity–Dynamic Relation problems, higher spatial visualization was associated with reduced odds of using ‘Quantitatively Incorrect Counting’ (OR = 0.51, 95% CI [0.29, 0.90], $p = .02$) and ‘Quantitatively Correct Counting’ (OR = 0.63, 95% CI [0.41, 0.96], $p = .03$) strategies relative to the baseline ‘Quantitatively Correct Number Fact’ strategy. Spatial visualization was not significantly associated with the odds of using ‘Non-Quantitative’ (OR = 0.85, 95% CI [0.43, 1.68], $p = .63$) or ‘Quantitatively Incorrect Number Fact’ (OR = 0.93, 95% CI [0.50, 1.73], $p = .83$) strategies relative to the baseline strategy.

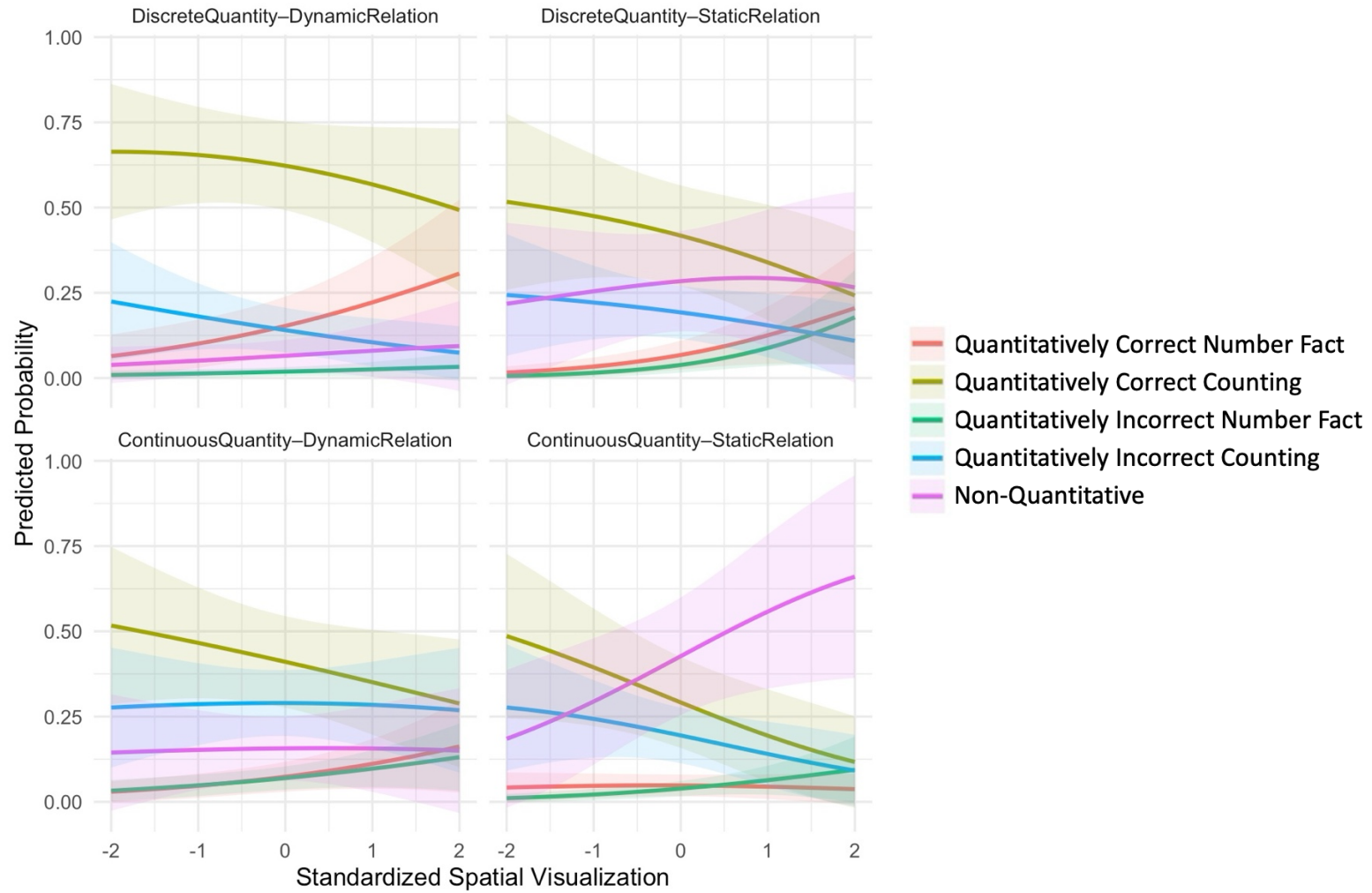
Three interaction terms approached conventional levels of statistical significance, two of which involved Continuous Quantity–Static Relation problems. For ease of interpretation, conditional odds ratios were calculated by multiplying main and interaction effects. There was a weak evidence that the association between spatial visualization and the likelihood of selecting a ‘Non-Quantitative’ (OR = 1.67, 95% CI [0.95, 2.94], $p = .07$) and a ‘Quantitatively Incorrect Number Fact’ strategy (OR = 1.90, 95% CI [0.99, 3.65], $p = .055$), relative to ‘Quantitatively Correct Number Fact,’ varied by problem type. The corresponding conditional odds ratio ($0.85 \times 1.67 \approx 1.42$; $0.93 \times 1.90 \approx 1.77$, respectively) suggest that, for Continuous Quantity–Static Relation problems, higher spatial visualization may be associated with increased odds of selecting these strategies.

A third interaction, observed for Discrete Quantity–Static Relation problems, suggested weak evidence that the association between spatial visualization and the likelihood of selecting a ‘Quantitatively Correct Counting’ strategy (relative to ‘Quantitatively Correct Number Fact’) varied by problem type (OR = 0.70, 95% CI [0.47, 1.03], $p = .07$). The corresponding conditional odds ratio ($0.63 \times 0.70 \approx 0.44$) suggests that for this problem type, higher spatial visualization may be associated with lower odds of using this strategy.

These effects should be interpreted with caution given their marginal statistical evidence and the number of interaction parameters estimated. Figure 4-6 presents predicted probabilities of each strategy across the observed range of spatial visualization for each problem type. Although the overall pattern of associations appears broadly similar—higher spatial visualization being associated with greater use of number-fact strategies and decreased use of the counting-based strategies—the magnitude of these associations varies modestly across problem types. These visual patterns are consistent with the omnibus interaction and the relatively small individual interaction coefficients.

Figure 4-6

Predicted Probabilities of Problem-Solving Strategies Use across Spatial Visualization by Word Problem Type



4.4.2.2.2 *Visuospatial Working Memory and Problem-Solving Strategy by Problem Type*

A pseudo-likelihood-ratio comparison between the full interaction model and a reduced model excluding the visuospatial working memory $\times$ problem type interaction term was significant, $\chi^2(12) = 35.37$, $p < .001$, indicating that the association between visuospatial working memory and strategy use differed across problem types. Several interaction terms reached or approached statistical significance, suggesting that the moderation may vary across strategy contrasts. Relative to the spatial visualization interaction, these effects appeared more localized to specific contrasts. Figure 4-7 presents predicted probabilities across the observed range of visuospatial working memory for each problem type.

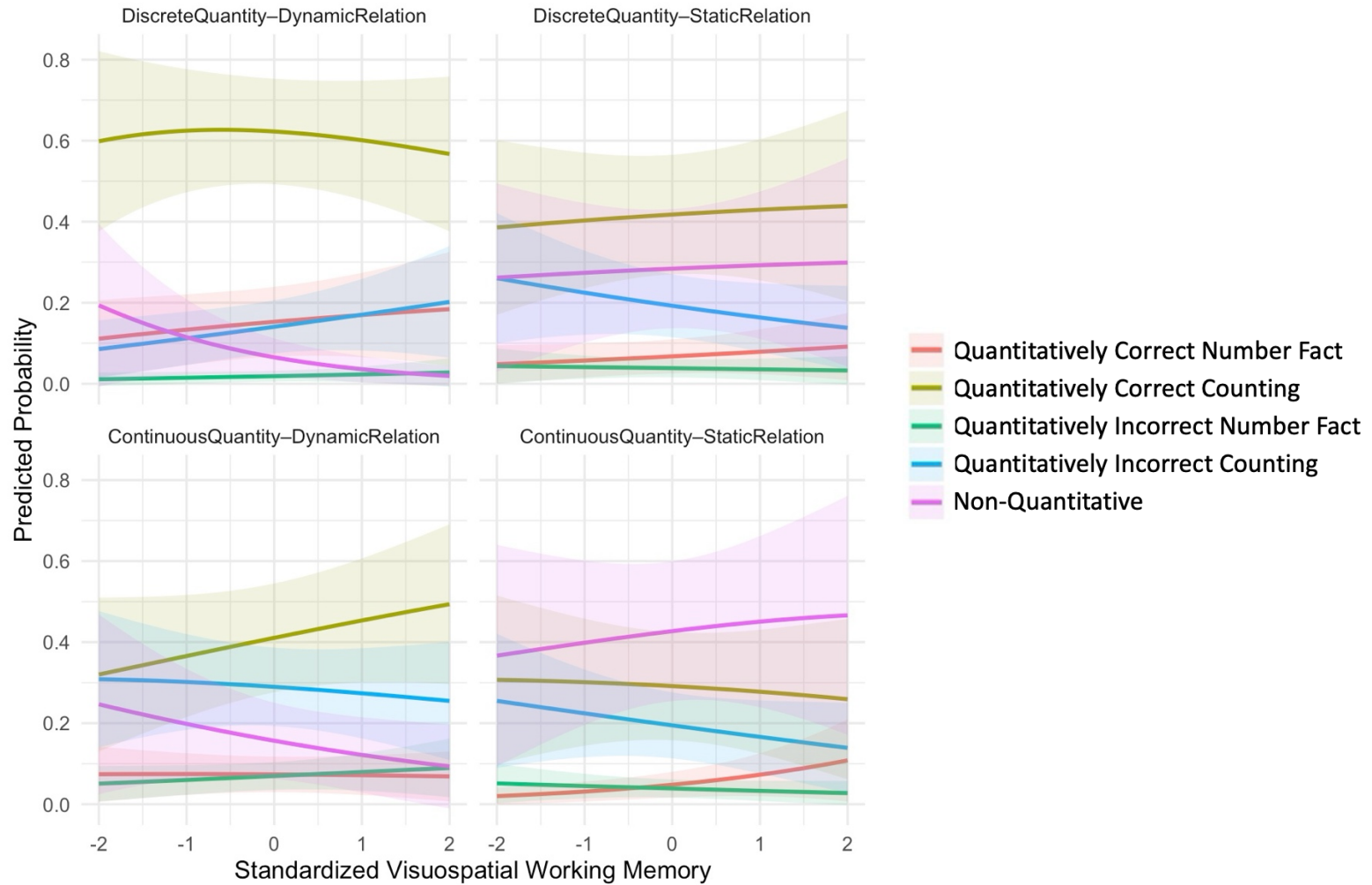
For the reference Discrete Quantity–Dynamic Relation problems, higher visuospatial working memory was associated with lower odds of using a ‘Non-Quantitative’ strategy relative to a ‘Quantitatively Correct Number Fact’ strategy (OR = 0.50, 95% CI [0.28, 0.87], $p = .01$). However, this association varied by problem type. For Discrete Quantity–Static Relation problems, the interaction term (OR = 1.78, 95% CI [1.10, 2.88], $p = .02$) yielded a conditional odds ratio of 0.89 (1.78×0.50), suggesting a weaker negative association. A similar pattern was observed for Continuous Quantity–Dynamic Relation problems (OR = 1.61, 95% CI [1.00, 2.60], $p = .05$), with a conditional odds ratio of 0.81 (1.61×0.50).

For the reference Discrete Quantity–Dynamic Relation problems, visuospatial working memory was not significantly associated with the likelihood of selecting a ‘Quantitatively Incorrect Counting’ (OR = 1.09, 95% CI [0.69, 1.72], $p = .70$), or ‘Quantitatively Incorrect Number Fact’ (OR = 1.11, 95% CI [0.59, 2.09], $p = .75$), or ‘Quantitatively Correct Counting’ strategy (OR = 0.87, 95% CI [0.60, 1.25], $p = .45$), relative to the baseline ‘Quantitatively Correct Number Fact’ strategy. However, a different pattern was observed for Continuous Quantity–Static Relation problems.

For this problem type, higher visuospatial working memory was associated with lower odds of using quantitatively-incorrect strategies and ‘Quantitatively Correct Counting’ strategies relative to ‘Quantitatively Correct Number Fact’. Specifically, the interaction for ‘Quantitatively Incorrect Counting’ was significant (OR = 0.52, 95% CI [0.32, 0.84], $p = .007$), yielding a conditional odds ratio of 0.57 (0.52×1.09), indicating lower odds of using this strategy.

Figure 4-7

Predicted Probabilities of Problem-Solving Strategies Use across Visuospatial Working Memory by Word Problem Type



The interaction for ‘Quantitatively Incorrect Number Fact’ approached statistical significance (OR = 0.51, 95% CI [0.25, 1.04], $p = .065$), producing a similar conditional odds ratio of 0.57 (1.11×0.51), suggesting a comparable tendency toward reduced use of this strategy. In addition, the interaction for ‘Quantitatively Correct Counting’ was marginal (OR = 0.72, 95% CI [0.50, 1.05], $p = .09$), yielding a conditional odds ratio of 0.63 (0.87×0.72), again indicating lower odds relative to the baseline category.

In sum, the results provide modest evidence that word problem type may moderate the relationship between spatial skills and strategy use. Spatial visualization exhibited a small and distributed moderating pattern, whereas visuospatial working memory showed more pronounced, problem-type-specific moderation for particular strategy contrasts. All interaction effects should be interpreted with caution given the number of comparisons and the modest magnitude of the observed effects.

4.5 Discussion

The present study sought to advance understanding of the spatial–mathematical link in two ways. First, it aimed to make a theoretical contribution by proposing a holistic framework for categorizing problem-solving strategies that considers both arithmetic strategies and representations of quantitative structure. Second, it aimed to make an empirical contribution by examining the relationship between lower primary school pupils’ (6–7-year-olds’) spatial visualization and visuospatial working memory and their use of problem-solving strategies in one-step additive word problems. Together, this study represents an initial step toward identifying problem-solving strategies as a potential target for future investigation into the association between young pupils’ spatial ability and arithmetic word problem solving.

The discussion begins with the findings related to RQ3, focusing on the association between spatial skills and problem-solving strategy use ([Subsection 4.5.1](#)). It then considers RQ4, examining whether and how word problem type moderates these associations ([Subsection 4.5.2](#)). Next, the section discusses the study’s limitations ([Subsection 4.5.3](#)), and concludes with suggestions for future research ([Subsection 4.5.4](#)).

4.5.1 RQ3: Interpreting the Relationship between Spatial Skills and Problem-Solving Strategy Use

Across all word problem types, higher spatial visualization was associated with greater relative use of ‘Quantitatively Correct Number Fact’ strategies compared with both ‘Quantitatively Incorrect Counting’ and ‘Quantitatively Correct Counting’ strategies, but not

compared with ‘Quantitatively Incorrect Number Fact’ or ‘Non-Quantitative’ strategies. In contrast, visuospatial working memory did not uniquely predict differences in problem-solving strategy use. This finding aligns with prior research suggesting that spatial visualization is more consistently associated with mathematical problem solving than visuospatial working memory (e.g., LeFevre et al., 2010; Li & Geary, 2013; Mix et al., 2016). In light of these findings, the following subsections discuss the association between spatial skills and arithmetic strategies (Subsection 4.5.1.1) and the association between spatial skills and the representation of quantitative structure (Subsection 4.5.1.2).

4.5.1.1 Spatial Skills and Arithmetic Strategy

The findings indicate that spatial visualization differentiates between counting-based and number-fact strategies when solving arithmetic word problems. Specifically, higher spatial visualization was associated with reduced reliance on both counting-based strategies relative to ‘Quantitatively Correct Number Fact’ strategies. This pattern aligns with prior findings linking spatial visualization to the use of higher-level arithmetic strategies (e.g., Laski et al., 2013; Möhring, 2024). Importantly, this pattern of associations was specific to spatial visualization, with fluid reasoning being associated with reduced reliance on ‘Quantitatively Incorrect Counting’ strategies, but not with the use of ‘Quantitatively Correct Counting’ strategies.

In contrast, visuospatial working memory was not related to increased use of ‘Quantitatively Correct Number Fact’ strategies, a finding that extends previous work suggesting a role for visuospatial working memory (Foley et al., 2017). One possible explanation for the differing patterns in the association between visuospatial working memory and arithmetic strategies is that Foley and colleagues (2017) controlled only for age and verbal working memory, whereas the present study included spatial visualization and other important cognitive covariates, including language ability, numerical ability and fluid reasoning. It is possible that the association between visuospatial working memory and strategy use reflects variance shared with spatial visualization or other cognitive abilities, such that its unique contribution diminished once these factors are taken into account.

An alternative explanation is that Foley and colleagues (2017) focused on derived number-fact strategies, whereas the present study grouped both derived and recalled number-fact strategies into a single category. Some researchers suggest that these two number-fact strategies may be related but not identical, particularly because deriving number facts involve

a series of analytical steps and active problem solving (Casey & Fell, 2018). For instance, studies indicate that children who use derived number-fact strategies demonstrate a deeper understanding of addition and subtraction, including fundamental properties such as commutativity and associativity (McNeil et al., 2025). Visuospatial working memory may therefore be more closely related to derived number-fact strategies, which rely on the storage of intermediate information during problem solving.

A related explanation concerns that Foley and colleagues' (2017) study and the present study differ in the type of visuospatial working memory and mathematics assessed. Foley et al. (2017) measured *simultaneous* visuospatial working memory (i.e., all information presented at the same time) and *computation tasks*, whereas the present study measured *sequential* visuospatial working memory (i.e., information presented sequentially) and *word problems*. A systematic review by Allen et al. (2019) suggests that visuospatial working memory may be more strongly associated with mathematical reasoning (e.g., word problems) than numerical operations (e.g., computation tasks), and provides some indication that different types of visuospatial working memory may relate differently to mathematical performance. This raises the possibility that sequential visuospatial working memory may be differentially associated with specific word problem types, and that grouping these problems may obscure such variation. However, the evidence remains limited and inconclusive. In particular, Allen et al. (2019) note that, no studies have examined “the influence of sequential visuospatial working memory on mathematical reasoning, highlighting a gap in the research” (p. 523). Thus, the more nuanced account of the association between sequential visuospatial working memory and word problem solving offered in the present study warrants further investigation.

The transition from counting-based to number-fact strategies represents a key developmental progression in early arithmetic (Jordan et al., 2008; Krenger & Thevenot, 2025). Number-fact strategies are generally considered more advanced and efficient than counting-based strategies (Casey & Fell, 2018). Importantly, research suggests that this transition becomes particularly important after approximately age 6.5: before this age, children who rely on finger counting tend to outperform those who do not, whereas after this age, this pattern reverses (Krenger & Thevenot, 2025). The pupils in the present study (6.52–7.74 years) therefore fall within this critical developmental window in which shifts toward number-fact strategies are especially consequential. In this context, the observed association between spatial visualization and greater use of number-fact strategies suggests that individual differences in spatial visualization are related to variation in arithmetic strategy use. This, in turn, may help

explain individual differences in problem-solving accuracy, which represents an important avenue for future investigation.

Taken together, these findings suggest that spatial visualization is associated with differences in arithmetic strategy use during a developmentally significant transition period from counting-based to number-fact strategies. Although the present data do not permit causal conclusions, the observed association suggests that spatial visualization may be meaningfully linked to the use of more advanced arithmetic strategies in word problem solving. Future research should investigate whether number-fact strategy use may act as a statistical mediator of the association between spatial visualization and word-problem-solving performance, ideally using longitudinal or experimental designs.

4.5.1.2 Spatial Skills and Representation of Quantitative Structure

Beyond its association with greater relative use of number-fact strategies, spatial visualization may also be linked to pupils' representations of quantitative structures. Spatial visualization—but not visuospatial working memory—uniquely predicted greater use of 'Quantitatively Correct Number Fact' strategies compared to 'Quantitatively Incorrect Counting' strategies, over and above maturational effects reflected in age, as well as cognitive abilities known to support arithmetic fluency, such as numerical ability and fluid reasoning (Daroczy et al., 2015; Fung & Swanson, 2017; Verschaffel et al., 2020; Vessonen et al., 2024). The association is not merely attributable to *storage capacity* or general *arithmetic proficiency*, as visuospatial working memory did not independently predict strategy use (Baddeley & Hitch, 1974), and numerical ability was controlled for in the analysis.

Instead, the contrast between 'Quantitatively Incorrect Counting' and 'Quantitatively Correct Number Fact' strategies may reflect not only a shift in arithmetic strategy (i.e., from counting-based to number-fact approaches) but also differences in how pupils represent quantitative structures. Spatial visualization may function as a "mental blackboard" (Hawes & Ansari, 2020, p. 472)—a representational space in which quantitative structure can be coordinated and transformed. In that sense, spatial visualization may facilitate the construction and manipulation of part-whole quantitative structure in one-step additive word problems, as well as the flexible transformation and segmentation of quantities and numerical magnitudes (Mix & Cheng, 2012; Möhring et al., 2024).

Importantly, the association between spatial visualization and problem-solving strategies showed a distinct pattern compared to the other cognitive covariates. All cognitive

covariates (i.e., language ability, numerical ability, and fluid reasoning), as well as spatial visualization, were associated with reduced use of ‘Non-Quantitative’ strategies, which is unsurprising. However, the covariates differed in how they related to quantitatively-incorrect and counting-based strategies: language ability was associated with reduced use of ‘Quantitatively Incorrect Number Fact’ strategies, whereas numerical ability and fluid reasoning were associated with reduced use of ‘Quantitatively Incorrect Counting’ strategies. In contrast, spatial visualization was uniquely associated with reduced use of both ‘Quantitatively Correct Counting’ and ‘Quantitatively Incorrect Counting’ strategies. This pattern suggests that spatial visualization may support the representation of quantitative structure in a way that reduces reliance on counting-based strategies.

However, the finding that spatial visualization predicted reduced use of ‘Quantitatively Incorrect Counting’ but not ‘Quantitatively Incorrect Number Fact’ strategies warrants careful consideration. At first glance, this appears inconsistent with prior research linking spatial visualization to quantitative reasoning (e.g., Boonen et al., 2013, 2014; de Koning et al., 2022; Hegarty et al., 1995; Ng & Lee, 2009; Van Garderen, 2006). However, much of this research has been conducted with older pupils within sequential models of problem solving that distinguish ‘representation’ and ‘computation’ as separate phases (Mayer, 1985).

One possible explanation is that ‘Quantitatively Incorrect Number Fact’ strategies reflect surface-level operation selection—essentially ‘guessing’ an operation based on the numbers present without representing the problem’s underlying quantitative structure. In this respect, they resemble ‘Non-Quantitative’ strategies in failing to engage meaningfully with the problem’s quantitative structure, despite appearing numerically oriented. This interpretation is consistent with the absence of an association between spatial visualization and reduced use of ‘Non-Quantitative’ strategies.

Alternatively, the null effect for ‘Quantitatively Incorrect Number Fact’ strategies may reflect restricted between-pupil variability. Although these strategies accounted for approximately 5% of responses, their occurrence was highly uneven across pupils, with a small subset of children contributing a disproportionate number of instances, as indicated by a high CV. While the mixed-effects models appropriately accounted for pupil-level variability through random effects (Snijders & Bosker, 2012), the uneven distribution of strategy use may have left limited systematic between-pupil variance for the model to explain. When variance is restricted, the statistical power to detect associations with pupil-level predictors (i.e., spatial visualization) is correspondingly reduced (Maas & Hox, 2005).

Taken together, these findings suggest that spatial visualization may be linked to the representation of quantitative structure in word problem solving. However, the absence of a detectable association with reduced use of ‘Quantitatively Incorrect Number Fact’ strategies highlights the complexity of this relationship and warrants cautious interpretation. It remains unclear whether this reflects a genuine theoretical distinction—positioning ‘Quantitatively Incorrect Number Fact’ strategies as a superficial approach akin to ‘Non-Quantitative’ strategies—or whether it reflects restricted between-pupil variability or other unmeasured factors. Future research with larger or more targeted samples is needed to clarify whether this null effect (i.e., the comparison between ‘Quantitatively Incorrect Counting’ and ‘Quantitatively Incorrect Number Fact’ strategies) reflects substantive differences in the representation of quantitative structure, limited statistical power, or alternative explanations.

In summary, the findings related to RQ3 indicate that spatial visualization plays a distinct role in shaping young pupils’ problem-solving strategy use. Across word problem types, higher spatial visualization was associated with greater relative use of number-fact strategies during a developmentally significant period in which transitions in arithmetic strategy use are especially consequential. This association appears to extend beyond general arithmetic proficiency and storage capacity. Moreover, the results provide modest evidence that spatial visualization is linked not only to the adoption of more advanced arithmetic strategies but also to how pupils construct and manipulate representations of a problem’s quantitative structure during word problem solving. In contrast, visuospatial working memory did not uniquely predict differences in strategy use once other cognitive factors were controlled.

4.5.2 RQ4: Interpreting Interaction Effects of Spatial Skills and Problem Type on Problem-Solving Strategy Use

Having established the overall association between spatial skills and strategy use, the discussion now turns to whether these associations vary across word problem types. Across most problem types, the pattern was broadly consistent with RQ3: higher spatial visualization was associated with greater use of ‘Quantitatively Correct Number Fact’ strategies relative to counting-based strategies, regardless of the correctness of the quantitative structure representation. In contrast, visuospatial working memory—which did not uniquely predict strategy use in the overall RQ3 model—was associated with reduced use of ‘Non-Quantitative’ strategies when variation across problem types was taken into account. Additionally, a different pattern emerged for Continuous Quantity–Static Relation problems, with visuospatial working

memory, rather than spatial visualization, showing a more consistent association with the use of ‘Quantitatively Correct Number Fact’ strategies.

Taken together, these findings provide modest evidence that the relative contributions of spatial visualization and visuospatial working memory vary as a function of the quantitative structure of the problems. These patterns should be interpreted in light of descriptive differences across problem types, whereby discrete-quantity and dynamic-relation problems were generally associated with greater use of quantitatively-correct strategies and lower use of ‘Non-Quantitative’ strategies than continuous-quantity and static-relation problems. First, [Subsection 4.5.2.1](#) discusses how, for most problem types, the roles of spatial visualization and visuospatial working memory appear to be complementary: spatial visualization differentiates between ‘Quantitatively Correct Number Fact’ strategies and counting-based strategies, whereas visuospatial working memory reduces reliance on ‘Non-Quantitative’ strategies. [Subsection 4.5.2.2](#) then examines how, for Continuous Quantity–Static Relation problems, this pattern shifts, with visuospatial working memory showing a more consistent association with the use of ‘Quantitatively Correct Number Fact’ strategies.

4.5.2.1 Patterns across Problem Types

Across Discrete Quantity–Dynamic Relation, Discrete Quantity–Static Relation, and Continuous Quantity–Dynamic Relation problems, spatial visualization and visuospatial working memory showed complementary patterns of association with problem-solving strategy use. Higher spatial visualization was associated with lower relative use of counting-based strategies, whereas higher visuospatial working memory was associated with lower relatively use of ‘Non-Quantitative’ strategies. This pattern suggests that the two spatial skills support distinct functional components of the problem-solving process, consistent with prior evidence that they are separable constructs, each with independent pathways to mathematical abilities (Hawes & Ansari, 2020; Mix et al., 2016, 2017; Mix, 2019).

One account proposes that spatial visualization supports the spatial modeling of *novel* content, whereas visuospatial working memory supports the procedural fluency of *familiar* content (Mix, 2019). The present findings offer partial support for this characterization. Notably, visuospatial working memory was most strongly associated with reduced use of ‘Non-Quantitative’ strategies for Discrete Quantity–Dynamic Relation problems, which are more familiar quantitative structures for children in this age group (Carpenter & Moser, 1984; Riley et al., 1983). This familiarity is reflected in the frequency with which this problem type appears

in the optional Key Stage 1 National Curriculum Assessments (see [Appendix 4C](#) for samples of this word problem type from the 2018–2025 exam papers, excluding 2020 and 2021). Although this association was present across other problem types, it was attenuated, suggesting that visuospatial working memory may indeed be especially recruited when children solve problems with more familiar quantitative structure.

However, this familiarity-based account does not fully explain the role of spatial visualization. Although its association with reduced reliance on counting-based strategies was broadly consistent across word problem types, the magnitude of this association varied modestly. For example, a marginally stronger effect was observed for Discrete Quantity–Static Relation problems, which may be considered less prototypical in early instructional contexts (see [Appendix 4D](#) for samples of this word problem type from the 2018–2025 Key Stage 1 National Curriculum Assessments, excluding 2020 and 2021). Yet this variation did not follow a consistent familiarity gradient comparable to that observed for visuospatial working memory. Notably, Continuous Quantity–Dynamic Relation problems are also relatively less frequently encountered in early instruction, but no corresponding interaction effect was observed (see [Appendix 4E](#) for samples of this word problem type). Thus, while spatial visualization exhibited some sensitivity to variation in quantitative structure, its influence does not appear to be systematically organized around familiarity *per se*.

This observation invites consideration of a second account, which proposes that spatial visualization supports the *coordination* and *transformation* of representations, whereas visuospatial working memory supports their *maintenance* (Hawes & Ansari, 2020). The present findings are broadly consistent with this distinction but suggest an important refinement. In particular, the marginally stronger association between spatial visualization and reduced reliance on ‘Quantitatively Correct Counting’ strategies in Discrete Quantity–Static Relation problems may reflect the specific representational demands of this word problem type.

That is, static-relation problems require representing comparative relations between disjoint quantities simultaneously (e.g., more/less marbles or farther away/closer to reference) within a coherent representation, which may not be readily reducible to stepwise temporal simulation. In contrast, dynamic-relation problems (e.g., gain/loss or movement toward/away) afford sequential mental simulation, allowing the representation to unfold procedurally over time and making counting-based strategies structurally viable. Consistent with this interpretation, prior research has shown that children more readily use direct modeling and counting-based strategies for dynamic-relation (i.e., change) problems than for static-relation

(i.e., comparison) problems (Carpenter & Moser, 1984; Riley et al., 1983). Under this interpretation, spatial visualization may be especially beneficial when successful problem solving depends on constructing a representation of quantitative structure, whereas visuospatial working memory may play a stabilizing role by maintaining that model during problem solving.

4.5.2.2 Continuous Quantity–Static Relation Problems: A Distinct Pattern

Continuous Quantity–Static Relation problems showed a distinct pattern of associations. Although spatial visualization continued to be associated with reduced reliance on counting-based strategies, its advantage over ‘Quantitatively Incorrect Number Fact’ and ‘Non-Quantitative’ strategies appeared attenuated, with marginal tendencies toward increased relative odds of these strategies. In contrast, for this problem type, visuospatial working memory was more consistently associated with reduced use of quantitatively-incorrect and counting-based strategies. This pattern suggests that successful strategy selection in this problem type may depend more heavily on visuospatial working memory than spatial visualization. Given the modest magnitude of the interaction effects, the interpretations that follow should be considered exploratory.

The attenuated protective effect of spatial visualization for this problem type is particularly striking given the inherent spatial nature of this problem type. These problems require reasoning about the relative positions of two people with respect to a reference point (e.g., farther from or closer to a wall), situations that are fundamentally spatial in nature (Mix & Cheng, 2012). If spatial visualization primarily supports the construction of spatial representations (Hawes & Ansari, 2020), one might expect its influence to be especially pronounced in this context, akin to findings in older pupils’ performance on spatial–visual problems (Boonen et al., 2014). Moreover, although prior research has demonstrated that other cognitive variables—such as language ability—account for substantial variance in word-problem performance (Spencer et al., 2022), the structural demands of this problem type appear to be deeply rooted in spatial reasoning. The attenuation of spatial visualization’s protective effect in such a spatially grounded context therefore warrants further explanation.

One possible explanation for the attenuated effect of spatial visualization is that the presence of countable units may have induced a “counting bias,” shifting attention away from the continuous-quantity structure and toward salient numerical information (Vasilyeva et al., 2025, p. 1161). For example, Jeong and colleagues (2007) found that young children (6-, 8- and 10-year-olds) were able to judge proportions in the context of a probability reasoning task

under continuous conditions (i.e., when the target and non-target regions were not divided) by relying on the relative amounts of target and nontarget elements. However, under discrete conditions (i.e., when the target and non-target regions were divided into same-sized subregions), children adopted an erroneous counting strategy, focusing on the number of target elements while neglecting the relational proportion between target and non-target elements. In other words, when discrete units are available, children tend to focus on them rather than on the underlying spatial-relational information, even when that relational information is more central to solving the problem. This tendency to prioritize discrete, countable units over relational magnitude information is consistent with research demonstrating a developmental bias toward whole-number and discrete representations in early numerical cognition (Siegler & Booth, 2004; Opfer & Siegler, 2007).

In the present context, the imposed discrete units in Continuous Quantity–Static Relation problems may similarly have drawn attention toward countable elements at the expense of relational structure, thereby diminishing the role of continuous quantity, as “they are presented in such a way that there is no consideration for the possibility of taking anything other than a single one as a unit” (Sophian, 2008, p. 61). As discussed earlier, ‘Quantitatively Incorrect Number Fact’ strategies may reflect superficial operation selection—applying an operation based on salient numerical cues without constructing a coherent representation of the problem’s quantitative structure. Although spatial visualization may facilitate the coordination and transformation of such representation, the coexistence of continuous quantities and discrete cues may introduce representational conflict, thereby increasing the likelihood of superficial responses, including ‘Quantitatively Incorrect Number Fact’ and ‘Non-Quantitative’ strategies.

At the same time, the distinct pattern observed for visuospatial working memory may reflect increased maintenance demands. Unlike Continuous Quantity—Dynamic Relation problems, which involve transformation of a single position relative to the reference, Continuous Quantity—Static Relation problems require simultaneously tracking the relative positions of two entities (see [Appendix 4F](#) for samples of this word problem type). Maintaining multiple spatial relations may impose greater working memory demands (Baddeley & Hitch, 1974). In such contexts, the ability to maintain a representation of the quantitative structure may be especially critical. External visual representations can reduce cognitive load and support relational reasoning (Boonen et al., 2013; Jitendra & Woodward, 2019), yet few pupils in the present study spontaneously generated such representations. As a result, pupils with

stronger visuospatial working memory may have been better equipped to maintain the internal representation, which enabled them to use ‘Quantitatively Correct Number Fact’ strategies.

Taken together, Continuous Quantity–Static Relation problems may place heightened demands on maintaining a representation of the quantitative structure while simultaneously increasing the risk of surface-level strategy use due to discrete cues. Under such conditions, spatial visualization may continue to support coordination and transformation of representations, but its effectiveness may depend, in part, on maintenance of the representation—a function more closely associated with visuospatial working memory. Given the modest magnitude of the interaction effects, these interpretations should be considered exploratory. Future research is needed to determine whether this pattern reflects genuine structure-sensitive differences in the roles of spatial skills based on the quantitative structure of the problems or is a byproduct or contextual (e.g., educational or cultural), statistical, or measurement limitations.

In summary, the findings for RQ4 suggest that the association between spatial skills and problem-solving strategy use is sensitive to differences in the quantitative structure of word problems, though the magnitude of these differences was generally modest. For most quantitative structures, spatial visualization and visuospatial working memory showed complementary patterns of association. For Continuous Quantity–Static Relation problems, however, this pattern shifted, with visuospatial working memory showing a more consistent association with ‘Quantitatively Correct Number Fact’ strategy use. Further research is needed to clarify the robustness and theoretical implications of these quantitative-structure-dependent patterns.

4.5.3 Limitations

Despite the present study’s theoretical and methodological strengths, several limitations should be acknowledged. First, several limitations relate to the scope of data collection. Although this study controlled for age and gender, it did not include other potentially important demographic variables such as socioeconomic status or information on pupils’ SEN and EAL status (Casey et al., 2011; Verschaffel et al., 2020). It also did not account for other potentially relevant factors, such as emotional factors (e.g., confidence and anxiety), which are known to mediate the spatial–mathematical link (Casey & Ganley, 2021; Mayer, 1998), or teachers’ instructional practices, which may influence the prevalence of particular problem-solving strategies (Jitendra & Woodward, 2019). Although the mixed-

effects design partially accounted for classroom-level differences, future research should consider additional individual factors as well as teachers' instructional practices.

Second, several limitations arise from measurement and coding decisions. Because no video, audio, or photographic recordings were made during data collection, the coding of problem-solving strategies was based on the researcher's field notes. This decision was made to minimize potential participant discomfort and to avoid reducing the sample size or introducing systematic bias in participation. Although detailed notes were taken (e.g., precise drawings of pupils' use of MathLink cubes and verbatim notes of pupils' retrospective reports), the possibility of human error cannot be completely eliminated. High interrater reliability mitigates concerns about coding consistency, but the absence of multimedia recordings and reliance on field notes may limit the precision of strategy classification.

In addition, the granularity of the strategy classification may have constrained the sensitivity of the analyses. The study focused broadly on the use of number-fact versus counting-based strategies, and no fine-grained distinctions were made between the different subtypes of these strategies. For example, derived number facts and recalled number facts were aggregated into a single category because they are both advanced strategies using internal mental solutions (Casey et al., 2017), and it was not always possible to reliably distinguish between them based on behavioral observations and pupils' retrospective reports. However, prior research suggests that spatial ability may be more strongly associated with deriving number facts than with recalling number facts (e.g., Casey et al., 2015, 2017; Foley et al., 2017; Möhring et al., 2024) and that deriving number facts may represent a more adaptive strategy because it involves active decision making and problem solving (Casey et al., 2017). Aggregating these subtypes may therefore have obscured potentially differential associations between spatial skills and specific forms of number-fact strategy use.

Similarly, all counting-based strategies were treated as a single category. Yet research indicates that spatial ability may relate more strongly to advanced counting-based strategies (e.g., count-on) than to more rudimentary ones (e.g., count-all) (e.g., Laski et al., 2013). A more fine-grained categorization of counting-based strategies may therefore reveal additional structure-sensitive patterns.

Measurement of response time may have helped differentiate between problem-solving strategies, as they are often associated with different processing speeds (Siegler, 1987). Although observable behavior and retrospective reports were used to infer strategies, the lack

of latency data limits the ability to triangulate strategy classification using temporal indicators. Instead, as noted above, detailed field notes were taken during word problem solving to ensure accurate documentation of children's responses and strategies.

Furthermore, the present study did not distinguish explicitly between 'addition' and 'subtraction' word problems. Prior research suggests that direct number-fact strategies are used less frequently in subtraction than in corresponding addition tasks (Barrouillet et al., 2008; Robinson, 2001). However, in the present study, all word problems involving part-whole relations were grouped together as additive problems, because pupils may conceptualize and solve them in different ways—some interpreting a problem as requiring addition, others as requiring subtraction. For example, given the same problem (e.g., $3 + __ = 7$), one pupil may approach it using a count-on strategy (i.e., 'How much must be added to 3 to obtain 7?'), whereas another may conceptualize it using a count-down approach (i.e., 'How much remains after taking 3 away from 7?'). Collapsing across these solution interpretations may have introduced additional heterogeneity into the strategy categories.

Third, limitations relate to the study's correlational, cross-sectional design. The present study represents an initial step toward identifying potential links between spatial skills and problem-solving strategy use. Longitudinal or experimental research is needed to determine whether spatial skills prospectively influence strategy use and whether differences in strategy use precede improvements in performance (Bailey, 2017).

Finally, several limitations relate to the magnitude and statistical robustness of the observed effects. The use of spline terms allows for flexible modelling of non-linear relationships; however, results may be sensitive to modelling choices (e.g., knot placement and degrees of freedom), which could influence their replicability. Additionally, the moderation effects observed in the present study were generally modest in magnitude. While the patterns were theoretically interpretable, any differences in word problem types in terms of spatial-strategy associations should be interpreted with caution and require replication with larger samples in future research.

4.5.4 Conclusion

The present study serves as an initial step toward identifying problem-solving strategies as a potential target for future investigation into the association between lower primary school pupils' spatial skills and arithmetic word problem solving. It contributes valuable insights by proposing a holistic problem-solving framework that integrates the dimensions of arithmetic

strategies and representations of quantitative structure. It also provides a more fine-grained analysis of this association, distinguishing between different spatial skills (i.e., spatial visualization and visuospatial working memory) and word problem types (i.e., based on varying quantitative structures).

Across word problem types, spatial visualization was consistently associated with variation in arithmetic strategy use during a developmentally significant transition from counting-based to number-fact strategies, as well as with accurate representations of the problem's quantitative structure. In contrast, higher visuospatial working memory was associated with reduced use of 'Non-Quantitative' strategies across most problem types; for comparison problems between disjoint quantities measured using an imposed unit, it was instead more consistently associated with reduced use of quantitatively-incorrect and counting-based strategies. Although the observed moderation effects were modest, they indicate that the relationship between spatial skills and problem-solving strategy use may be sensitive to differences in quantitative structure. Taken together, the findings suggest that spatial visualization and visuospatial working memory may be differentially recruited depending on the demands of the arithmetic word problem—specifically, whether representations of the problem's quantitative structure need to be coordinated and transformed or maintained. At the same time, the correlational, cross-sectional design and modest interaction magnitudes underscore the need for cautious interpretation and further investigation.

By identifying potential structure-sensitive links between spatial skills and problem-solving strategy use, this study provides valuable insight into the spatial–mathematical link in arithmetic word problem solving in the early primary years. Extending the present findings, future research should test whether problem-solving strategies serve as mediators of the spatial–mathematical association. Moreover, longitudinal and experimental designs are needed to clarify the directionality of the relationships observed. Replication across diverse samples and with more fine-grained strategy classifications may further shed light on this relationship.

5 Study 3: Young Children’s Problem-Solving Strategies as a Mediator between Spatial Skills and Arithmetic Word-Problem-Solving Accuracy

Abstract

Despite increasing interest in the mechanisms underlying the spatial–mathematical link, few studies have examined the processes that mediate the relationship between spatial ability and young pupils’ one-step additive word problem solving. The present study investigates whether ‘*Quantitatively Correct Number Fact*’ strategy use—a mental arithmetic strategy relying on the accurate representation of a problem’s quantitative structure—mediates the relationship between spatial skills (spatial visualization and visuospatial working memory) and problem-solving accuracy. Using path analyses with data from 137 Year 2 pupils (6–7-year-olds) in England and controlling for key covariates (language ability, numerical ability, fluid reasoning, age, and gender), this study found that the indirect effects of spatial skills differed by word problem type. The findings indicate that spatial visualization has an indirect effect on arithmetic word-problem-solving performance, mediated through problem-solving strategy use, for word problem types that place greater demands on coordinating and transforming the representation of the problem’s quantitative structure. In contrast, visuospatial working memory has an indirect effect for word problem types that place greater demands on maintaining these representations.

5.1 Introduction

Spatial ability, which concerns “the locations of objects, their shapes, their relations to each other, and the paths they take as they move” (Newcombe, 2017, p. 6), has a strong connection with mathematical achievement (for a review, see Lourenco et al., 2018; Mix, 2019; Young et al., 2018). The spatial–mathematical link may be “one of the most robust and well-established findings in cognitive psychology” (Mix & Cheng, 2012, p. 198). Cross-sectional studies have established their associative relationship (for a review, see Cui & Guo, 2022). Cross-lagged panel models in longitudinal studies indicate that early spatial ability predicts later mathematical achievement (e.g., Fung et al., 2020; Kahl et al., 2022) and that this relationship may be bidirectional (Geer et al., 2019; Lombardi et al., 2019). Experimental studies of spatial training further suggest a far-transfer effect to mathematical performance (i.e., improvements in mathematical performance following spatial training; for a recent meta-analysis, see Hawes et al., 2022).

However, there is a limited understanding of *how* spatial ability and mathematical achievement are related (Lowrie et al., 2020; Mix, 2019; Mix & Cheng, 2012), which could, in turn, inform *how* the relationship can be better leveraged in education. Although experimental methods are required for establishing causal mechanisms (Bailey, 2017), examining potential mediators cross-sectionally is an important first step in informing the design and target of subsequent experimental work. Arithmetic word problem solving in the early primary years remains an overlooked area within mechanistic accounts of the spatial–mathematical link, partly due to the challenges posed by the diversity of word problems and spatial taxonomies (for a review, see [Subsection 3.2](#) of Study 1).

This study aims to contribute to addressing this gap in the literature by examining whether young pupils’ (6–7-year-olds’) spatial skills are related to arithmetic word-problem-solving accuracy through their usage of problem-solving strategies. Specifically, this study adopts the *one-step additive word problems* from [Study 1](#), which can be solved with a single addition or subtraction calculation. It focuses on two spatial skills that have been consistently identified as having a particularly strong association with mathematical thinking: *spatial visualization*, “the ability to generate, recall, maintain, and manipulate visual-spatial images and solutions” (Hawes & Ansari, 2020, p. 466), and *visuospatial working memory*, the ability to temporarily maintain and process visuospatial information (Allen et al., 2019).

This empirical investigation is important for both theoretical and practical reasons. Theoretically, it advances understanding of the mechanisms underlying the relationship between children's spatial skills and mathematical performance. Practically, it provides insights into how the spatial–mathematical link can be leveraged to inform mathematical instruction.

5.2 Literature Review

[Subsection 5.2.1](#) reviews literature on the relationship between spatial ability and arithmetic word problem solving. In particular, it examines two mathematical abilities that are critical for word problem solving—quantitative reasoning and arithmetic skills (Nunes et al., 2012). It concludes by synthesizing how these strands of research inform the study's focus on problem-solving strategies as a potential mediator linking spatial ability and one-step additive word problem solving. [Subsection 5.2.2](#) reviews literature on path analysis, one of the most frequently used approaches for estimating mediation models. [Subsection 5.2.3](#) presents this study's two research questions.

5.2.1 Arithmetic Word Problems

Solving arithmetic word problems is an important mathematical skill (OECD, 2023). They are different from arithmetic computation tasks in that the problems are presented in verbal descriptions of a quantitative structure rather than number sentences (Verschaffel et al., 2020). Young children can solve arithmetic word problems even before formal schooling through informal strategies involving direct modeling of the situation described in the text of the problem (Carpenter & Moser, 1982; Nunes & Bryant, 1996). Yet, solving arithmetic word problems remains an area of mathematics in which pupils experience great difficulty in schools (Thevenot & Barrouillet, 2015).

Solving arithmetic word problems requires both quantitative reasoning and arithmetic skills (Nunes et al., 2012, 2015). *Quantitative reasoning* is needed to represent the problem's quantitative structure (Nunes et al., 2015), whereas *arithmetic skills* are required to understand numerical relations and operate on them (Guedj, 1997). Prior research suggest that spatial ability is associated with both quantitative reasoning ([Subsection 5.2.1.1](#)) and arithmetic skills ([Subsection 5.2.1.2](#)). The following subsections review these links in turn, with the aim of integrating evidence on quantitative reasoning and arithmetic skills to justify the study's focus on problem-solving strategies as the mechanism linking spatial ability and word-problem-solving performance ([Subsection 5.2.1.3](#)).

5.2.1.1 Quantitative Reasoning

Quantitative reasoning is important in word problem solving because it is required to analyze the problem's quantitative structure (i.e., network of quantities and quantitative relations; Thompson, 1993). A *quantity* is an intrinsic attribute of an object, such as the set size (e.g., Ann has 6 books) or position (e.g., Ann is 3 steps away from the wall), that is best represented by a natural number (Vergnaud, 1982). In contrast, a *quantitative relation* describes the connection between quantities. It may be additive (describing a whole as the sum of its parts, e.g., Bob has 8 more books than Ann) or multiplicative (describing one-to-many correspondence; Bob is twice as far from the wall as Ann) (Nunes et al., 2012). Such quantitative relations are best represented by “directed numbers” (Vergnaud, 1982, p. 46) and necessarily have an inverse (e.g., if Bob has 8 more books than Ann, then Ann has 8 fewer books than Bob; if Bob is twice as far from the wall as Ann, then Ann is half as far from the wall as Bob (or twice as close to the wall) (Dos Santos, 2022).

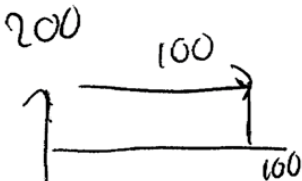
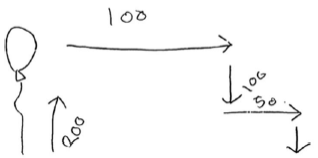
The relationship between spatial ability, quantitative reasoning, and arithmetic word-problem-solving performance has been studied most widely in secondary pupils, although authors have not explicitly used the term quantitative reasoning. Instead, this relationship has typically been examined through analyses of the *visual representations* that problem solvers produce, which provide insight into their representation of the problem's quantitative structure. These visual representations indicate whether the problem solver is engaging with the quantitative structure and, if so, whether it is being represented accurately.

For example, in a seminal study, Hegarty and Kozhevnikov (1999) administered *visual-spatial word problems*—problems that can be solved “relatively easily by visual-pictorial means if the correlation of the given elements in the problem is expressed visually” (Krutetskii, 1976, p. 156)—to 11–13-year-olds, and found that some visual representations were more helpful than others in solving the problems (Boonen et al., 2013, 2014; de Koning et al., 2022; Ng & Lee, 2009). As illustrated in Table 5-1, some pupils produced *iconic representations*, that is, visual representations that focused on depicting the surface features of the objects and persons described in the problem rather than its quantitative structure. In contrast, other pupils produced *schematic representations*, which depicted the “abstract mathematical gist” (Wong, 2021, p. 20), that is, the structure of quantities and the quantitative relations relevant to the solution. Moreover, schematic representations could be either accurate or inaccurate. Whereas accurate schematic representations correctly capture the quantitative

structure inaccurate ones contain errors, which are detrimental to successful problem solving (de Koning et al., 2022).

Table 5-1

Types of Visual Representations for Visual–Spatial Word Problems

Iconic Representation	Inaccurate Schematic Representation	Accurate Schematic Representation
“A balloon first rose 200 meters from the ground, then moved 100 meters to the east, then dropped 100 meters. It then travelled 50 meters to the east, and finally dropped straight to the ground, how far was the balloon from its original starting point?” (Hegarty & Kozhevnikov, 1999, p. 680)		
		

Note. Adapted from Table 4-1 (Chapter 4).

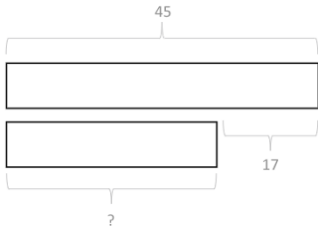
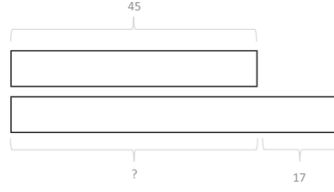
Spatial ability may support children’s representation of the quantitative structure of a problem. Solving arithmetic word problems requires identifying the relevant quantities and quantitative relations described in the text and organizing them into a coherent representation of the problem situation (Thevenot, 2010). Accordingly, spatial ability may facilitate solving visual–spatial word problems—which explicitly describe situations containing spatial elements (Vasilyeva et al., 2009)—by supporting the construction of a representation of the problem’s quantitative structure during problem solving. Consistent with this account, Boonen et al. (2013), using adapted versions of Hegarty and Kozhevnikov’s (1999) problems, found that spatial visualization predicted the production of accurate schematic representations, which in turn predicted successful problem solving. This evidence suggests that spatial ability may influence word-problem-solving performance partly through its association with how children represent and organize the quantitative structure of the problem.

An important question, however, concerns *verbal–analytical word problems*, which can be solved using logical or algebraic reasoning without explicit reliance on visualizing the problem’s quantitative structure (Krutetskii, 1976). As illustrated in Table 5-2, although verbal–analytical word problems do not describe situations containing overt spatial elements, their quantitative structure may nevertheless be represented spatially by the problem solver (Further details regarding the distinction between visual–spatial and verbal–analytical word

problems are provided in [Subsection 3.2.2](#) of Study 1). From this perspective, spatial ability may also support quantitative reasoning in verbal–analytical word problems by facilitating the construction of representations of this type of quantitative structure.

Table 5-2

Types of Visual Representations for Verbal–Analytical Word Problems

Iconic Representation	Inaccurate Schematic Representation ^a	Accurate Schematic Representation ^a
Problem 1: “At the Bijenkorf, jeans cost 45 euros. That is 17 euro less than at C&A. How much will you pay at C&A?” (de Koning et al., 2022, p. 31)		
		

Note. Adapted from Table 4-1 (Chapter 4).

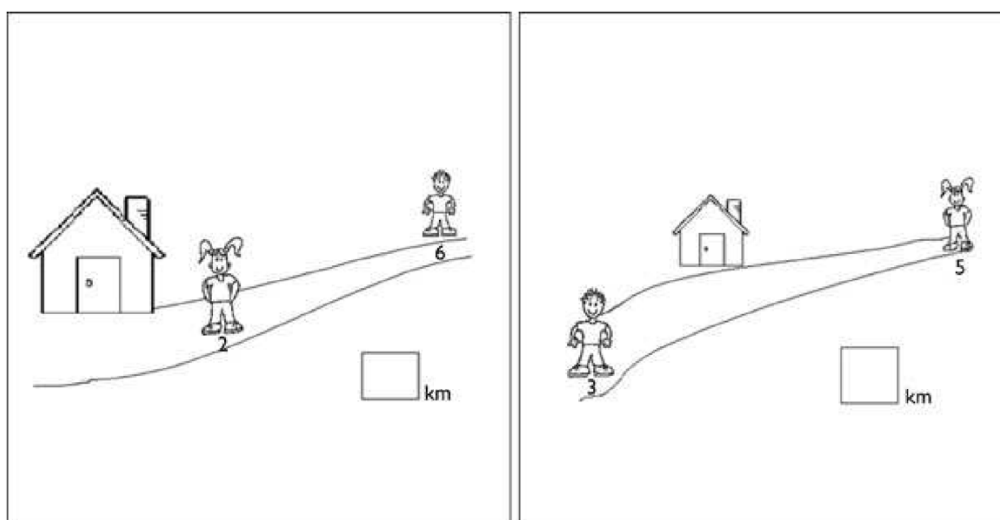
The relationship between spatial ability, quantitative reasoning, and arithmetic word-problem-solving performance has received relatively little attention in younger primary pupils, perhaps due to the challenges of capturing their representations of the quantitative structure. For instance, measuring quantitative reasoning through visual representations is difficult to do with young pupils because they rarely produce such representations when solving mathematical word problems (Jitendra & Woodward, 2019). Instead, studies involving young pupils have assessed quantitative reasoning in alternative ways (e.g., Nunes et al., 2009a, 2012; Wong & Ho, 2017; Yip et al., 2020). However, most do not explicitly use the term *quantitative reasoning*; an exception is the Avon Longitudinal Study of Parents and Children (ALSPAC) project conducted by Nunes and colleagues (2009a). The relationship between spatial ability and quantitative reasoning has not been the primary focus of these studies, even though it may be inferred from their findings. Three illustrative studies are outlined below.

First, as part of the ALSPAC project, Nunes and colleagues (2009a) assessed the quantitative reasoning of a sub-sample of pupils at Year 4 (8–9-year-olds) using specifically designed word problems that required simple calculations but posed a challenge in determining which calculation (or counting strategy) was needed. Consider the following example, illustrated in Figure 5-1 (right), from Nunes et al.’s (2009a) study: “Two friends were walking on the beach. They started walking from the house. The boy walked 3 km one way. The girl walked 5 km the other way. How far apart are the two friends?” (p. 20). Now consider a comparable problem with a similar context, illustrated in Figure 5-1 (left): “Two friends were

walking on the beach. They started walking from the house. The boy walked 6 km one way. The girl walked 2 km in the same way. How far apart are the two friends?’ As these examples illustrate, the computations required in both problems ($3 + 5$ and $6 - 2$, respectively) are well within the range of arithmetic capabilities of 8-year-olds. Rather, the primary challenge lies in reasoning about the relational structure of the problem, which determines the operation (or the counting strategy) required for its solution.

Figure 5-1

Examples of Additive Quantitative Reasoning Problems Used in the ALSPAC Project



*Note. Reprinted from “The relative importance of two different mathematical abilities to mathematical achievement,” by T. Nunes, P. Bryant, R. Barros, & K. Sylva, 2011, *British Journal of Educational Psychology*, 82(1), p. 137 (<https://doi.org/10.1111/j.2044-8279.2011.02033.x>). Copyright 2011 by The British Psychological Society. Reprinted with permission.*

In the ALSPAC project, quantitative reasoning was measured by scoring the correctness of the final numerical answer to these word problems (Nunes et al., 2012). However, in this case, quantitative reasoning cannot be fully disentangled from arithmetic skills, as pupils were still required to produce a numerical answer, which inevitably involves arithmetic skills. Nonetheless, Nunes and colleagues (2009a) found that Year 4 pupils’ spatial visualization was correlated with their quantitative reasoning, and more strongly than with their arithmetic skills. However, this association was not the primary focus of the project, as the authors were more interested in whether arithmetic skills, quantitative reasoning, and spatial ability independently predicted overall mathematical attainment.

Another way to measure young pupils’ quantitative reasoning is by examining their number-sentence construction. Wong and Ho (2017) gave kindergarten pupils ($M = 6.08$ years, $SD = 0.33$ years) 14 one-step and four multistep additive word problems and instructed them

to construct a number-sentence for each. Scores were awarded for correctly written number sentence (1 point for one-step problems and 2 points for multistep problems, with partial scoring allowed). However, the limitation of this measure is that young children can use informal methods to solve arithmetic word problems; thus, constructing a number sentence is neither a necessary nor a typical step in their spontaneous problem-solving process (Hughes, 1986). As such, there may be a dissociation between pupils' representation of the problem's underlying quantitative structure and their construction of number sentence. Although the association between spatial ability and quantitative reasoning was again not the focus of this study, the authors reported significant bivariate correlations between visuospatial working memory and number-sentence construction ($r = .22, p < .01$), as well as between visuospatial working memory and word-problem-solving accuracy ($r = .20, p < .05$).

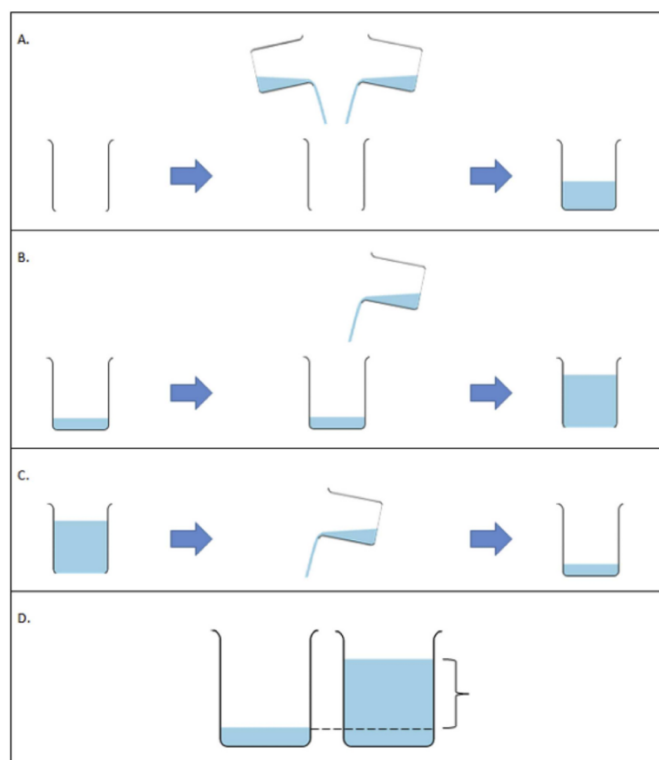
Yet another way to measure quantitative reasoning is through what Wong (2021) terms 'problem representation,' defined as the process of understanding the problem and forming a mental representation of the situation that "extracts the targeted features of the word problems (i.e., the operation involved as well as the semantic role of the unknown)" (Wong & Yip, 2023, p. 5). For example, in a study by Yip et al. (2020), Grade 2 pupils (7–8-year-olds) in Hong Kong were given a set of one-step additive word problems and were asked to match each problem with a conceptual diagram that best represented it. For instance, given the word problem 'Arnold has 5 candies and Lily has 4 candies. How many more candies does Arnold have compared to Lily?', conceptual diagram D in Figure 5-2 best represents this compare word problem. In contrast, the problem 'Peter has 6 candies. His teacher has given him some more, and he now has 9 candies. How many candies did his teacher give him?', is best represented by conceptual diagram B in Figure 5-2, as it describes a change problem involving an increase over time. Although Wong, Yip and their colleagues do not explicitly refer to quantitative reasoning, their measure—assessing pupils' representation of arithmetic word problems without involving computation—effectively captures pupils' ability to identify the underlying quantitative structure of word problems.

A limitation of this measure is that it provides a finite set of conceptual diagrams based on the traditional classification of one-step additive word problems proposed by Riley and colleagues (1983): combine (conceptual diagram A in Figure 5-2), change (conceptual diagrams B and C in Figure 5-2), and compare (conceptual diagram D in Figure 5-2). Setting aside the ecological strength of this classification system (Thevenot, 2010), the measure may artificially constrain young pupils' representation of quantitative structure by requiring them

to conform to representations deemed appropriate by the researchers based on the semantic structure of the problems.

Figure 5-2

Additive Conceptual Diagrams of Problem Representations from Yip et al. (2020)



*Note. Reprinted from “Do children with mathematics learning disability in Hong Kong perceive word problems differently?,” by E. S.-K. Yip, T. T.-Y. Wong, S.-H. Cheung, & K. K.-W. Chan, 2020, *Learning and Instruction*, 68, p. 3 (<https://doi.org/10.1016/j.learninstruc.2020.101352>). Copyright 2020 by Elsevier Ltd. Reprinted with permission.*

As in the previous two studies, the relationship between spatial ability and quantitative reasoning was not the primary focus. Nevertheless, the authors reported that visuospatial working memory positively predicted problem-representation matching ($r = .27, p < .001$), and did so more strongly than verbal working memory ($r = .18, p < .05$). Furthermore, in a similar study, Leung and Wong (2023) found that Grade 6 pupils’ (11–12-year-olds) performance on matching one-step multiplicative word problems with conceptual representations mediated the relationship between spatial ability (spatial visualization, perspective taking, spatial scaling, and hidden figures) and overall mathematical performance, comprising numerical operations and problem solving.

In summary, with the exception of Nunes et al. (2009a), the other studies do not explicitly refer to quantitative reasoning, although they effectively measure it through various

means, such as visual representations (Boonen et al., 2013), number-fact construction (Wong & Ho, 2017), or problem-representation matching (Leung & Wong, 2023; Yip et al., 2020). None of these studies focused specifically on the relationship between spatial ability and quantitative reasoning, despite measuring either spatial visualization or visuospatial working memory.

To contribute to this gap in the literature, [Study 2](#) of this thesis took an initial step by investigating the relationship between spatial ability and quantitative reasoning in 6–7-year-olds. In that study, quantitative reasoning was assessed by observing how pupils solved one-step additive word problems—specifically, whether they engaged with the problem’s quantitative structure and, if so, whether they reasoned about it accurately (i.e., non-quantitative, quantitatively-incorrect, and quantitatively-correct strategies). Pupils were free to express their representations of the quantitative structure in whatever way they chose, including drawing, writing number sentences, using MathLink cubes, or using their fingers.

Study 2 found that, across all problem types, stronger spatial visualization was associated with an increased likelihood of correctly representing the quantitative structure of the problem while using a mental number-fact strategy (i.e., ‘Quantitatively Correct Number Fact’ strategies). In contrast, higher visuospatial working memory was associated with a reduced likelihood of failing to engage with the quantitative structure for most problem types (i.e., ‘Non-Quantitative’ strategies). Moreover, the role of visuospatial working memory in facilitating quantitative reasoning was especially pronounced in word problems involving comparisons between disjoint quantities measured using an imposed unit, where it was associated with reduced use of ‘Quantitatively Incorrect Counting’ strategies.

5.2.1.2 Arithmetic Skills

In addition to quantitative reasoning, arithmetic skills are also important for arithmetic word problem solving (Nunes et al., 2012, 2015). The relationship between spatial ability and arithmetic skills is well documented (Frick, 2019). Beyond simple association, intervention studies suggest a potential causal relationship between spatial ability and arithmetic skills (e.g., Cheng & Mix, 2014; Vasilyeva et al., 2025). For instance, one intervention study found that a single 40-minute training session involving a two-dimensional mental rotation task improved 6–8-year-old’s performance on a missing-term arithmetic task (e.g., $3 + _ = 7$; Cheng & Mix, 2014). However, Hawes et al. (2015) did not replicate this finding in a delayed post-test when

the intervention was digitized and delivered over a longer period of time (six weeks), indicating inconsistency in the evidence.

Other intervention studies have focused less on far transfer from spatial training to mathematical performance and more on training at the spatial–mathematical intersection (e.g., embedding spatial supports within mathematical instruction). For example, in an arithmetic fluency intervention study with 6–7-year-olds, Vasilyeva et al. (2025) found that eight 30-minute small-group training sessions using spatial materials (i.e., materials with lengths proportional to numbers they represent) was more effective than training with non-spatial materials (i.e., equal-sized square tiles regardless of the numerical value). Another example comes from a substantive body of research building on Siegler and Ramani’s (2008) ‘The Great Race’ boardgame, which indicates that playing linear board games (or computerized derivatives) improves arithmetic skills, along with other mathematical competencies (for examples of studies stemming from this paradigm, see Hawes et al., 2019a; Sella et al., 2017).

One way to better understand the link between spatial ability and arithmetic skills is to examine the strategies children use when solving arithmetic tasks. Children’s arithmetic strategies can be broadly grouped into two categories: *counting-based* strategies and *number-fact* strategies. With development, children transition from laborious, external counting strategies (e.g., Count-All, Separate-From), to more economical strategies (e.g., Count-On, Count-Down; Carpenter & Moser, 1982) and non-perceptual strategies (e.g., nodding; Steffe et al., 1982), and eventually to internalized number-fact strategies (Möhring et al. 2024). (See [Subsection 4.2.2.1](#) of Study 2 for a comprehensive list of counting-based strategies.) *Recalled* number facts are typically described as being retrieved directly from long-term memory (Ashcraft & Fierman, 1982). However, alternative accounts propose that number facts with small operands (≤ 4) may rely on automatized counting-based procedures (Baroody, 1983; Barrouillet & Thevenot, 2013). In contrast, *derived* number facts involve decomposing a problem into known facts (e.g., solving $6 + 8$ as $[6 + 6] + 2$). Regardless of the theoretical stance (for a review, see Andras & Macizo, 2026; Thevenot & Prado, 2026), number-fact strategies are generally considered more advanced and sophisticated than counting-based strategies because they are more efficient and less error prone (Casey & Fell, 2018). Furthermore, derived number-fact strategies involve active problem solving that relies on the application of prior mathematical knowledge (Casey & Fell, 2018; Geary, 2011).

The relationship between strategy choice (counting-based vs. number-fact strategies) and arithmetic skills is developmental, a conclusion drawn from studies examining differences

between finger users and non-finger users. In a longitudinal study, Jordan et al. (2008) followed pupils from the beginning of kindergarten ($M = 5.7$ years) to the end of second grade across 11 testing points and assessed their addition and subtraction performance. They found that until the beginning of the second grade (around age 7.8), finger-use frequency was positively correlated with arithmetic accuracy, although this positive association gradually weakened over time. By mid-second grade, however, finger counting was negatively correlated with accuracy.

More recently, Krenger and Thevenot (2025) reported a similar pattern in a longitudinal study focusing on addition tasks. They followed pupils from age 4.5 to 7.5 across seven testing points and found that until the age 6.5, pupils who relied on finger counting outperformed those who did not. After age 6.5, however, this relationship reversed. As Krenger and Thevenot (2025) note, although the two studies differ in the age at which the relationship reverse (7.8 vs. 6.5 years), this discrepancy is likely attributable to differences in task materials and analytic approaches. The critical point, however, is that the association between strategy choice (counting-based vs. mental arithmetic strategies) and arithmetic accuracy changes across development. Although both studies focused specifically on finger counting—without providing pupils with alternatively representational tools—the findings may potentially extend to other counting strategies involving concrete objects (e.g., counters, manipulatives), non-perceptual objects (e.g., nodding or pointing the air, eye-tracking, and gesturing), or marks and drawings on paper (Lê et al., 2024).

Krenger and Thevenot (2025) further observed that many pupils who were non-finger users at age 7 had previously relied on finger counting. These pupils appear to have internalized the strategy such that they no longer required overt finger use to solve addition tasks. Notably, they also outperformed both persistent finger users and pupils who had never relied on finger counting. Although the authors did not explicitly determine whether these former finger users were employing genuine number-fact strategies (recalled or derived facts) as opposed to covert counting (e.g., non-perceptual, vocal, or subvocal counting), the findings suggest that pupils who internalize arithmetic strategies may exhibit higher arithmetic accuracy past age 6.5.

Importantly, both studies locate this shift in the relationship between counting-based and number-fact strategies and arithmetic accuracy within the early primary school years (approximately ages 6–8). This period therefore represents a particularly informative developmental window for examining factors that may influence children's strategy use, including individual differences in spatial ability. Indeed, a complementary body of research

suggests that spatial ability may play an important role in explaining these individual differences in arithmetic strategy use. In related research with children in this same age range, number-fact strategy use has been shown to link spatial ability and arithmetic computation accuracy (for a review, see Casey & Fell, 2018). Spatial ability positively predicts the frequency of more sophisticated mental strategies (particularly derived number-fact strategies) and negatively predicts the frequency of less sophisticated counting-based strategies such as count-all strategies (e.g., Laski et al., 2013; Foley et al., 2017; [Study 2](#)). In addition, mediation analyses from both cross-sectional and longitudinal studies suggest that derived number-fact strategy use accounts for the association between spatial ability and arithmetic performance (Casey et al., 2017; Foley et al., 2017; Möhring et al., 2024).

Among spatial skills, spatial visualization and visuospatial working memory have most frequently been examined in relation to arithmetic strategies. Several explanations may account for why number-fact strategies mediate the relationship between these spatial skills and arithmetic accuracy. Spatial visualization has been implicated in mapping numbers onto space via the mental number line, an internal spatial representation of numerical magnitude (Gunderson et al., 2012). Through this spatial structuring of numerical magnitudes, spatial visualization may support the consolidation and strengthening of numerical representations, enabling children to rely more on efficient number-fact strategies rather than effortful counting-based strategies (Barrouillet & Thevenot, 2013). Moreover, spatial visualization may facilitate the organization of numerical relations in ways that support efficient arithmetic processing, such as decomposing problems into smaller or simpler components (Möhring et al., 2024). In addition, visuospatial working memory may support the maintenance of numerical information—including intermediate numerals and relational transformations—during mental calculation, which may facilitate the use of number-fact strategies (Casey & Fell, 2018; Foley et al., 2017).

Taken together, the literature suggests that spatial ability is associated with children's arithmetic skills (Frick, 2019), but the mechanisms underlying this relationship remain unclear. Research on arithmetic strategies indicates that children rely on a range of approaches when solving arithmetic problems and that the effectiveness of these strategies changes across development. Counting-based procedures can support early arithmetic performance (before approximately age 7.8 or 6.5, Jordan et al., 2008; Krenger & Thevenot, 2025) but tend to give way to more efficient number-fact strategies with development. Notably, this shift occurs during the early primary school years, which is also the developmental period in which higher

spatial ability has been shown to predict greater reliance on number-fact strategies and less reliance on counting-based strategies (Casey et al., 2017; Foley et al., 2017; Laski, 2013; Möhring et al., 2024). Together, these findings suggest that strategy use is a potential mechanism linking spatial ability and arithmetic skills, as spatial skills may support how children organize and operate on numerical relations during problem solving. However, most existing studies have focused on arithmetic computation tasks rather than arithmetic word problems. Whether similar mechanisms operate in the context of word problem solving therefore remains unclear.

5.2.1.3 Spatial Ability and Problem-Solving Strategies

To summarize the literature reviewed above, spatial ability may be important for both quantitative reasoning and arithmetic skills, which in turn are important for arithmetic word problem solving (Nunes et al., 2012, 2015). Spatial visualization and visuospatial working memory may support the representations of a problem's quantitative structure (Hawes & Ansari, 2020; Mix & Cheng, 2012). These two spatial skills may also relate to children's arithmetic strategy use, as spatial processes involved in organizing numerical relations may facilitate reliance on more efficient number-fact strategies rather than effortful counting-based strategies (e.g., Foley et al., 2017; Laski et al., 2013).

Based on the reviewed literature on young pupils' quantitative reasoning and arithmetic skills, their problem-solving strategies for one-step additive word problems can be synthesized as shown in Table 5-3. Importantly, this synthesis reflects the fact that young pupils do not necessarily solve arithmetic word problems in two distinct sequential phases of 'representation' and 'computation' (Carpenter & Moser, 1982; Nunes & Bryant, 1996), as is often assumed for older pupils and adults (Mayer, 1985). This framework of problem-solving strategies, first proposed in Study 2, offers a more holistic perspective by integrating two dimensions: (1) representation of quantitative structure (i.e., whether the quantitative structure is represented correctly, incorrectly, or not at all), and (2) arithmetic strategy (counting-based versus number-fact strategies). (See [Subsection 4.2.2](#) and [Subsection 4.3.4](#) of Study 2 for a fuller discussion of the problem-solving strategies and for examples of each strategy, respectively).

When pupils engage with the problem's quantitative structure, they may represent it either accurately or inaccurately. Independently of this accuracy, pupils may employ either a counting-based strategy or a number-fact strategy. Considering these dimensions results in four strategy types: '*Quantitatively Incorrect Counting*,' '*Quantitatively Incorrect Number Fact*,'

‘*Quantitatively Correct Counting*,’ and ‘*Quantitatively Correct Number Fact*’ strategies. In addition, ‘*Non-Quantitative*’ strategy is characterized by no clear evidence of engagement with the problem’s quantitative structure (e.g., iconic representation).

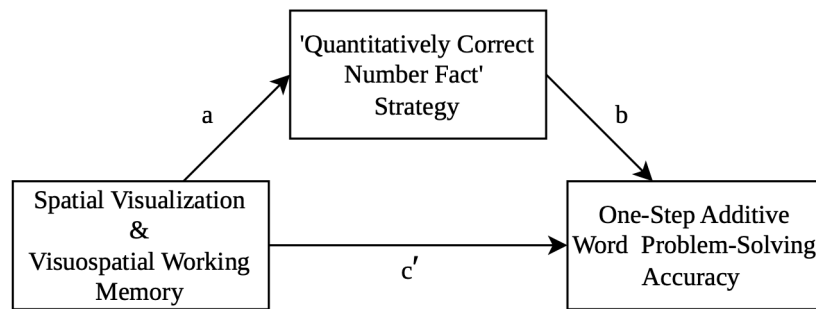
Table 5-3

Problem-Solving Strategies for One-Step Additive Word Problems

Strategy	Definition
Non-Quantitative	Does not appear to actively engage with the problem’s quantitative structure (e.g., guessing; repeating a number from the problem statement)
Quantitatively Incorrect Counting	Represents the problem’s quantitative structure incorrectly while using a counting-based strategy (e.g., counting-all for a subtraction problem)
Quantitatively Incorrect Number Fact	Represents the problem’s quantitative structure incorrectly while using a number-fact strategy (e.g., mentally subtracting for an addition problem)
Quantitatively Correct Counting	Represents the problem’s quantitative structure correctly while using a counting-based strategy (e.g., counting-down for a subtraction problem)
Quantitatively Correct Number Fact	Represents the problem’s quantitative structure correctly while using a number-fact strategy (e.g., mentally adding for an addition problem)

Note. Reproduced from Table 1-3 (Chapter 1).

In light of the reviewed literature, this study hypothesizes that ‘Quantitatively Correct Number Fact’ strategies mediate the relationship between spatial ability and problem-solving accuracy. Problem solvers with higher spatial ability may be better able to correctly represent a problem’s quantitative structure and rely more on number-fact strategies during word problem solving. The hypothesized mediation model is illustrated in Figure 5-3. The present study investigates this hypothesized pathway using path analysis.

Figure 5-3*Hypothesized Mediation Model*

Note. *Path a* represents the effect of spatial visualization and visuospatial working memory, respectively, on 'Quantitatively Correct Number Fact' strategy use, controlling for the other independent variable and all demographic and cognitive covariates (age, gender, language ability, numerical ability, and fluid reasoning). *Path b* represents the effect of 'Quantitatively Correct Number Fact' strategy use on one-step additive word-problem-solving accuracy, controlling for spatial visualization and visuospatial working memory, and other covariates. *Path c'* represents the direct effect of spatial visualization and visuospatial working memory, respectively, on one-step additive word-problem-solving accuracy, controlling for 'Quantitatively Correct Number Fact' strategy use and other covariates.

5.2.2 Path Analysis

Mediation analysis is used to examine hypothesized causal mechanisms by which an independent variable (i.e., spatial skills; hereafter denoted as X) is related to a dependent variable (i.e., problem-solving accuracy; hereafter denoted as Y) through a mediator (i.e., 'Quantitatively Correct Number Fact' strategy; hereafter denoted as M), by decomposing the total effect (c) into indirect effect (ab) and direct effect (c') (Cole & Maxwell, 2003). The indirect effect explains the mechanism of interest (i.e., the degree with which a change in X produces a change in Y by means of M) while the direct effect explains all the other unspecified mechanisms. Path analysis is one of the most commonly used approaches for estimating mediation models when examining manifest (observed) variables using a structural equation modeling framework (MacKinnon, 2008; Rijnhart et al., 2021). This method can accommodate multiple independent variables, mediators, and dependent variables, and it estimates all paths simultaneously, in contrast to Baron and Kenny's (1986) causal-steps approach.

Whereas path analysis directly estimates the indirect effect as the product of associated paths, a ($X \rightarrow M$) and b ($M \rightarrow Y$) (VanderWeele, 2016), the Baron and Kenny approach does not directly estimate or test the indirect effect but infers mediation based on a sequence of statistical significance across multiple regression models (Baron & Kenny, 1986). This traditional causal-steps approach of inferring mediation based on a significant total effect and evidence that direct effect is reduced when the mediator is included in the model, has since

been criticized for the following reasons. First, the claim that the total effect must be significant has been criticized for overlooking the possibility that direct and indirect effects may counteract and cancel each other out (MacKinnon, 2008). Instead, it has been argued that only the indirect effect needs to be significant to establish mediation, and that the strength of mediation should be measured by the size of the indirect effect only (Zhao et al., 2010). Second, contrary to Baron and Kenny's (1986) claim, it has been argued that the significance of the direct effect neither establishes nor negates mediation; rather, it serves to classify the type of mediation (Zhao et al., 2010). If the direct effect is non-significant, this is referred to as *indirect-only mediation*, corresponding to what Baron and Kenny (1986) described as full mediation. If the direct effect is significant, this corresponds to partial mediation. However, it is important to note that the presence or absence of a direct effect does not by itself indicate whether additional mediators are present, as mediation may occur through multiple pathways. There are two cases of partial mediation, depending on whether the direct and indirect effects have the same sign (i.e., complementary mediation) or the opposite sign (i.e., competitive mediation).

The path-analytic approach to mediation, which focuses directly on the indirect effect, addresses several limitations of the traditional causal-steps approach. Nevertheless, it entails important caveats of its own. First, although confidence intervals (CIs) provide more information than p -values by indicating a range of plausible values for the indirect effect (Cumming, 2008; Sobel, 1982), traditional normal-theory CIs have been criticized because they assume that the indirect effect is normally distributed. In practice, however, sampling distribution of the indirect effect is typically asymmetric and skewed toward zero (MacKinnon, 2008). Second, evaluating mediation solely through a test of the indirect effect can yield inflated Type I error rates (i.e., falsely concluding that there is mediation effect) in situations where one component path is large while the other is absent or negligible (Biesanz et al., 2010; Yzerbyt et al., 2018).

Several recommendations have been proposed to address these limitations. One recommendation is the use of resampling methods, such as percentile-based and bias-corrected and accelerated (BCa) bootstrap CIs, which estimate the sampling distribution empirically and therefore do not rely on normality assumptions (MacKinnon, 2008). Although BCa intervals are often preferred because they adjust for both bias and skewness in the bootstrap distribution (Efron & Tibshirani, 1993; Yzerbyt et al., 2018), simulation studies suggest that percentile-based intervals may exhibit more stable Type I error rates in small samples (Biesanz et al.,

2010; Chen & Fritz, 2021). A second recommendation is to evaluate the significance of both path a and path b (Yzerbyt et al., 2018), given that evidence for mediation is most convincing when both constituent paths are statistically significant (MacKinnon, 2008).

While path analysis based on bootstrapped confidence intervals addresses the issue of inflated Type I error, arguably the biggest methodological limitation of path analysis is its reliance of associational data for causal interpretation (Gelfand et al., 2009). As stated above, mediation analysis is often used to examine hypothesized *causal* mechanisms. However, causal interpretations of mediation analyses require several necessary but not sufficient conditions, including evidence of association, non-spuriousness (i.e., the absence of unmeasured confounding), and appropriate temporal ordering (Bollen, 1989). As summarized in Table 5-4, these conditions must hold for each hypothesized causal link underlying the mediation model (i.e., $X \rightarrow M$, $M \rightarrow Y$, and $X \rightarrow Y$). These conditions are most convincingly satisfied through rigorous study designs; accordingly, some argue that the strongest evidence for mediation comes from randomized controlled longitudinal studies that are replicated by independent researchers across different contexts and measures (MacKinnon, 2008), while others take a step even further and suggest double randomization (i.e., randomization of both X and M) when possible along with other study design suggestions (Preacher, 2015).

Table 5-4

Necessary Conditions for Causal Interpretation

	$X \rightarrow M$	$M \rightarrow Y$	$X \rightarrow Y$
Association	Variation in spatial skills is associated with variation in problem-solving strategy use	Variation in problem-solving strategy use is associated with variation in problem-solving accuracy	Variation in spatial skills is associated with variation in problem-solving accuracy
Non-Spuriousness	No unmeasured variable that are correlated with spatial skills and affect problem-solving strategy use	No unmeasured variable that are correlated with problem-solving strategy use and affect problem-solving accuracy	No unmeasured variable that are correlated with spatial skills and affect problem-solving accuracy
Temporal Ordering	Spatial skills temporally precede changes in problem-solving strategy use	Problem-solving strategy use temporally precedes changes in problem-solving accuracy	Spatial skills temporally precede changes in problem-solving accuracy

Note. X = spatial skills; M = ‘Quantitatively Correct Number Fact’ strategy; Y = problem-solving accuracy.

Cross-sectional study designs can support evidence of association and address some sources of confounding, but they provide limited evidence for appropriate temporal ordering.

The evidence of association can be assessed statistically by examining the associations among X, M, and Y, including the significance of path a and path b within the mediation model (Gelfand et al., 2009). While the non-spuriousness condition is fundamentally untestable on empirical data (Alfons & Schley, 2025), measuring potential covariates and controlling for them statistically can partially address this concern. For instance, key covariates known to explain differences in word-problem-solving strategy and accuracy, such as demographic (age and gender), and cognitive factors (language ability, numerical ability, and fluid reasoning) (for a review, see Atit et al., 2022; Daroczy et al., 2015; Vessonen et al., 2024), need to be measured and controlled for. Moreover, testing the assumption of no interaction between X and M in the outcome model helps assess whether the mediation model has been correctly specified, ensuring that the estimated indirect effect is not distorted by an omitted moderation effect—specifically, that the effect of the M on the Y does not depend on the level of X (Alfons & Schley, 2025; MacKinnon et al., 2020). Alternatively, sensitivity analysis techniques help to assess how strong an unmeasured confound would have to be related to both the M and Y to substantially change conclusions being drawn about the indirect effects (VanderWeele, 2016).

The appropriate temporal ordering (i.e., X comes before M, which comes before Y), however, cannot be established in a cross-sectional study design since it is not possible to rule out that variables are changing simultaneously, or that the true causal directions are reversed (Cole & Maxwell, 2003). When the study design does not give information regarding temporal ordering, it must be assumed based on theory and previous empirical research (MacKinnon, 2008). Guided by prior literature and the generative structure of the task, the model is parameterized such that spatial skills (spatial visualization and visuospatial working memory), reflecting relatively stable cognitive abilities, are treated as antecedent to problem-solving strategies enacted during task execution, with problem-solving accuracy treated as the task outcome. This ordering reflects the logical sequence of the task rather than a claim of causal direction. One approach to evaluating the plausibility of the assumed temporal ordering is to compare the hypothesized model with theoretically plausible alternative orderings (Gelfand et al., 2009). However, more recently, Rohrer et al. (2022) cautioned that a larger indirect effect in the hypothesized model does not constitute evidence that the model is correct, as misspecified models may also produce substantial indirect effects.

5.2.3 Present Study

The present study goes beyond establishing *whether* a relationship exists between spatial ability and word problem solving (Study 1) to examining *how* this relationship may operate by identifying potential underlying mechanisms, namely children's problem-solving strategy use. Specifically, this is investigated by testing whether 'Quantitatively Correct Number Fact' strategy use mediates the relationship between spatial skills (spatial visualization and visuospatial working memory) and word-problem-solving accuracy (MacKinnon et al., 2020). Accordingly, the first research question is as follows:

RQ5) To what extent do spatial visualization and visuospatial working memory exert indirect effects on young primary school pupils' (6–7-year-olds') accuracy on one-step additive word problems through 'Quantitatively Correct Number Fact' strategy use?

In addition to RQ5, Studies 1 and 2 found that the relationships between spatial skills and problem-solving accuracy and strategies, respectively, vary depending on the word problem type. As such, the mediation model may operate differently across word problem types. Accordingly, the second research question is as follows:

RQ6) Do the indirect effects of spatial visualization and visuospatial working memory on problem-solving accuracy through 'Quantitatively Correct Number Fact' strategy use vary across different word problem types?

To investigate RQ6, this study adopts the 2×2 conceptual framework (Quantity $\times$ Quantitative Relation) for categorizing one-step additive word problems proposed in Study 1. This conceptual framework categorizes word problems based on their quantitative structure (Thompson, 1993) by considering two categories of quantity (Davydov, 1982) and two categories of quantitative relation (Vergnaud, 1982), which results in four word problem types: *Discrete Quantity–Dynamic Relation*, *Discrete Quantity–Static Relation*, *Continuous Quantity–Dynamic Relation*, and *Continuous Quantity–Static Relation*. Figure 5-4 summarizes the definitions and examples of each word-problem type.

Figure 5-4*Word Problem Types Based on Quantitative Structure*

		Quantitative Relation	
		Dynamic Relation	Static Relation
Quantity	Discrete Quantity	Problems involving transformations of quantities measured through one-to-one counting e.g., Ann had 6 marbles. Then she played a game of marbles and won 8 more marbles. How many marbles does she have now?	Problems involving comparisons of disjoint quantities measured through one-to-one counting e.g., Ann has 5 marbles. Bob has 7 more marbles than Ann. How many marbles does Bob have?
	Continuous Quantity	Problems involving transformations of quantities measured using an imposed unit e.g., Ann was 4 steps from the wall. Then she walked 7 steps away from the wall. How far is Ann from the wall now?	Problems involving disjoint comparisons of quantities measured using an imposed unit e.g., Ann is 9 steps from the wall. Bob is 3 steps farther from the wall than Ann. How far is Bob from the wall?

Note. Reproduced from Figure 1-1 (Chapter 1).

5.3 Methods

To investigate whether lower primary pupils' spatial visualization and visuospatial working memory relate to their one-step additive word-problem-solving accuracy via their 'Quantitatively Correct Number Fact' strategy use, a sample of Year 2 pupils in England were examined. This research has been approved by the Central University Research Ethics Committee and adhered to the ethical guidelines for educational research as recommended by the British Educational Research Association (2018) and the University of Oxford Policy on Research Involving Human Participants and Data Protection (Central University Research Ethics Committee, 2022). This section outlines the participants ([Subsection 5.3.1](#)), measures ([Subsection 5.3.2](#)), procedure ([Subsection 5.3.3](#)), and analytic approach ([Subsection 5.3.4](#)).

5.3.1 Participants

Participants were 149 Year 2 pupils attending state schools across South East England. To capture a wider range of mathematical skills and arithmetic word-problem-solving strategies and accuracy, as well as enhance the generalizability of the findings to the broader population, pupils were included in the study as long as they had sufficient proficiency in English, and the required Special education Needs (SEN) support, if any, were minimal, such

that there was no reason to believe that they could not engage meaningfully with the tasks. Among the 149 pupils tested, 12 pupils discontinued either because they could not understand the tasks or expressed a desire to stop. Among the remaining 137 pupils (50.4% female, 49.6% male; age range: 6.52–7.74 years; mean age: 7.10 years), 27 (19.7%) were reported as SEN and 21 (15.3%) were reported as English as an additional Language (EAL) pupils. As SEN and EAL information was not always available, it is possible that an even higher percentage of SEN and EAL pupils were included in the study. The schools were recruited first based on a non-probabilistic convenience sampling method, followed by the pupils, either through opt-in or opt-out parental consent, as deemed appropriate by the respective schools.

5.3.2 Measures

Pupils completed seven tasks designed to assess spatial skills (spatial visualization and visuospatial working memory), problem-solving accuracy, problem-solving strategy use, and relevant covariates. Spatial visualization was assessed using two measures: the Wechsler Abbreviated Scale of Intelligence–Second Edition (WASI–II) Block Design subtest (Atit et al., 2022) and the Visual–Spatial Reasoning task (Hawes et al., 2015, 2017, 2019). The Block Design subtest involved recreating a design using red-and-white cubes within a specified time limit. The Visual–Spatial Reasoning task involved reasoning about and visualizing solutions to problems involving parts of objects in relation to their whole. Visuospatial working memory was assessed using a digitized version of the Corsi Block-Tapping task on an iPad (Hawes et al., 2019b), which involved replicating a sequence of flashing dots by tapping them in the same order.

Problem-solving accuracy and strategy use were assessed using the 24 one-step additive word problems adopted from Study 1. Of the 24 items, six belonged to each of the four word problem types outlined in [Subsection 5.2.3](#). The four parallel versions of the full problem set (Forms A–D) are presented in [Subsection 2.3.3.2](#).

The three covariates, language ability, numerical ability, and fluid reasoning were measured using the following standardized tests, respectively: the Wechsler Individual Achievement Test–Second UK Edition (WIAT–II^{UK}) Listening Comprehension and Numerical Operations subtests, and WASI–II Matrix Reasoning subtest. The Listening Comprehension subtest involved selecting the picture that matched a word or sentence and generating a word that matched a picture and oral description. The Numerical Operations subtest involved identifying and writing numbers, counting, and solving written calculation problems with basic

operations of addition, subtraction, multiplication, and division. The Matrix Reasoning subtest involved completing an incomplete matrix or series by selecting one of five response options. (The detailed descriptions of the measures are presented in [Subsection 2.3](#) of the Methodology chapter.)

5.3.3 Procedure

Each pupil completed all seven tasks individually with the researcher at a quiet location at school. The seven measures were administered in a randomized order across two 30-minute sessions that were typically spaced one-hour apart, or a minimum, 30 minutes apart. Pupils were randomly assigned to one of the four forms of one-step additive word problems (Forms A–D). The 24 one-step additive word problems were divided between the two testing sessions (i.e., 12 items per session). The administration order of the word-problem items was randomized subject to predefined sequencing constraints. The problems were presented on A4-size papers, with four problems arranged in each quadrant of a page. Each problem was read aloud slowly and repeated if necessary. Pupils were provided with 30 MathLink cubes and a pencil, informed that the materials could be used if helpful, and reminded that the materials were available if they were stuck on a problem for more than one minute. The interview-like procedure for administering word problems was adapted from Hegarty and Kozhevnikov (1999) and Boonen et al. (2013, 2014). (The detailed administration procedure for the word problems, including the randomization of item order, is outlined in [Subsection 2.3.3.3](#) of the Methodology chapter.)

5.3.4 Analytic Approach

Pupils' numerical solutions to the word problems were scored as either incorrect (0) or correct (1). Independently of the correctness of the numerical solutions, problem-solving strategies were coded based on pupils' observable behavior and retrospective reports (Carr et al., 2008; Cho et al., 2011). When discrepancies arose between the two sources, the researcher's observations were used (Casey et al., 2017; Laski et al., 2013). As summarized in Table 5-3, each problem-solving strategy was coded based on the dimensions of (1) the representation of the quantitative structure (none, incorrect, or correct) and, where applicable, (2) arithmetic strategy (counting-based or number-fact strategy): 'Non-Quantitative,' 'Quantitatively Incorrect Counting,' 'Quantitatively Incorrect Number Fact,' 'Quantitatively Correct Counting,' and 'Quantitatively Correct Number Fact' strategies.

A second rater independently coded 20.4% of the data (28 pupils $\times$ 24 items), selected at random, using the primary author's field notes. The two raters showed 97.2% agreement, and Cohen's κ was .96, 95% CI [.95, .98], $p < .001$, indicating excellent inter-rater reliability (Siegel & Castellan, 1988). (The coding process is described in detail in [Subsection 2.3.3.4](#) of the Methodology chapter.)

The mediation models were estimated using path analysis to examine whether the use of 'Quantitatively Correct Number Fact' strategies, in particular, mediates the relationship between spatial skills (spatial visualization and visuospatial working memory) and problem-solving accuracy (RQ5), and whether the mediation effect varied across word problem types (RQ6). Although pupils were nested within classrooms, multilevel modeling was not pursued because the number of clusters was very small (six classrooms). With so few clusters, some with very small number of pupils, variance component estimates and cluster-robust standard errors are known to be unstable (Huang & Li, 2022; McNeish & Stapleton, 2016). Therefore, models were estimated at the individual level. Similarly, cluster-level fixed effects were not included, as the small number of clusters and limited within-cluster sample sizes would substantially reduce degrees of freedom and yield unstable estimates. To evaluate the potential impact of nesting, Intraclass Correlation Coefficients (ICCs) were calculated for problem-solving strategy use and problem-solving accuracy and reported in the Results section.

Statistical analyses were conducted in R version 4.4.0 (R Core Team, 2024) using RStudio version 2023.12.1.402 (Posit Team, 2024), and with the *lavaan* package (Rosseel, 2012). Spatial visualization and visuospatial working memory were specified as the independent variables. Problem-solving accuracy, operationalized as the total number of correctly solved problems, served as the dependent variable. For RQ5, accuracy was calculated across all 24 word problems. For RQ6, accuracy was calculated within each subset of 6 items belonging to the same word problem type. The mediator was defined as the proportion of word problems solved using the 'Quantitatively Correct Number Fact' strategy, computed across the same set of items as the corresponding accuracy measure (i.e., 24 items for RQ5 and six items for each RQ6 model). Key cognitive and demographic variables—language ability, numerical ability, fluid reasoning, age, and gender—were included as covariates. Continuous predictor variables were mean centered prior to analysis to facilitate interpretation of model estimates while preserving the original metric of the variables. Standardization was not applied, as comparisons of standardized effect sizes were not the focus of the analysis. The composite spatial visualization score was derived from a principal component analysis of the Visual–

Spatial Reasoning task and the WASI-II Block Design subtest (detailed in [Subsection 2.3.2.1.1](#) of Methodology chapter).

Best practice recommendations outlined in [Subsection 5.2.2](#) were followed, where possible when evaluating the indirect effects of spatial visualization and visuospatial working memory (MacKinnon, 2008). The mediation models were estimated using maximum likelihood with 95% bias-corrected and accelerated bootstrapped confidence intervals (BCa CIs) based on 5,000 resamples to account for the potentially non-normal distribution of indirect effects. BCa CIs were deemed appropriate given the moderate sample size and the large number of bootstrap resamples. Following current methodological recommendations for mediation (Preacher & Hayes, 2008), statistical significance was determined by whether the BCa CIs excluded zero, as this approach does not assume a normal sampling distribution for the indirect effect. Although mediation was evaluated primarily based on the indirect effect, the component paths were also examined to aid interpretation, as per recommendation by Yzerbyt et al. (2018).

5.4 Findings

The findings from the path analyses are presented in this section. [Subsection 5.4.1](#) addresses RQ5 by examining the indirect effects of spatial skills (spatial visualization and visuospatial working memory) on problem-solving accuracy via ‘Quantitatively Correct Number Fact’ strategy use across all one-step additive word problems. [Subsection 5.4.2](#) addresses RQ6 by examining these indirect effects separately by word problem type.

5.4.1 RQ5: Indirect Effects of Spatial Skills across Word Problem Types

RQ5) To what extent do spatial visualization and visuospatial working memory exert indirect effects on young primary school pupils’ (6–7-year-olds’) accuracy on one-step additive word problems through ‘Quantitatively Correct Number Fact’ strategy use?

5.4.1.1 Preliminary Analyses

Prior to presenting the mediation model, preliminary analyses are reported, including the descriptive statistics and correlations ([Subsection 5.4.1.1.1](#)), and examination of assumptions and data structure ([Subsection 5.4.1.1.2](#)).

5.4.1.1.1 Descriptive Statistics and Correlations

Table 5-5 presents descriptive statistics and bivariate correlations for all variables, with zero-order correlations presented above the diagonal and partial correlations presented below

the diagonal. Descriptively, pupils exhibited substantial variability in their use of ‘Quantitatively Correct Number Fact’ strategies, using this strategy relatively infrequently on average ($M = 4.85$, $SD = 5.98$). Problem-solving accuracy also varied widely across pupils ($M = 9.77$, $SD = 7.12$).

At the zero-order level, both spatial visualization and visuospatial working memory were positively associated with ‘Quantitatively Correct Number Fact’ strategy use ($r = .55$, $p < .001$; $r = .37$, $p < .001$, respectively) and with problem-solving accuracy ($r = .44$, $p < .001$, $r = .35$, $p < .001$, respectively). When controlling for covariates and the other spatial skill, the unique relationship between spatial visualization and ‘Quantitatively Correct Number Fact’ strategy use remained significant but was substantially reduced in magnitude ($r = .22$, $p < .05$), whereas the unique association between visuospatial working memory and strategy use became non-significant ($r = .10$, $p = .25$).

‘Quantitatively Correct Number Fact’ strategy use was strongly positively associated with problem-solving accuracy at both the zero-order ($r = .74$, $p < .001$) and partial ($r = .50$, $p < .001$) levels, indicating a strong association that remained after controlling for spatial skills and covariates. In contrast, the associations between the spatial skills and problem-solving accuracy were no longer evident after adjustment ($r = -.03$, $p = .74$; $r = .10$, $p = .25$, respectively), suggesting that their zero-order relationships with problem-solving accuracy largely overlap with other cognitive factors and problem-solving strategy. These preliminary partial correlations are consistent with the possibility that the association between spatial skills and problem-solving accuracy is accounted for by problem-solving strategy use and other control variables.

Table 5-5*Descriptive Statistics and Bivariate Correlations*

	Mean (<i>SD</i>)	Min	Max	1	2	3	4	5	6	7	8	9
1. Spatial Visualization	0.00 (1.00)	-1.90	3.53	-	.42***	.55***	.44***	.26**	.55***	.53***	.02	.11
2. Visuospatial Working Memory	4.39 (1.21)	2	8	-	-	.37***	.35***	.16	.39***	.29***	.06	.07
3. Strategy	4.85 (5.98)	0	23	.22*	.10	-	.74***	.32***	.64***	.54***	.15	.20*
4. Problem-Solving Accuracy	9.77 (7.12)	0	23	-.03	.10	.50***	-	.46***	.68***	.61***	.24**	.14
5. Language Ability	17.78 (3.89)	8	27	-	-	-	-	-	.36***	.42***	-.01	-.17*
6. Numerical Ability	12.71 (3.26)	5	23	-	-	-	-	-	-	.57***	.19*	.16
7. Fluid Reasoning	9.03 (4.22)	2	20	-	-	-	-	-	-	-	.23**	.02
8. Age (years)	7.10 (0.30)	6.52	7.74	-	-	-	-	-	-	-	-	.09
9. Gender (0 = Female, 1 = Male)	0.50 (0.50)	0	1	-	-	-	-	-	-	-	-	-

Note. $N = 137$. SD = standard deviation; Strategy = ‘Quantitatively Correct Number Fact’ strategy. The composite spatial visualization score was derived from principal component analysis of the Visual–Spatial Reasoning task and the WASI–II Block Design subtest. ‘Quantitatively Correct Number Fact’ strategy use and problem-solving accuracy are expressed as raw frequencies (out of 24 problems). Above diagonal: zero-order correlations; below diagonal: partial correlations. Partial correlations involving the independent variables (spatial visualization and visuospatial working memory) were estimated by controlling for the other independent variable and all covariates (age, gender, language ability, numerical ability, and fluid reasoning). The partial correlation between the mediator (‘Quantitatively Correct Number Fact’ strategy use) and the outcome (problem-solving accuracy) was estimated by controlling for both independent variables and all covariates. *** $p < .001$, ** $p < .01$, * $p < .05$.

5.4.1.1.2 Assumption Checks and Data Structure

Assumption diagnostics were conducted for the constituent regression models underlying the mediation analysis. Specifically, linearity, homoscedasticity, normality of residuals, multicollinearity, influential observations, homogeneous mediation, and clustering were examined.

First, linearity was evaluated using component-plus-residual plots (see [Appendix 5A](#)). Most predictors exhibited approximately linear and monotonic associations with their respective outcomes. Some deviations from linearity were observed, particularly for spatial visualization and numerical ability. However, these deviations did not follow a clear or systematic pattern and were not considered sufficiently pronounced to warrant nonlinear specification, particularly given that numerical ability served as a covariate rather than a focal predictor.

In contrast, the association between ‘Quantitatively Correct Number Fact’ strategy use and problem-solving accuracy—the focal path in the mediation model—showed a more systematic pattern suggestive of diminishing returns. A supplementary analysis therefore compared linear and quadratic specifications of this relationship. The inclusion of a quadratic term significantly improved model fit, ($F(1, 127) = 6.55, p = .01$). Consistent with this finding, the quadratic SEM specification exhibited slightly lower AIC and BIC values (AIC = 704.21, BIC = 756.77) than the linear specification (AIC = 709.11, BIC = 758.74). Nevertheless, the linear specification was retained in the primary mediation model to facilitate interpretation of the indirect effect. The quadratic specification was examined separately as a sensitivity analysis (see [Subsection 5.4.1.3](#)).

Inspection of residual-versus-fitted plots indicated no substantial evidence of heteroscedasticity. Normality of residuals was evaluated using Q-Q plots. Although residuals exhibited modest departures from normality, particularly in the tails, these deviations were not considered sufficiently severe to invalidate the analyses. Moreover, indirect effects were evaluated using BCa CIs, which do not rely on normal-theory assumptions regarding the sampling distribution of the indirect effect.

Multicollinearity was assessed using variance inflation factors (VIFs). All predictors exhibited acceptable VIF values (all < 5), ranging from 1.10 to 1.89 in the mediator model and from 1.12 to 2.14 in the outcome model, indicating no evidence of problematic multicollinearity. Influential observations were evaluated using Cook’s distance. No cases

exceeded conventional cutoff of ($4/n \approx 0.03$) in either the mediator or outcome models, indicating that results were not driven by influential observations.

To evaluate the assumption of homogeneous mediation, interactions between spatial skills (spatial visualization and visuospatial working memory) and ‘Quantitatively Correct Number Fact’ strategy use were added to the outcome model. Neither interaction was significant ($ps > .05$), indicating no evidence that the strategy–accuracy association varied as a function of spatial skills.

Finally, intraclass correlation coefficients (ICCs) were examined to assess clustering by classroom. ICCs indicated modest clustering for ‘Quantitatively Correct Number Fact’ strategy use ($ICC = .08$) and moderate clustering for problem-solving accuracy ($ICC = .18$). Although the latter suggests some classroom-level variance in problem-solving accuracy, the small number of available clusters ($k = 6$) precluded the use of multilevel modeling. Consequently, the subsequent path analysis was conducted at the individual level. This limitation is considered further in the Discussion.

5.4.1.2 Path Analysis

A path model was estimated to examine whether ‘Quantitatively Correct Number Fact’ strategy use mediated the relationship between spatial skills (spatial visualization and visuospatial working memory) and problem-solving accuracy across all one-step additive word problems. Model fit indices indicated acceptable to good fit ($CFI = .98$, $TLI = .90$, $RMSEA = .09$, $SRMR = .04$). The model explained a substantial proportion of the variance in both ‘Quantitatively Correct Number Fact’ strategy use ($R^2 = .50$) and problem-solving accuracy ($R^2 = .69$).

Table 5-6 summarizes the path coefficients predicting ‘Quantitatively Correct Number Fact’ strategy use and problem-solving accuracy. With respect to strategy use, neither spatial visualization ($B = 0.05$, 95% CI $[-0.002, 0.10]$) nor visuospatial working memory ($B = 0.02$, 95% CI $[-0.01, 0.04]$) was significantly associated with strategy use. Among the covariates, numerical ability ($B = 0.03$, 95% CI $[0.01, 0.04]$) and gender ($B = 0.06$, 95% CI $[0.002, 0.12]$) were significantly associated with strategy use.

With respect to problem-solving accuracy, neither spatial visualization ($B = -0.84$, 95% CI $[-1.74, 0.18]$) nor visuospatial working memory ($B = 0.22$, 95% CI $[-0.42, 0.79]$) was significantly associated with problem-solving accuracy when strategy use and other covariates were included in the model, indicating no direct effect. In contrast, ‘Quantitatively Correct

Table 5-6*Path Coefficients Predicting ‘Quantitatively Correct Number Fact’ Strategy use and Problem-Solving Accuracy*

	‘Quantitatively Correct Number Fact’ Strategy				Problem-Solving Accuracy			
	<i>B</i>	β	SE	95% BCa CI	<i>B</i>	β	SE	95% BCa CI
Spatial Visualization	0.05	.21	0.03	[−0.002, 0.10]	−0.84	−.12	0.49	[−1.74, 0.18]
Visuospatial Working Memory	0.02	.08	0.01	[−0.01, 0.04]	0.22	.04	0.31	[−0.42, 0.79]
Language Ability	0.01	.09	0.004	[−0.003, 0.01]	0.32	.18	0.11	[0.12, 0.54]
Numerical Ability	0.03	.35	0.01	[0.01, 0.04]	0.55	.25	0.20	[0.17, 0.95]
Fluid Reasoning	0.01	.16	0.01	[−0.0002, 0.02]	0.31	.19	0.12	[0.08, 0.54]
Age	0.03	.03	0.05	[−0.08, 0.13]	1.73	.07	1.32	[−0.94, 4.20]
Gender	0.06	.13	0.03	[0.002, 0.12]	0.57	.04	0.74	[−0.83, 2.06]
‘Quantitatively Correct Number Fact’ Strategy	-	-	-	-	13.04	.45	2.27	[8.41, 17.35]

Note. *B* = unstandardized coefficient; β = standardized coefficient. SE = standard error. The mediator is expressed as a proportion (0–1), such that a one-unit increase represents a shift from no use to exclusive use of the strategy. In contrast, other predictors are mean centered but retain their original units. Accordingly, unstandardized coefficients (*B*) may differ in magnitude due to scaling and should be interpreted in this context. Confidence intervals (CIs) are bias-corrected and accelerated (BCa) bootstrap intervals based on 5,000 resamples. Statistical significance was evaluated based on 95% BCa CIs and is indicated by boldface.

Number Fact' strategy use was strongly associated with problem-solving accuracy ($B = 13.04$, 95% CI [8.41, 17.35]). Among the covariates, language ability ($B = 0.32$, 95% CI [0.12, 0.54]), numerical ability ($B = 0.55$, 95% CI [0.17, 0.95]), and fluid reasoning ($B = 0.31$, 95% CI [0.08, 0.54]) were significantly associated with problem-solving accuracy.

As summarized in Table 5-7, neither the indirect effect of spatial visualization ($B = 0.67$, 95% BCa CI [−0.003, 1.58]) nor that of visuospatial working memory ($B = 0.22$, 95% BCa CI [−0.10, 0.65]) was statistically significant. Percentile-based confidence intervals yielded the same results (see [Appendix 5B](#)). These findings do not provide evidence for mediation in the overall model.

Table 5-7

Indirect Effects of Spatial Skills on Problem-Solving Accuracy

Indirect Effect	B	β	SE	95% BCa CI
Spatial Visualization				
→ 'Quantitatively Correct Number Fact' strategy	0.67	.09	0.39	[−0.003, 1.58]
→ Problem-Solving Accuracy				
Visuospatial Working Memory				
→ 'Quantitatively Correct Number Fact' strategy	0.22	.04	0.18	[−0.10, 0.65]
→ Problem-Solving Accuracy				

Note. B = unstandardized coefficient; β = standardized coefficient. SE = standard error. Confidence intervals (CIs) are bias-corrected and accelerated (BCa) bootstrap intervals based on 5,000 resamples. Statistical significance was evaluated based on 95% BCa CIs.

5.4.1.3 Sensitivity Analysis

To assess whether the mediation findings depended on imposing a directional mediator–outcome path, an alternative model was estimated in which 'Quantitatively Correct Number Fact' strategy use and problem-solving accuracy were treated as correlated processes, while both were predicted by spatial skills and covariates. In this model, strategy use and problem-solving accuracy exhibited a substantial residual correlation ($r = .50$), and the pattern of associations between spatial skills and strategy use remained substantively unchanged. No alternative pattern emerged that contradicted the hypothesized model, indicating that the main findings were not driven by the directional specification of the mediator–outcome path.

In addition, a sensitivity analysis was conducted to examine whether the mediation findings were robust to nonlinear specification of the mediator–outcome relationship. Specifically, a quadratic term for 'Quantitatively Correct Number Fact' strategy use was included in the outcome model. Results differed depending on model specification. In the linear

model, indirect effects were not statistically significant. In contrast, when a quadratic term was included, the indirect effect of spatial visualization reached statistical significance ($B = 0.94$ 95% BCa CI [0.01, 2.08]). However, because the mediator–outcome relationship was modeled nonlinearly, this indirect effect reflects only the linear component of the relationship and is therefore not directly comparable to that from the primary linear model. Overall, these findings indicate that evidence for mediation is sensitive to model specification and should be interpreted with caution.

5.4.2 RQ6: Indirect Effects of Spatial Skills by Word Problem Type

RQ6) Do the indirect effects of spatial visualization and visuospatial working memory on problem-solving accuracy through ‘Quantitatively Correct Number Fact’ strategy use vary across different word problem types?

5.4.2.1 Descriptive Statistics by Word Problem Type

Table 5-8 presents descriptive statistics of ‘Quantitatively Correct Number Fact’ strategy use and problem-solving accuracy by word problem type. Descriptively, pupils were, on average, most accurate when solving Discrete Quantity–Dynamic Relation problems ($M = 3.35$, $SD = 1.94$), followed by Discrete Quantity–Static Relation ($M = 2.40$, $SD = 2.10$), Continuous Quantity–Dynamic Relation ($M = 2.15$, $SD = 1.86$), and Continuous Quantity–Static Relation ($M = 1.87$, $SD = 2.06$) problems. A similar descriptive pattern was observed for ‘Quantitatively Correct Number Fact’ strategy use, with more frequent use in Discrete Quantity–Dynamic Relation problems and less frequent use in Continuous Quantity–Static Relation problems.

Table 5-8

Descriptive Statistics for Problem-Solving Accuracy and Strategy by Word Problem Type

	Discrete Quantity –Dynamic Relation			Discrete Quantity –Static Relation			Continuous Quantity –Dynamic Relation			Continuous Quantity –Static Relation		
	<i>M</i> (<i>SD</i>)	Min	Max	<i>M</i> (<i>SD</i>)	Min	Max	<i>M</i> (<i>SD</i>)	Min	Max	<i>M</i> (<i>SD</i>)	Min	Max
Strategy	1.65 (2.01)	0	6	1.15 (1.68)	0	6	1.09 (1.53)	0	5	0.96 (1.56)	0	6
Problem-Solving Accuracy	3.35 (1.94)	0	6	2.40 (2.10)	0	6	2.15 (1.86)	0	6	1.87 (2.06)	0	6

Note. M = mean; SD = standard deviation; Strategy = ‘Quantitatively Correct Number Fact’ strategy. ‘Quantitatively Correct Number Fact’ strategy use and problem-solving accuracy are expressed as raw frequencies (out of 6 problems for each word problem type).

5.4.2.2 Path Analyses by Word Problem Type

Four separate mediation models were estimated for each word problem type to examine whether the mediation effect varied as a function of the quantitative structure of the problems. Table 5-9 summarizes the path coefficients for the mediation models, including paths from spatial visualization and visuospatial working memory to ‘Quantitatively Correct Number Fact’ strategy use (i.e., path a), paths from strategy use to problem-solving accuracy (i.e., path b), and the corresponding indirect effects. Full covariate results are reported in [Appendix 5C](#). Across problem types, the mediation models accounted for a substantial proportion of variance in both ‘Quantitatively Correct Number Fact’ strategy use ($R^2 = .44, .46, .35$, and $.36$, respectively) and problem-solving accuracy ($R^2 = .54, .63, .62$, and $.58$, respectively). Figure 5-5 illustrates the mediation models.

As summarized in Table 5-9, neither spatial visualization nor visuospatial working memory had an indirect effect on accuracy of solving Discrete Quantity–Dynamic Relation problems. On the other hand, spatial visualization showed a significant indirect effect on problem-solving accuracy via ‘Quantitatively Correct Number Fact’ strategies for Discrete Quantity–Static Relation ($B = 20$, 95% BCa CI [0.04, 0.45]) and Continuous Quantity–Dynamic Relation ($B = 0.29$, 95% BCa CI [0.05, 0.57]) problems.

For both word problem types, higher spatial visualization was associated with greater use of ‘Quantitatively Correct Number Fact’ strategies ($B = 0.07$, 95% CI [0.01, 0.13] for both). In turn, ‘Quantitatively Correct Number Fact’ strategy use was positively associated with problem-solving accuracy ($B = 2.99$, 95% CI [1.85, 4.07]; $B = 4.00$, 95% CI [2.90, 5.07], respectively). These mediated models are illustrated in Figure 5-5 (top-right and bottom-left).

In contrast, for Continuous Quantity–Static Relation problems, visuospatial working memory showed a significant indirect effect on problem-solving accuracy via ‘Quantitatively Correct Number Fact’ strategy use ($B = 0.15$, 95% BCa CI [0.01, 0.34]). Higher visuospatial working memory was associated with a small increase in ‘Quantitatively Correct Number Fact’ strategy use ($B = 0.04$, 95% BCa CI [0.0003, 0.08]), which, in turn, was strongly associated with higher problem-solving accuracy ($B = 3.89$, 95% BCa CI [2.27, 5.54]). This mediated model is illustrated in Figure 5-5 (bottom-right).

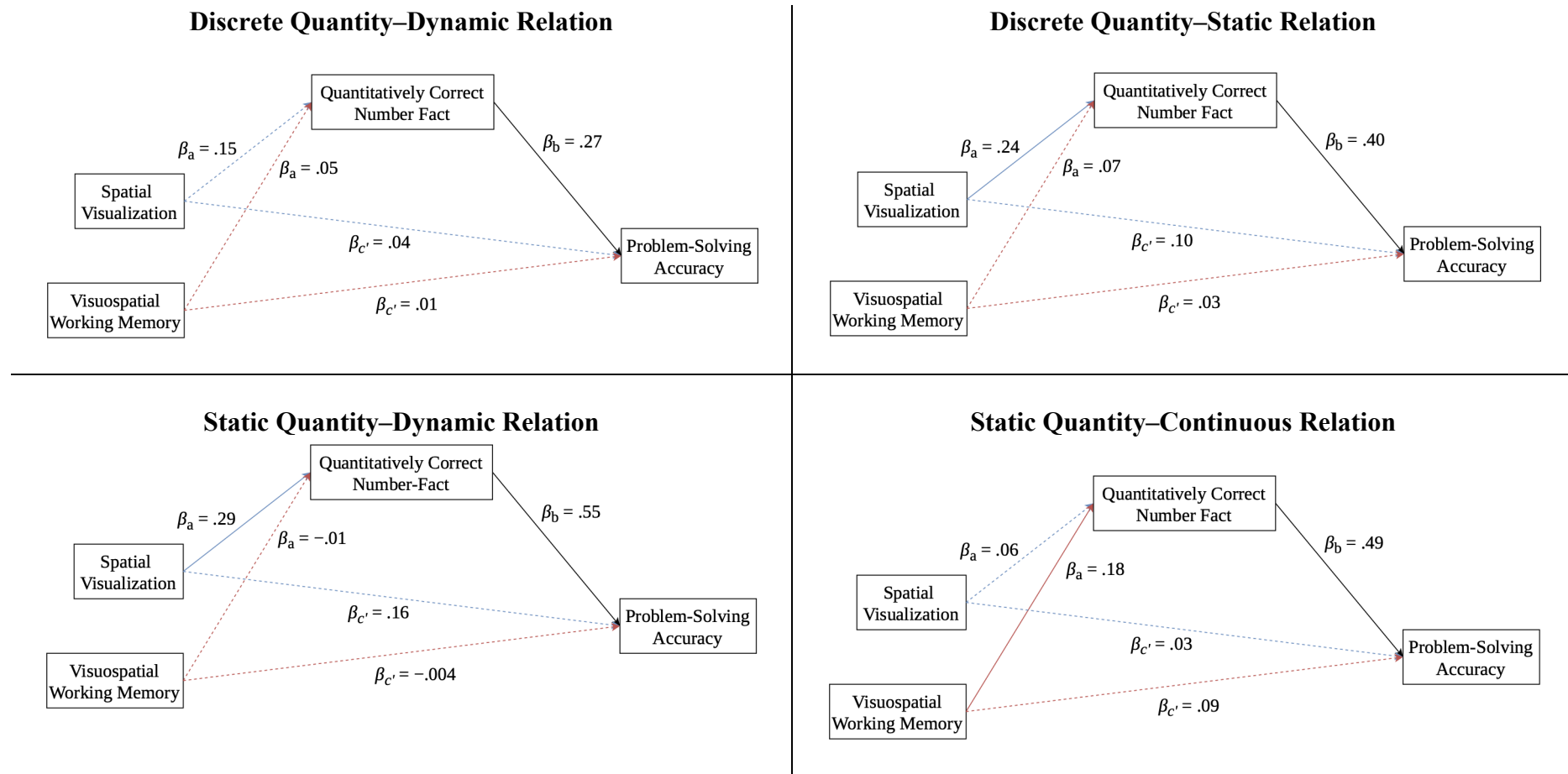
Table 5-9*Path Coefficients and Indirect Effects for Mediation Models by Problem Type*

	Strategy				Problem-Solving Accuracy				Indirect Effect			
	<i>B</i>	β	SE	95% BCa CI	<i>B</i>	β	SE	95% BCa CI	<i>B</i>	β	SE	95% BCa CI
Discrete Quantity–Dynamic Relation												
SV	0.05	.15	0.04	[−0.02, 0.12]	0.0004	.0004	0.16	[−0.31, 0.30]	0.08	.04	0.07	[−0.03, 0.24]
VSWM	0.02	.05	0.02	[−0.02, 0.06]	0.04	.02	0.11	[−0.20, 0.25]	0.02	.01	0.03	[−0.03, 0.10]
Strategy	-	-	-	-	1.56	.27	0.41	[0.75, 2.34]	-	-	-	-
Discrete Quantity–Static Relation												
SV	0.07	.24	0.03	[0.01, 0.13]	−0.29	−.14	0.15	[−0.57, 0.03]	0.20	.10	0.10	[0.04, 0.45]
VSWM	0.02	.07	0.02	[−0.02, 0.05]	0.03	.02	0.10	[−0.18, 0.22]	0.05	.03	0.05	[−0.04, 0.16]
Strategy	-	-	-	-	2.99	.40	0.57	[1.85, 4.07]	-	-	-	-
Continuous Quantity–Dynamic Relation												
SV	0.07	.29	0.03	[0.01, 0.13]	−0.27	−.15	0.15	[−0.55, 0.02]	0.29	.16	0.13	[0.05, 0.57]
VSWM	−0.001	−.01	0.02	[−0.03, 0.03]	0.15	.10	0.09	[−0.04, 0.33]	−0.01	−.004	0.07	[−0.14, 0.12]
Strategy	-	-	-	-	4.00	.55	0.55	[2.90, 5.07]	-	-	-	-
Continuous Quantity–Static Relation												
SV	0.02	.06	0.03	[−0.04, 0.07]	−0.24	−.12	0.16	[−0.53, 0.09]	0.06	.03	0.12	[−0.16, 0.31]
VSWM	0.04	.18	0.02	[0.0003, 0.08]	0.01	.003	0.10	[−0.18, 0.20]	0.15	.09	0.08	[0.01, 0.34]
Strategy	-	-	-	-	3.89	.49	0.84	[2.27, 5.54]	-	-	-	-

Note. SV = spatial visualization; VSWM = visuospatial working memory; Strategy = ‘Quantitatively Correct Number Fact’ strategy; *B* = unstandardized coefficient; β = standardized coefficient; SE = standard error. The mediator is expressed as a proportion (0–1), such that a one-unit increase represents a shift from no use to exclusive use of the strategy. In contrast, other predictors are mean centered but retain their original units. Accordingly, unstandardized coefficients (*B*) may differ in magnitude due to scaling and should be interpreted in this context. Confidence intervals (CIs) are bias-corrected and accelerated (BCa) bootstrap intervals based on 5,000 resamples. Statistical significance was evaluated based on 95% BCa CIs and is indicated by boldface. Covariate effects are reported in [Appendix 5B](#).

Figure 5-5

Mediation Models for each Word Problem Type



Note. Path coefficients are standardized estimates. Solid lines indicate effects whose 95% bias-corrected and accelerated (BCa) confidence intervals exclude zero; dashed lines indicate non-significant relations. Blue lines represent paths involving spatial visualization, and red lines represent paths involving visuospatial working memory. Covariates are omitted from the figure for clarity.

Taken together, the findings indicate that the mediating role of ‘Quantitatively Correct Number Fact’ strategy use varied across word problem types. No indirect effects were observed for Discrete Quantity–Dynamic Relation problems. In contrast, spatial visualization showed significant indirect effect for Discrete Quantity–Static Relation and Continuous Quantity–Dynamic Relation problems, whereas visuospatial working memory showed a significant indirect effect for Continuous Quantity–Static Relation problems.

A sensitivity analysis was conducted to examine whether these findings were robust to nonlinear specification of the mediator–outcome relationship. Quadratic terms for ‘Quantitatively Correct Number Fact’ strategy use were included in each model. Across all word problem types, the quadratic terms were not statistically significant, indicating that the mediator–outcome relationship was approximately linear within each problem type. The inclusion of quadratic terms did not alter the pattern of results, suggesting that the findings are robust to the functional form of the mediator–outcome relationship.

5.5 Discussion

The primary aim of the present study was to contribute to mechanistic accounts of the spatial–mathematical link by examining whether the relationship between spatial skills (spatial visualization and visuospatial working memory) and word-problem-solving accuracy was mediated by the use of a specific problem-solving strategy, namely the ‘Quantitatively Correct Number Fact’ strategy. The ‘Quantitatively Correct Number Fact’ strategy is a mental arithmetic strategy (involving derived or recalled number facts) that relies on an accurate representation of a problems’ quantitative structure. To address RQ5, all word problem types were analyzed together to test the general mediation model; these findings are discussed in [Subsection 5.5.1](#). For RQ6, word problem types were analyzed separately to provide a more nuanced understanding of how the indirect pathway functions within specific word problem types; these findings are discussed in [Subsection 5.5.2](#). Finally, educational implications in light of the findings are presented ([Subsection 5.5.3](#)), the study’s limitations are discussed ([Subsection 5.5.4](#)), and suggestions for future research are provided ([Subsection 5.5.5](#)).

5.5.1 RQ5: *Interpreting Indirect Effects of Spatial Skills across Word Problem Types*

The findings for RQ5 provide limited evidence for the mediating role of ‘Quantitatively Correct Number Fact’ strategy use in the relationship between spatial skills and problem-solving accuracy. In the primary linear model, indirect effects were not statistically significant, suggesting that this relationship is not explained by strategy use under a linear assumption.

However, when a quadratic term was included, the indirect effect of spatial visualization—but not visuospatial working memory—became statistically significant. This discrepancy likely reflects a nonlinear association between strategy use and problem-solving accuracy, as indicated by both visual inspection and the significant quadratic term in preliminary analyses.

It is noteworthy that a complementary analysis using a potential outcomes framework, which simulates counterfactual scenarios based on observed data (Imai et al., 2010), also provided evidence for an indirect effect of spatial visualization on problem-solving accuracy via strategy use (Seo et al., in press). The discrepancy between these findings and those obtained in the present linear path-analytic model suggests that conclusions regarding mediation may depend on the analytical framework and underlying assumptions. In particular, whereas the potential outcomes approach estimates average causal mediation effects, the present model relies on assumptions of linearity and model specification, which may not fully capture the conditional nature of the mediator–outcome relationship observed in the data.

Taken together, the findings do not provide conclusive evidence for a mediating role of problem-solving strategy use between spatial visualization and problem-solving accuracy, suggesting that any such role may be complex and sensitive to modeling assumptions. One possible explanation is that problem-solving accuracy was aggregated across heterogeneous word problem types that differ in their quantitative structure. Prior findings from Studies [1](#) and [2](#) indicate that the relationship between spatial skills and both problem-solving accuracy and problem-solving strategy use varies across word problem types. Aggregating across these heterogeneous structures may therefore obscure mediation effects that are specific to the quantitative structure of word problem types. Accordingly, RQ6 examines whether mediation patterns differ across word problem types.

5.5.2 RQ6: Interpreting Indirect Effects of Spatial Skills by Word Problem Type

Consistent with the suggestion that mediation effects may be obscured when aggregating across problem types, RQ6 examined whether indirect effects differed by word problem type. The results indicated that mediation effects were not uniform across problem types but instead emerged selectively depending on the quantitative structure of the problem. Specifically, ‘Quantitatively Correct Number Fact’ strategy use mediated the relationship between spatial visualization and problem-solving accuracy for Discrete Quantity–Static Relation and Continuous Quantity–Dynamic Relation problems. In contrast, this strategy use mediated the relationship between visuospatial working memory and problem-solving

accuracy for Continuous Quantity–Static Relation problems. No indirect effects were observed for Discrete Quantity–Dynamic Relation problems.

5.5.2.1 Spatial Visualization Supports the Coordination of Representation

The two word problem types for which indirect effects of spatial visualization were observed share a common feature: both place greater demands on the coordination of representations of quantitative structure, either through comparison or transformation. Thus, as problem solvers construct a representation of a quantitative structure, spatial visualization may provide a “mental blackboard” on which such representations can be constructed, manipulated, or transformed (Hawes & Ansari, 2020, p. 472).

For Discrete Quantity–Static Relation problems, spatial visualization may provide a common representational space in which two disjoint sets can be compared (i.e., more or less). Importantly, similar to verbal–analytical problems (Casey et al., 2015), although these problems do not explicitly describe spatial situations, the network of quantities and quantitative relations may nevertheless be organized spatially in the representation. In this way, spatial visualization may be involved in the organization of representations of quantitative structure.

For Continuous Quantity–Dynamic Relation problems, which describe changes along a continuous dimension (i.e., walking toward or away from a wall), spatial visualization may support the transformation of representation. Solving such problems requires coordinating initial and final positions and the change between them. Spatial visualization may facilitate this process by supporting the organization and manipulation of these quantitative structures within a representational space, thereby contributing indirectly to successful problem solving.

Notably, for these two problem types only, language ability predicted problem-solving accuracy but did not predict the use of the ‘Quantitatively Correct Number Fact’ strategy. The influence of linguistic task features (e.g., lexical, semantic, and structural complexity) and problem solvers’ language ability in mathematical word problem solving has been well documented (Jaffe & Bolger, 2023; Verschaffel et al., 2020; Vessonen et al., 2024), which is unsurprising given that mathematical word problems are presented as verbal descriptions. It is therefore noteworthy that the supportive role of spatial visualization extends beyond language ability and appears to be unique in its mechanism. Spatial visualization predicted the use of the ‘Quantitatively Correct Number Fact’ strategy but did not directly predict accuracy (consistent with prior findings from [Study 1](#)); instead, its influence on performance was indirect, operating through strategy use.

This dissociation—language ability predicting accuracy and spatial visualization predicting strategy use—suggests that language ability and spatial visualization contribute differently to mathematical word problem solving. Language ability may support comprehension of the problem statement, whereas spatial visualization may support the representation of quantitative structure, particularly in terms of coordinating and transforming these representations. Future research is needed to further investigate the mechanisms through which language ability and spatial visualization make distinct contributions to mathematical word problem solving. Nonetheless, the present findings suggest that young pupils’ spatial visualization supports quantitative reasoning, particularly in problem types that require coordinating comparisons or transforming representations of quantitative structure, consistent with previous descriptions of spatial visualization as enabling individuals to “perceive, maintain, *manipulate* [emphasis added], and generate visual-spatial solutions” (Hawes & Ansari, 2020, p. 475).

5.5.2.2 Visuospatial Working Memory Supports the Maintenance of Representation

Visuospatial working memory’s indirect effect on solving Continuous Quantity–Static Relation problems mediated through ‘Quantitatively Correct Number Fact’ strategy use may be attributed to the representational demands of this word problem type, which emphasize maintaining multiple quantitative components simultaneously. Continuous Quantity–Static Relation problems describe a relational comparison (i.e., farther from or closer to the wall) between two disjoint positions and require holding the reference point (the wall) and two continuous quantities in mind at the same time. In contrast to word problem types that place greater demands on the coordination and transformation of representations of quantitative structure, successful problem solving on this problem type depends more on the capacity to maintain multiple quantitative components within a representation—a function more closely associated with visuospatial working memory.

The present findings therefore indicate distinct but complementary roles for spatial visualization and visuospatial working memory in mathematical word problem solving, consistent with the interpretation advanced in [Study 2](#). Spatial visualization is primarily associated with problems whose quantitative structure places greater demands on the coordination and transformation of representations of quantitative structure, whereas visuospatial working memory appears more strongly associated with problems whose quantitative structure places greater demands on the maintenance of multiple quantitative

components within representations. This interpretation may also help explain why this word problem type had the highest levels of failure to engage with the quantitative structure but not the highest incidence of quantitatively-incorrect representations (see [Study 2](#)). That is, visuospatial working memory may function as a capacity- constraint: a sufficient level may be necessary for problem solvers to meaningfully engage with the problem's quantitative structure at all. When visuospatial working memory resources are sufficient, solvers may be able to maintain a representation, although this does not guarantee its accuracy. In contrast, when such resources are insufficient, solvers may fail to maintain the representation and therefore fail to engage with the problem's quantitative structure altogether.

This distinction between spatial visualization and visuospatial working memory is consistent with prior evidence suggesting that these two spatial skills contribute to mathematical performance through partially independent pathways (Hawes & Ansari, 2020). Factor-analytic studies indicate that, although the two spatial skills are highly related, they are separable constructs. For example, Hawes et al. (2019b) identified spatial visualization and visuospatial working memory as distinct latent factors contributing independently to mathematical performance. Developmental findings further suggest that these abilities may play different roles across age groups (Mix et al., 2016), although results have been mixed. Nevertheless, converging evidence indicates that spatial visualization and visuospatial working memory make unique contributions to mathematical problem solving (e.g., Holmes & Adams, 2006; Li & Geary, 2017; McKenzie et al., 2003).

One account proposes that spatial visualization supports the spatial modeling of novel content, whereas visuospatial working memory supports the procedural fluency of familiar content (Mix, 2019). The present findings do not fully align with this characterization. Assuming that the Key Stage 1 National Curriculum Assessments reflect the types of problems pupils typically encounter in classrooms, an analysis of these assessments suggests that Continuous Quantity–Static Relation problems are relatively rare (see [Study 2](#)). As such, this problem type may be better characterized as novel rather than familiar for this study's sample of pupils. A more precise interpretation is therefore that spatial visualization and visuospatial working memory play distinct but complementary roles in mathematical problem solving, corresponding respectively to the coordination and transformation of representations of quantitative structure and to their maintenance.

The indirect effect of visuospatial working memory should nevertheless be interpreted with caution. Although indirect effects were evaluated using bias-corrected and accelerated

(BCa) bootstrap confidence intervals (Cumming, 2008; Efron & Tibshirani, 1993), which are generally considered more appropriate for mediation than reliance on p -values alone, the p -value associated with the indirect effect was marginal ($p = .07$). Furthermore, recent work suggests that percentile-based bootstrap intervals may yield more conservative Type I error rates in certain mediation contexts (Alfons & Schley, 2025; Biesanz et al., 2010; Chen & Fritz, 2021). In the present analyses, the indirect effect of visuospatial working memory was statistically significant when evaluated using BCa intervals but not when evaluated using percentile-based intervals (see [Appendix 5D](#)), indicating sensitivity to the choice of bootstrap method. BCa intervals were retained because they adjust for both bias and skewness in the sampling distribution of the indirect effect (Biesanz et al., 2010). Nevertheless, the indirect effect should be interpreted with caution and warrants replication. It remains unclear whether the discrepancy between the BCa intervals, the marginal p -value, and the percentile-based intervals reflects limited statistical power or a potential Type I error. Future studies with larger samples are needed to clarify this issue.

Finally, although not central to the study's aims, gender emerged as a significant predictor of 'Quantitatively Correct Number Fact' strategy use for Continuous Quantity–Static Relation problems only, with boys more likely than girls to employ this strategy, even after controlling for spatial skills, and other demographic and cognitive variables. Notably, gender did not predict problem-solving accuracy, suggesting that gender-related differences may be expressed at the level of strategy use rather than performance outcomes. One possibility is that girls may be more likely to rely on 'Quantitatively Correct Counting' strategies for this problem type. While both strategies involve accurate representations of quantitative structure, they differ in their arithmetic approach. This interpretation is consistent with prior findings indicating that girls are more likely to use counting-based strategies than boys (Casey et al., 2017; Laski et al., 2013; Vasilyeva et al., 2025), partly because they tend to rely on strategies that are more closely tied to formal instruction, whereas boys tend to use more inventive strategies (Casey & Ganley, 2021).

As Casey et al. (2017) suggest, given parallel gender differences in spatial skills and the use of advanced arithmetic strategies, arithmetic strategy may be a suitable target for early intervention at the spatial–mathematical intersection to support girls' mathematical development. Such interventions may help reduce gender disparities in STEM-related fields, particularly in domains that rely heavily on spatial reasoning (Wai et al., 2009). The reasons why this gender difference emerged only for this problem type remain unclear. Because gender

was not a primary focus of the present study, further research is needed to determine the robustness and generality of this effect.

5.5.2.3 Limits of Strategy Mediation in Discrete Quantity–Dynamic Relation Problems

An indirect effect of either spatial skills was not found for Discrete Quantity–Dynamic Relation problems, which describe the winning or losing of marbles (e.g., a change in set size). One potential explanation for this is that this word problem type allows for strategy heterogeneity, inviting multiple successful approaches. For instance, pupils may directly model the problem situation using counters or drawing (Verschaffel et al., 2020). Indeed, descriptively, this problem-type had the highest incidence of ‘Quantitatively Correct Counting’ strategies (see [Study 2](#)). It is therefore plausible that, in this word problem type—where problem can be solved using an alternative strategy (e.g., ‘Quantitatively Correct Counting’)—the indirect effect of spatial skills via ‘Quantitatively Correct Number Fact’ strategy use is attenuated. Instead, it may be that the relationship between spatial skills and problem-solving accuracy in this problem type is mediated through another intervening process.

Another possibility is that for this word problem type, which is frequently encountered at Key Stage 1, problem solvers may have developed a repertoire of representational schemas in long-term memory (Kintsch & Greeno, 1985; Riley et al., 1983), such that spatial skills are not required to the same extent. However, this proposition appears inconsistent with the findings of [Study 1](#), which showed that, at least descriptively, the association between spatial visualization and problem-solving accuracy was most positive for Discrete Quantity–Dynamic Relation problems. Future research should further investigate whether the limits of mediation observed for this word problem type arise from reduced representational demands, increased familiarity, or the availability of alternative successful strategies.

5.5.3 Educational Implications

In addition to its theoretical significance for understanding how young pupils’ spatial skills and arithmetic word-problem-solving performance are related, the present study also has practical implications for instruction. Specifically, the findings identify problem-solving strategies as a promising target for leveraging the spatial–mathematical link in educational contexts and future intervention studies. The following instructional recommendations highlight the potential value of embedding spatial supports within word-problem instruction.

First, rather than targeting spatial skills directly or in isolation, young pupils may benefit from support in using a range of tools (e.g., drawings, part-whole models, number lines) that scaffold the transition from counting-based strategies to number-fact strategies while also supporting the coordination and transformation of representations of a problem's quantitative structure. Some educational practitioners may be wary that such externalized tools for modeling the problems could inhibit progression towards internalized strategies. However, findings from recent studies showing that finger counting facilitates the transition to internalized mental arithmetic strategies (Krenger & Thevenot, 2025; Poletti et al., 2022) suggest that externalized tools may help make explicit quantitative and numerical relationships that are still fragile in children's minds at an early age. As McNeil et al. (2025) put effectively, "the pathway toward developing expertise does not begin with teaching children to emulate experts. It begins with nurturing children's understanding of the foundations and linking them to the formal math symbols" (p. 27). The use of such tools should be especially beneficial for pupils with lower working memory capacity (Thevenot & Barrouillet, 2015).

More than merely providing access to these tools, teachers should explicitly model and teach how to use them, as well as explain their benefits. For example, young pupils are unlikely to spontaneously produce visual representations when solving word problems (Jitendra & Woodward, 2019), even when such representations would facilitate the coordination, transformation, and maintenance of representations of quantitative structure. Indeed, the present study also found that although pupils had access to paper and pencil, very few made use of them, even after being explicitly told that they may be helpful. It is therefore important to directly teach young pupils how to generate visual representations that highlight the quantitative structure of the word problem, even when the problems do not describe situations that are overtly spatial in nature.

Second, young pupils should be exposed to a wide variety of problems to challenge the habitual use of familiar strategies. For example, problems should vary in terms of the quantitative structure. Rather than practicing only final-quantity-unknown problems, pupils should also encounter initial-quantity-unknown and dynamic-relation-unknown problems (Verschaffel & De Corte, 1997). Similarly, instead of being given only comparison-quantity-unknown problems, they should gain familiarity with referent-quantity-unknown and static-relation-unknown problems (Múñez et al., 2013). Moreover, rather than focusing solely on one-step additive word problems comprising three quantities, or two quantities and one

quantitative relation, pupils should also be exposed to problems comprising three quantitative relations (Vergnaud, 1982).

Additionally, classroom instruction should include problems that can be solved using both spatial and analytical strategies. For example, the continuous-quantity problems used in this study require constructing a representation of the underlying spatial relation. In the currently mathematics curriculum, however, emphasis is often placed on discrete-quantity problems—particularly Discrete Quantity–Dynamic Relation problems—which the present study suggests may be solved without the indirect effects of spatial ability. Another type of problems that may encourage the use of spatial strategies is measurement problems, which “bridges two main areas of school mathematics—geometry and number” (National Council of Teachers of Mathematics, 2000, p. 103). More specifically, young pupils may benefit from greater exposure to measurement problems that encourage the construction of mental models (e.g., distance-estimation map problem), rather than solely the application of formulas (e.g., calculation of the perimeter) (Vasilyeva et al., 2009).

5.5.4 Limitations

It is important to acknowledge the study’s limitations when interpreting the findings and drawing conclusions. First, the most prominent limitation is its correlational, cross-sectional design. Although mediation analyses may be consistent with a causal process, causality cannot be definitively established in cross-sectional data (Preacher, 2015). Moreover, some research suggests that the relationship between spatial and mathematical abilities may be bidirectional, further complicating causal interpretations (e.g., Geer et al., 2019; Lombardi et al., 2019). Nevertheless, the present study grounds its hypothesized causal mechanism in strong theoretical assumptions and demonstrates that the mediation results are consistent with the proposed mechanism. Moreover, because “inference regarding a true mediating mechanism is most likely to emerge based on a body of research evidence rather than one study” (MacKinnon, 2008, p. 10), the present findings provide a direction for future longitudinal and experimental research.

Second, given the study’s sample size, the analysis employed manifest variables rather than latent variable mediation models. Although latent variable approaches are not a panacea for all sources of measurement error (Casey et al., 2017), they are generally more effective than path analysis in accounting for measurement error, which can bias estimates (Hoyle & Kenny, 1999). In the present study, this limitation was partially mitigated by using well-validated

measures (Alfons & Schley, 2005). Moreover, simulation studies suggest that unreliability in the measures typically leads to underestimation of indirect effects (Cole & Maxwell, 2003). Thus, the fact that significant indirect effects were observed despite the use of manifest variables suggests that the mediated associations were sufficiently strong to be detected even in the presence of measurement error.

Third, the assumption of independent observations was technically violated because pupils were nested within classrooms. Multilevel modeling was not pursued due to the small number of clusters and limited sample size within each cluster. When independence of assumptions is violated, standard errors may be underestimated, potentially inflating Type I error rates and producing overly narrow confidence intervals (Preacher, 2015). Conceptually, it is possible that teachers in the present study differed in their instructional practices regarding the teaching of representations of quantitative structure and arithmetic strategies. Future research should therefore consider incorporating classroom-level variables to better understand the influence of instructional context.

Finally, the present study examined mediation at the participant (between-person) level. Strategy use was operationalized as the proportion of problems solved using the ‘Quantitatively Correct Number Fact’ strategy, aggregated across all 24 items for RQ5 and across subsets of six items for each of the four mediation models in RQ6. Although this approach was appropriate for addressing the research questions and examining associations at the level of individual differences, aggregating across items necessarily obscures within-person variability and limits the examination of item-level processes. Consequently, the findings reflect associations between pupils’ spatial skills and their overall tendency to use particular strategies, rather than testing whether spatial skills have an indirect effect on accuracy at the item level through the use of ‘Quantitatively Correct Number Fact’ strategies on those same items (Curran & Bauer, 2011). In addition, strategy proportions are compositional in nature, meaning that increases in one strategy necessarily imply decreases in others (Aitchison, 1982). Although the mediation models focused on a theoretically central strategy, this approach does not fully capture the interdependent structure of children’s broader strategy profiles. Future research employing item-level or multilevel modeling approaches could provide a more fine-grained test of the proposed mechanism by examining within-person processes and the full composition of strategy use (Preacher et al., 2010).

5.5.5 Conclusion

There are two different temporal perspectives for understanding the mechanistic account of spatial–mathematical link. One concerns how the two domains relate in real-time, and the other concerns how they relate across developmental time (Mix, 2019). The present study focused on the former, examining the cognitive processes that may provide intervening pathways between young pupils’ spatial ability and arithmetic word-problem-solving performance. The findings indicate that spatial visualization had an indirect effect on one-step additive word-problem-solving performance via problem-solving strategy use for word problem types that place greater demands on the coordination and transformation of representation of the quantitative structure. Visuospatial working memory, in contrast, had an indirect effect for word problem types that place greater demands on maintaining the representation.

Future research should investigate whether the proposed mechanism is truly causal through carefully designed longitudinal and experimental studies. First, longitudinal designs in which spatial ability is measured at Time 1 and problem-solving strategies and performance are assessed at a later time point would help clarify the temporal ordering of spatial skills and arithmetic word problem solving via strategy use. Likewise, intervention studies with active control groups targeting the instructional approaches outlined above would provide stronger evidence for causal claims.

Second, future research should examine the qualitative aspect of pupils’ representations of quantitative structure. In the present study, young pupils’ quantitative reasoning was assessed by observing how they represented the quantitative structure of word problem, rather than requiring number-sentence construction (Wong & Ho, 2017) or restricting them to a fixed set of conceptual diagrams (Yip et al., 2020). A next step would be to analyze these rich data more systematically to gain a deeper understanding of children’s problem-solving strategies and how they may relate to spatial skills.

6 General Discussion

Arithmetic word problem solving, as a key component of early mathematics, is an important mathematical skill that predicts later academic and career success (Duncan et al., 2007; Murnane et al., 2001). Despite robust evidence of a spatial–mathematical link (Hawes & Ansari, 2020; Uttal et al., 2013), there is relatively limited research that systematically investigates the relationship between early primary school pupils’ spatial ability and arithmetic word problem solving. Moreover, relatively little research has examined the cognitive mechanisms underlying this relationship (Mix & Cheng, 2012), although some studies have explored potential mechanisms in other mathematical domains (e.g., Foley et al., 2017; Gunderson et al., 2012; Laski et al., 2013; Möhring et al., 2024).

The aim of this integrated thesis was to contribute to addressing this gap in the spatial–mathematical literature by examining the relationship between young primary school pupils’ (6–7-year-olds’) spatial visualization and visuospatial working memory, and their one-step additive word-problem-solving strategies and accuracy. To conclude this thesis, [Subsection 6.1](#) summarizes the theoretical contributions. [Subsection 6.2](#) summarizes the main empirical findings from each of the three studies. [Subsection 6.3](#) synthesizes the findings to highlight their contribution to the nuanced understanding of the relationship between young primary school pupils’ spatial ability and one-step additive word problem solving. [Subsection 6.4](#) discusses the educational implications in light of the findings. [Subsection 6.5](#) addresses limitations of the thesis and suggests directions for future research, and the thesis closes with [Subsection 6.6](#), which presents concluding remarks.

6.1 Summary of Theoretical Contributions

This thesis contributes to the literature on the relationship between young primary school pupils’ spatial ability and one-step additive word problem solving in two main theoretical ways. First, the thesis proposes a 2×2 conceptual framework for categorizing one-step additive word problems, comprising two quantities and a quantitative relation, based on the problem’s quantitative structure. Through the integration of Thompson’s (1993) distinction between quantity and quantitative relation, with Davydov’s (1982) discussion of quantification methods (discrete and continuous quantities), and Vergnaud’s (1982) classification of quantitative relations (dynamic and static relations), this 2×2 framework yields four distinct word problem types: Discrete Quantity–Dynamic Relation, Discrete Quantity–Static Relation, Continuous Quantity–Dynamic Relation, and Continuous Quantity–Static Relation. Moreover,

as shown in Figure 3-1, this thesis adapted Vergnaud's (1982) notation system such that upright squares ($\square$) represented discrete quantities, tilted squares ($\diamond$) represented continuous quantities, while circles with a horizontal arrow ($\overset{\circ}{\rightarrow}$) represented dynamic relations and circles with a vertical arrow ($\overset{\circ}{\uparrow}$) represented static relations. Furthermore, informed by the conceptual framework, four parallel forms of 24 word problems (six problems per word problem type) were developed.

Second, the thesis proposes a holistic framework for categorizing problem-solving strategies along the dimensions of arithmetic strategies and representations of quantitative structure. This framework is based on the view that both arithmetic skills and quantitative reasoning are important for solving mathematical word problems (Nunes et al., 2012), and young pupils' problem-solving processes do not necessarily unfold in sequential phases of 'representation' and 'computation' (Carpenter & Moser, 1984; Nunes & Bryant, 1996). This framework yields five problem-solving strategies: 'Non-Quantitative,' 'Quantitatively Incorrect Counting,' 'Quantitatively Incorrect Number Fact,' 'Quantitatively Correct Counting,' and 'Quantitatively Correct Number Fact.'

6.2 Summary of Empirical Findings

This thesis contributes to the literature on the relationship between young primary school pupils' spatial ability and one-step additive word problem solving through three main empirical research themes. The first research theme investigates the association between spatial skills and accuracy in word problem solving ([Subsection 6.2.1](#)). The second research theme investigates the association between spatial skills and problem-solving strategy ([Subsection 6.2.2](#)). Finally, the third research theme examines the indirect effect of spatial skills on problem-solving accuracy via problem-solving strategy ([Subsection 6.2.3](#)). Below, the main findings of each research theme are summarized.

6.2.1 Findings of Research Theme 1: Spatial Skills and One-Step Additive Word-Problem-Solving Accuracy

Neither spatial visualization nor visuospatial working memory independently predicted word-problem-solving accuracy after controlling for language ability, numerical ability, fluid reasoning, age, and gender. However, although the simple slopes within each word problem type were not statistically significant, the association between spatial visualization and problem-solving accuracy differed between Discrete Quantity–Dynamic Relation problems and Continuous Quantity–Static Relation problems. In contrast, the association between

visuospatial working memory and problem-solving accuracy did not vary significantly across word problem types.

6.2.2 Findings of Research Theme 2: Spatial Skills and Problem-Solving Strategies

Across all word problem types, higher spatial visualization was associated with greater relative use of ‘Quantitatively Correct Number Fact’ strategies compared to ‘Quantitatively Correct Counting’ and ‘Quantitatively Incorrect Counting’ strategies. In contrast, the association between visuospatial working memory and problem-solving strategies varied across word problem types. For most word problem types, visuospatial working memory was associated with reduced relative use of ‘Non-Quantitative’ strategies. For Continuous Quantity–Static Relation problems, visuospatial working memory was associated with reduced relative use of ‘Quantitatively Incorrect Counting’ strategies, compared to Discrete Quantity–Dynamic Relation problems.

6.2.3 Findings of Research Theme 3: Spatial Skills and One-Step Additive Word-Problem-Solving Accuracy: The Mediating Role of Problem-Solving Strategies

The mediating effects of spatial visualization and visuospatial working memory, via the use of ‘Quantitatively Correct Number Fact’ strategies, were not uniform across word problem types. Instead, the presence and strength of mediation varied as a function of the problems’ quantitative structure. Spatial visualization showed significant indirect effects for Discrete Quantity–Static Relation and Continuous Quantity–Dynamic Relation problems, indicating that higher spatial visualization was associated with greater use of ‘Quantitatively Correct Number Fact’ strategies, which in turn related to higher accuracy for solving these problem types. In contrast, visuospatial working memory had an indirect effect on problem-solving accuracy for Continuous Quantity–Static Relation problems via ‘Quantitatively Correct Number Fact’ strategies. Neither spatial visualization nor visuospatial working memory had an indirect effect on problem-solving accuracy via ‘Quantitatively Correct Number Fact’ strategies for Discrete Quantity–Dynamic Relation problems. Table 6-1 provides a summary of all the main empirical findings related to each research theme.

Table 6-1*Summary of Main Empirical Findings from each Study*

Study	Research Question	Spatial Visualization	Visuospatial Working Memory
Research Theme 1	RQ1: Spatial Skills and Word-Problem-Solving Accuracy	<i>ns</i>	<i>ns</i>
	RQ2: Spatial Skills and Word-Problem-Solving Accuracy by Problem Type	Association differ between DQ–DR and CQ–SR problems	<i>ns</i>
Research Theme 2	RQ3: Spatial Skills and Word-Problem-Solving Strategy Use	Relatively less ‘Quantitatively Correct Counting’ and ‘Quantitatively Incorrect Counting’ strategies	<i>ns</i>
	RQ4: Spatial Skills and Word-Problem-Solving Strategy Use by Problem Type	<i>ns</i>	Relatively less ‘Non-Quantitative’ strategies for DQ–DR, DQ–SR, and CQ–DR problems; Relatively less ‘Quantitatively Incorrect Counting’ strategies for CQ–SR problems
Research Theme 3	RQ5: Indirect Effect of Spatial Skills on Word-Problem-Solving Accuracy via ‘Quantitatively Correct Number Fact’ Strategies	<i>ns</i>	<i>ns</i>
	RQ6: Indirect Effect of Spatial Skills on Word-Problem-Solving Accuracy via ‘Quantitatively Correct Number Fact’ Strategies by Problem Type	Indirect effects in DQ–SR and CQ–DR problems	Indirect effect in CQ–SR problems

Note. DQ–DR: Discrete Quantity–Dynamic Relation; DQ–SR: Discrete Quantity–Static Relation; CQ–DR: Continuous Quantity–Dynamic Relation; CQ–SR: Continuous Quantity–Static Relation; *ns* = Non-Significant.

6.3 Synthesis of Findings

Results across the three studies are synthesized to derive three key conclusions about the role of spatial ability in young pupils' arithmetic word problem solving. The key conclusions are that: (1) the quantitative structure of the problem matters ([Subsection 6.3.1](#)), (2) the type of spatial skill matters ([Subsection 6.3.2](#)), and (3) the problem-solving strategy use matters ([Subsection 6.3.3](#)). They are elaborated below by first presenting the relevant findings from the individual studies, then synthesizing them to draw overarching conclusions, and finally discussing their broader significance. [Subsection 6.3.4](#) summarizes the synthesis and highlights the thesis's contribution to the literature.

6.3.1 *The Quantitative Structure of the Problem Matters*

Study 1 showed that the association between spatial visualization and problem-solving accuracy differed between Discrete Quantity–Dynamic Relation and Continuous Quantity–Static Relation problems. Although the simple slopes within each of these problem types were not statistically significant, the significant interaction between spatial visualization and word problem type indicates that this difference between these problem types was reliable. Similarly, Study 2 showed that the association between visuospatial working memory and problem-solving strategy use also differed between these same problem types. In addition, findings from Studies 1 and 3 showed that neither spatial visualization nor visuospatial working memory was consistently associated with arithmetic word problem solving across different quantitative structures after controlling for language ability, numerical ability, fluid reasoning, age, and gender, regardless of whether problem-solving strategy use was included as an additional covariate.

Taken together, these findings indicate that ***the quantitative structure of the problem matters***. In young pupils, spatial ability is not uniformly predictive of arithmetic word problem solving; instead, its association varies systematically as a function of the quantitative structure of the problems.

This finding contributes to the literature on the relationship between young pupils' spatial ability and arithmetic word problem solving, an area of research that has been underexplored. The limited research that does exist has yielded mixed findings with some studies reporting a significant association (e.g., Chan & Wong, 2019; Fung & Swanson, 2017; Gilligan et al., 2017; Kyttälä et al., 2014; Mix et al., 2017, 2021; Zhang & Lin, 2015, 2017; Zheng et al., 2011), while others report null findings (e.g., Campos et al., 2013; Carr et al.,

2008; Ching & Nunes, 2017; Fuchs et al., 2010; Rodán et al., 2019). This inconsistency may partly reflect the lack of integration between the spatial reasoning and quantitative reasoning literatures, which have often been examined in isolation.

By systematically considering the quantitative structure of problems through a 2×2 conceptual framework for categorizing one-step additive word problems, this thesis shows that the relationship between spatial ability and arithmetic word problem solving is not uniform across word problem types. By controlling for key covariates known to contribute to word problem solving (i.e., language ability, numerical ability, fluid reasoning, age, and gender; Peng et al., 2020; Verschaffel et al., 2020; Vessonen et al., 2024), the thesis reduces the risk of confounding due to shared variance, allowing for more precise estimation of the associations between spatial ability and arithmetic word problem solving.

6.3.2 The Type of Spatial Skill Matters

Study 1 showed that the association between spatial visualization and problem-solving accuracy varied as a function of quantitative structure, whereas visuospatial working memory was not differentially associated with accuracy across problem types. Study 2 further showed that spatial visualization was associated with the use of number-fact strategies and with the correct representation of quantitative structure. In contrast, visuospatial working memory was associated with engagement with the problem's quantitative structure for most problems and, for some problems, with the use of number-fact strategies and correct representation of quantitative structure.

Taken together, these findings indicate that ***the type of spatial skill matters***. **In young pupils, spatial visualization and visuospatial working memory contribute differently to arithmetic word problem solving.**

By examining these two spatial skills simultaneously, the thesis demonstrates that spatial visualization and visuospatial working memory have distinct and complementary roles in word problem solving, thereby extending prior work suggesting that spatial ability is multifaceted (Mix & Cheng, 2012) and that specific spatial skills share unique processing with specific mathematical skills (Hawes et al., 2019b; Mix, 2019). Spatial visualization and visuospatial working memory appear to support distinct cognitive processes during arithmetic word problem solving, and their relevance depends on the representational demands imposed by the quantitative structure of the problem.

In particular, the findings suggest that spatial visualization may contribute indirectly to performance on problems that place greater demands on the *coordination* and *transformation* of representations of quantitative structure, whereas visuospatial working memory may contribute indirectly to performance on problems that place greater demands on the *maintenance* of these representations. This refines existing characterizations of spatial visualization as supporting the generation of novel visuospatial solutions for more difficult problems and of visuospatial working memory as supporting the recall of familiar or easier problem solutions (Hawes et al., 2019b; Mix, 2019; Mix et al., 2016, 2017). Moreover, these findings challenge accounts that attribute the relationship between spatial visualization and mathematical performance solely to visuospatial working memory (Hawes & Ansari, 2020), suggesting instead that spatial visualization contributes beyond the functions typically associated with working memory.

6.3.3 The Problem-Solving Strategy Use Matters

Study 2 showed that spatial visualization was consistently associated with reduced use of counting-based and quantitatively-incorrect strategies. Visuospatial working memory, in contrast, was associated with reduced use of ‘Non-Quantitative’ strategies across most problem types, and for Continuous Quantity–Static Relation problems, was more consistently associated with reduced use of counting-based and quantitatively-incorrect strategies. Study 3 showed that strategy use mediated the relationship between spatial visualization and problem-solving accuracy for Discrete Quantity–Static Relation and Continuous Quantity–Dynamic Relation problems, which may place greater demands on the coordination and transformation of representations of quantitative structure. In contrast, strategy use mediated the relationship between visuospatial working memory and problem-solving accuracy for Continuous Quantity–Static Relation problems, which may place greater demands on the maintenance of these representations. Study 3 further showed that, for Discrete Quantity–Dynamic Relation problems, spatial skills did not operate through the strategy pathway examined in the present thesis. One possible explanation is that these problems afford multiple successful solution pathways, including counting-based approaches.

Taken together, these findings indicate that **the problem-solving strategy use matters**. **In young pupils, problem-solving strategy use provides a pathway linking spatial ability and arithmetic word-problem-solving accuracy, as a function of the quantitative structure of the problems and the type of spatial skill.**

More specifically, these findings indicate that the ‘Quantitatively Correct Number Fact’ strategy provides a pathway linking spatial ability and arithmetic word-problem-solving accuracy, but this pathway is not uniform across problem types. Instead, the extent to which problem-solving strategy use mediates this relationship depends on the quantitative structure of the problem and the type of spatial skill.

More broadly, the results indicate that there is not a single underlying cognitive mechanism that explains the spatial–mathematical link, reflecting the non-unitary nature of spatial and mathematical abilities (Lourenco et al., 2018; Mix & Cheng, 2012). Instead, problem-solving strategy use represents one of several pathways through which spatial ability may relate to arithmetic word problem solving. This interpretation is consistent with prior research in other mathematical domains, primarily arithmetic computation tasks, which has identified additional mechanisms underlying the spatial–mathematical link, such as arithmetic strategy use (Casey et al., 2017; Foley et al., 2017; Laski et al., 2013; Möhring et al., 2024) and number line estimation (Gunderson et al., 2012; LeFevre et al., 2013; Tam et al., 2019). In contrast, numerical magnitude knowledge has not been found to mediate the relationship (Vasilyeva et al., 2025). Taken together, these findings suggest that spatial ability contributes to word problem solving through multiple, structure-dependent pathways rather than a single general process.

Regardless of whether strategy use functions as a mediator, it remains an important component of word problem solving, consistent with literature highlighting the role of quantitative reasoning (Nunes et al., 2012). In this thesis, the inclusion of the ‘Quantitatively Correct Number Fact’ strategy was associated with a substantial increase in explained variance, from a marginal pseudo- R^2 of .29 in the RQ1 model from Study 1 to an R^2 of .69 in the RQ5 model from Study 3, highlighting the potential importance of strategy use in accounting for variability in word-problem-solving accuracy (Nunes et al., 2015). However, these values are not directly comparable because they were derived from different statistical models (i.e., a mixed-effects logistic regression model and a path-analytic model), represent different measures of explained variance (marginal pseudo- R^2 versus R^2), and were estimated at different levels of analysis (item-level versus participant-level). Consequently, the comparison should be interpreted as a broad indication of the potential contribution of strategy use rather than a precise estimate of incremental variance explained.

6.3.4 Contribution to Literature

In summary, these three conclusions highlight the importance of considering how the quantitative structure of word problems shapes their representational demands and the spatial skills recruited during problem solving. Rather than exerting a uniform influence, spatial ability appears to support different aspects of problem solving depending on the problem's quantitative structure. In particular, different problem types place varying demands on how representation of quantitative structure is coordinated, transformed, and maintained, which in turn determines which spatial skills are most relevant and whether strategy use functions as a pathway to successful problem solving. This perspective moves beyond viewing spatial ability as a general predictor of mathematical performance and instead emphasizes the interaction between spatial skills and mathematical task structure.

As such, a central contribution of this thesis is the more nuanced understanding it offers of the spatial–mathematical link in arithmetic word problem solving during the early primary years. By integrating quantitative structure, spatial skills, and problem-solving strategy use within a unified framework, the thesis provides a more differentiated account of how spatial ability contributes to one-step additive word problem solving.

This more nuanced perspective offers important implications for educational practice, particularly in relation to how to harness the spatial–mathematical link in early primary arithmetic word problem instruction. At the same time, it highlights important theoretical and methodological considerations that delimit the scope and generalizability of these findings. These considerations inform the educational implications discussed next, as well as the limitations and directions for future research that follows.

6.4 Educational Implications

Two main approaches have been proposed for harnessing the spatial–mathematical link in education: (1) directly training spatial ability to yield mathematical gains and (2) *spatializing* the mathematics curriculum (Newcombe, 2017). This thesis lends support to the latter approach, insofar as the present findings permit such claims. The empirical findings suggest that problem-solving strategies may be a promising but structure-dependent intervening process at the intersection of the spatial–mathematical link. As such, problem-solving strategies may provide a potential means of spatializing the early primary mathematics curriculum, particularly in the context of arithmetic word problem solving instruction. Although the correlational, cross-sectional design of this thesis cannot provide causal evidence

(Bailey, 2017), it identifies a promising target that may serve as a starting point for future research and intervention studies. Moreover, problem-solving strategies represent a relatively low-risk target: even if they are not ultimately viable for leveraging the spatial–mathematical link, they may still improve arithmetic word problem solving and support later mathematical competence.

The discussion section in Study 3 ([Subsection 5.5.3](#)) offers several suggestions for using problem-solving strategies in the instruction of arithmetic word problems in the early primary years. These include (1) providing access to, and explicitly modeling the use of, various tools that scaffold the transition from counting-based strategies to number-fact strategies while supporting the coordination, transformation, and maintenance of representations of a problem’s quantitative structure, and (2) exposing young pupils to a wide range of problems that vary in quantitative structure and in the type of strategies (spatial and analytical strategies) useful for solving them. It also recommends using measurement problems in early primary mathematics classrooms as they bridge “two main areas of school mathematics—geometry and number” (National Council of Teachers of Mathematics, 2000, p. 103).

What was not explicitly discussed in Study 3, however, is how the continuous-quantity problems used in this thesis may serve as useful, measurement-like problems that encourage the representation of quantitative structures, rather than the types of measurement problems that rely more on the application of formulas (e.g., finding area; Vasilyeva et al., 2009). Despite their potential value, a clear disparity exists in the types of word problems young pupils encounter in early primary mathematics education (see [Appendices 4C–4F](#)). In essence, Key Stage 1 National Curriculum Assessments (2018–2025) indicate that young pupils encounter far fewer continuous-quantity problems than discrete-quantity problems.

Continuous-quantity problems may offer several additional advantages for early primary mathematics instruction. First, they foreground the concept of unit as it applies to enumeration, a foundational aspect of early quantitative reasoning that is often overlooked because children may not be aware that counted objects are treated as units (e.g., Sophian, 2008). Second, as demonstrated by the arithmetic approaches pupils used to solve these problems (see [Subsection 2.3.3.4.1](#)), they emphasize the *iteration* of units, a foundational concept in the development of measurement understanding that young pupils find challenging (Nunes et al., 2009b). Third, they may support the transition from whole-number reasoning to rational-number reasoning—a transition that young pupils often find difficult (Barros, 2020)—

by foregrounding the partitioning and iteration of units, alongside the representation of quantities as continuous magnitudes, which form the basis of the rational number system (Lamon, 2020). Fourth, they mimic real-life measurement situations that encourage the use of spatial thinking. Finally, they draw attention away from ‘keyword strategies,’ which remain prevalent in mathematics classrooms despite evidence that such strategies focus narrowly on vocabulary (Peng et al., 2020), and redirect attention toward representing a problem’s quantitative structure. Taken together, these features suggest that continuous-quantity problems may serve as a means of spatializing the mathematics curriculum through a measurement-based approach to early mathematics.

A final note on using problem-solving strategies as a means of spatializing the curriculum concerns the implications of this approach for equity. Here, equity is understood as encompassing (1) identity (i.e., whose ways of thinking mathematics validates; Gutiérrez, 2007) and (2) access (i.e., whether all pupils have equal opportunities to acquire the tools needed to participate in mathematical activities; Esmonde, 2009).

The first implication is that emphasizing problem-solving strategies shifts the focus from arithmetic skills alone to broader forms of mathematical proficiency, including quantitative reasoning and spatial reasoning (National Research Council, 2001). This aligns with calls for more equitable mathematics instruction that values multiple ways of thinking (National Council of Teachers of Mathematics, 2014; Boaler, 2016). From an equity perspective, this shift is significant, as mathematics is often narrowly framed in terms of numerical computation, which may disadvantage learners whose strengths lie in other domains, such as spatial reasoning. Indeed, it is not uncommon for individuals who report strong spatial reasoning to also report difficulties in mathematics, possibly because the mathematics they have encountered overemphasizes arithmetic at the expense of quantitative reasoning.

Reorienting early mathematics toward quantitative reasoning, rather than primarily focusing on counting, may therefore enable young pupils to engage with mathematics as more than just numbers, draw on their intuitive thinking, and develop more positive mathematical identities (Boaler, 2016; Gutiérrez, 2007). Although shifting the prevailing view that equates mathematics with numbers may take time, as reflected in the tendency of even researchers to use mathematical cognition and numerical cognition interchangeably (Davis et al., 2015b), using problem-solving strategies to spatialize the curriculum may represent a step in this direction.

A second, related implication for equity is that broadening what *counts* as mathematics may also broaden who is able to access it. As Casey and Ganley (2021) note, pupils with higher spatial ability and/or richer spatial experiences outside school (e.g., playing with Lego) may have an advantage in mathematical problem solving because “they can more easily draw on spatial representations, spatial imagery, mental manipulation of images and other spatial approaches for problem-solving strategies” (p. 12). Such differences reflect broader opportunity gaps, whereby some pupils have greater access to experiences that support the development of spatial thinking (Levine et al., 2012; National Council of Teachers of Mathematics, 2014). In other words, these pupils may possess a broader repertoire of adaptive strategies involving spatial thinking. By contrast, pupils with lower spatial ability or those who rely primarily on school-taught strategies may have access to a more limited set of approaches, often restricted to analytical procedures or algorithms (Casey et al., 2017; Casey & Fell, 2018).

However, spatial skills are malleable and responsive to instruction (Uttal et al., 2013), suggesting that these disparities are not fixed. Providing explicit instruction in spatially oriented problem-solving strategies from the outset of mathematics learning may therefore serve as an equitable pedagogical approach, supporting pupils who have the capacity to reason spatially but have not yet been taught how to develop or apply this capacity (Gutiérrez, 2007; National Council of Teachers of Mathematics, 2014). This also underscores the importance of exposing pupils to a wide variety of word problems. As this thesis shows, different problem types may differentially afford the use of spatial strategies, thereby shaping pupils’ opportunities to engage in diverse forms of mathematical reasoning. In turn, this may influence whose strengths are recognized and rewarded within mathematics education.

6.5 Limitations and Suggestions for Future Research

While this thesis contributes valuable insights into the relationship between young primary school pupils’ spatial ability and their one-step additive word-problem-solving accuracy and strategies, the findings should be interpreted in light of the thesis’s limitations. With the aim of providing scope for future research, this section highlights the thesis’s overarching limitations in addition to the ones already discussed in each of the respective studies ([Subsections 3.5.3, 4.5.3, and 5.5.4](#)). [Subsection 6.5.1](#) concerns theoretical limitations and [Subsection 6.5.2](#) concerns methodological limitations.

6.5.1 Theoretical Limitations and Related Suggestions for Future Research

First, consistent with the aims of this thesis, the studies focus on the cognitive aspects of arithmetic word problem solving. However, as Mayer (1998) argues, problem solving requires three components: cognition (*skill*), metacognition (*metaskill*), and motivation (*will*). Cognition refers to the problem-solver's domain-specific knowledge, metacognition to the ability to monitor cognitive processes, and motivation to the individual's interest in the task. Indeed, some studies suggest that emotional factors (e.g., confidence and anxiety) partially mediate the spatial–mathematical link (Casey & Ganley, 2021). For example, Sokolowski et al. (2019) suggest that commonly reported sex differences in mathematics anxiety may, in part, stem from anxiety about or avoidance of spatial strategies when solving mathematical tasks.

Although this thesis did not formally assess pupils' motivation, its definition of problem-solving strategies as *goal*-directed procedures for producing numerical answers raises the possibility that pupils' goals may sometimes have been extrinsic (e.g., motivated to complete the task quickly rather than understanding and solving the problem meaningfully). In this sense, problem-solving strategies cannot be considered entirely independent of motivation. Therefore, to better understand the role of problem-solving strategies as a potential intervening process in the spatial–mathematical link, future research should consider the metacognitive and motivational aspects of arithmetic word problem solving.

Second, this thesis focused on one-step additive word problems comprising two quantities and a quantitative relation for reasons highlighted in [Subsection 1.4.1](#). Due to that limited focus, the findings of this thesis cannot attest to how spatial ability may relate to other word problem categories, such as multi-step problems, multiplicative relations, or one-step additive word problems involving different quantitative structures (e.g., three quantities or three quantitative relations; Vergnaud, 1982). These remain important avenues for future research.

For instance, consider the following problem comprising three discrete quantities: 'Ann has 8 books in her left hand. She has 6 books in her right hand. How many books does she have altogether?' Now consider a problem comprising three dynamic quantitative relations: 'Ann played two games of marbles. She lost 3 marbles in the first game and won 5 marbles in the second game. What was the change in the number of marbles?' It may be hypothesized that problems involving three quantitative relations, such as the second problem, are more strongly associated with spatial ability than those involving three quantities, as they place greater

demands on the representation of the problem's quantitative structure. However, such problems may also impose greater verbal or linguistic demands, which could partly account for differences in performance. This hypothesis warrants further investigation in future research, particularly to disentangle the relative contributions of spatial and verbal processing demands.

A similar study examining the relationship between spatial ability and multiplicative word problems may be a worthwhile direction for future research. Multiplicative relations are fundamentally proportional, involving one-to-many correspondences between two variables (Nunes et al., 2016). Given this, spatial scaling—another type of spatial skill that involves “mentally mapping between spaces of different scale” (Gilligan-Lee et al., 2022, p. 2) and therefore relies on proportional reasoning (Newcombe et al., 2018)—may be particularly relevant. Together, these points suggest that spatial scaling and multiplicative reasoning may be linked through their shared reliance on proportional reasoning. Similar to how this thesis systematically examined the quantitative structure of one-step additive word problems, future research could undertake a similar effort, although this would be a meaningful yet challenging endeavor, given that less is known about the categorization of multiplicative word problems (Nunes & Bryant, 2015).

Finally, it may be argued that the continuous quantities used in this thesis are, in fact, relational, as they describe positions that are defined relative to a reference point. However, because these positions were specified with respect to a fixed zero (e.g., the wall), they function more as quantities than as quantitative relations. Future research could build on these competing perspectives by examining continuous quantities across a broader range of contexts (e.g., weight, volume).

6.5.2 Methodological Limitations and Related Suggestions for Future Research

First and foremost, as described throughout the thesis, it uses a correlational, cross-sectional design. While the findings add valuable insights about the association between spatial ability, problem-solving strategies, and problem-solving accuracy, they cannot establish causation (Bailey, 2017). The findings should therefore be interpreted with appropriate caution. However, one key contribution of this thesis is to take the first steps toward identifying problem-solving strategies as a potential target for future longitudinal and experimental studies.

Relatedly, future research could address the aforementioned limitation by implementing a mediation design that involves “random assignment to conditions, theoretically specified mediating mechanisms targeted that differ across interventions, longitudinal

measurement, and good measures of mediating and outcome variables” (MacKinnon, 2008, p. 382). In addition, while this thesis concerned how the domains of spatial ability and word problem solving relate in real-time, longitudinal studies should be employed to investigate how they relate across developmental time (Mix, 2019).

Second, the findings of this thesis should be interpreted in light of limitations related to sample size and statistical power. As noted in relation to the *a priori* power analysis (Subsection 2.2.1), this thesis has a relatively small sample size, which may have resulted in insufficient statistical power to detect small or moderate effects, particularly in mixed-effects models and interaction terms, thereby increasing the likelihood of Type II error.

Third, a further limitation relates to measurement, particularly the reliability and validity of the one-step additive word problems developed for this thesis. While most instruments used in this thesis were either standardized tests or tasks previously validated in research with young children, the one-step additive word problems developed for this thesis raise potential concerns about their reliability and validity. However, to the best of the researcher’s knowledge, no comparable measures exist for young children within this specific conceptual framework, necessitating the development of a new instrument. As detailed in Subsection 2.3.3.5, multiple steps were taken to strengthen its reliability and validity; nevertheless, it cannot be assumed that the four alternative forms are fully equivalent. Future research should therefore aim to replicate these findings with a larger sample, further validate the one-step additive word problems measure, and examine the equivalence of the different forms.

Fourth, as with most educational research, there are limitations concerning generalizability. A non-probabilistic convenience sampling method was used to recruit participants, which may limit generalizability (Cohen et al., 2018; De Vaus, 2001). Participating schools differed from national averages in terms of pupil characteristics (e.g., ethnicity, first language, eligibility for free meals, SEN provision), and complete SEN and EAL data was not always available (Subsection 2.2.3). To mitigate this limitation, testing proceeded only when children were able to meaningfully engage with the tasks. At the same time, including pupils with a wide range of spatial and mathematical abilities captures meaningful variability, which may partially enhance generalizability.

A final set of limitations concerns several methodological trade-offs that may enhance ecological validity at the expense of internal validity. There was minor variation in testing

conditions across schools and classrooms: in some cases, testing occurred in quieter hallway spaces, whereas in others, it took place in classroom-adjacent areas. This variation reflected an ethical priority to ensure that children were assessed in environments where a familiar adult was nearby and that testing did not disrupt essential classroom instruction.

In addition, the administration of the one-step additive word problems involved a deliberate trade-off between strict experimental control and the ethical responsibility to prioritize children's positive experience with mathematics. Although it would have been methodologically preferable to administer each problem under strictly timed conditions (e.g., limiting responses to one minute and then proceeding to the next item), this approach was not adopted to minimize potential negative effects on participants' motivation and engagement. Instead, when a child struggled with a problem for more than one minute, they were reminded of the available supports. While such reminders may have influenced problem-solving strategies and accuracy, these reminders may also have helped sustain engagement and reduce the likelihood that frustration with one item would discourage attempts on later items. This procedural decision therefore represents a limitation, but one made deliberately to prioritize children's positive engagement with mathematics over strict procedural control.

6.6 Concluding Remarks

Spatializing the curriculum may be an effective way of leveraging the spatial-mathematical link in early primary mathematics classrooms without adding spatial reasoning as an additional topic. To effectively harness this link in educational contexts, it is important to first identify the mechanisms underlying it. Despite increased attention to this question in recent years, the spatial-mathematical link remains an underexplored area, partly because both spatial and mathematical abilities are non-unitary. Importantly, arithmetic word problem solving, which is an important early mathematics skill, has received relatively limited attention in this regard.

To contribute to addressing this gap, this thesis took a first step toward investigating the relationship between spatial visualization, visuospatial working memory, and one-step additive word problem solving in young primary pupils. To this end, it examined: first, the association between spatial skills and problem-solving accuracy; second, the association between spatial skills and problem-solving strategies; and third, the association between spatial skills and problem-solving accuracy, mediated through problem-solving strategies.

By proposing a conceptual framework for categorizing one-step additive word problems based on quantitative structure, this thesis offers a more nuanced account of the spatial–mathematical link in early word problem solving. Although distinctions between quantity and quantitative relation (Thompson, 1983), different types of quantities (Davydov, 1982), and different types of quantitative relations (Vergnaud, 1982) have been articulated in prior theoretical work, they have rarely been systematically integrated into research on spatial ability in early mathematics. In addition, by proposing a framework for categorizing problem-solving strategies that integrates arithmetic strategies (e.g., Laski et al., 2013) and representations of quantitative structure (Boonen et al., 2013), this thesis helps bridge the spatial reasoning and quantitative reasoning literatures.

By systematically examining spatial skills that are strongly linked to mathematical thinking, theoretically grounded word-problem structures, and problem-solving strategies that integrate arithmetic skills and quantitative reasoning (Nunes et al., 2012), this thesis demonstrates that the roles of spatial visualization, visuospatial working memory, and problem-solving strategies are structure-dependent rather than uniform across word problem types. Although the correlational, cross-sectional design precludes definitive causal claims (Bailey, 2017), the findings identify problem-solving strategies as a promising, structure-sensitive target for future research and intervention. More broadly, the thesis highlights the importance of considering quantitative structure in understanding how spatial ability contributes to early mathematical reasoning.

References

- ACT (2009). *The path to career success: High school achievement, certainty of career choice, and college readiness make a difference*. ACT.
- Ahmad, A., Tarmizi, R. A., & Nawawi, M. (2010). Visual representations in mathematical word problem solving among form four students in Malacca. *Procedia-Social and Behavioral Sciences*, 8, 356–361. <https://doi.org/10.1016/j.sbspro.2010.12.050>
- Aiken, L. S., & West, S. G. (1991). *Multiple Regression: Testing and Interpreting Interactions*. Sage Publications.
- Aitchison, J. (1982) The statistical analysis of compositional data. *Journal of the Royal Statistical Society. Series B*, 44(2), 139–177. <https://doi.org/10.1111/j.2517-6161.1982.tb01195.x>
- Aleman, A., de Haan, E. H., & Kahn, R. S. (2005). Object versus spatial visual mental imagery in patients with schizophrenia. *Journal of Psychiatry and Neuroscience*, 30(1), 53–56. <https://doi.org/10.1139/jpn.0509>
- Alfons, A., & Schley, D. R. (2025). Robust mediation analysis: What we talk about when we talk about robustness. *Wiley Interdisciplinary Reviews: Computational Statistics*, 17(4), Article e70051. <https://doi.org/10.1002/wics.70051>
- Allen, K., Higgins, S., & Adams, J. (2019). The relationship between visuospatial working memory and mathematical performance in school-aged children: A systematic review. *Educational Psychology Review*, 31(3), 509–531. <https://doi.org/10.1007/s10648-019-09470-8>
- Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, 103(3), 411–423. <https://doi.org/10.1037/0033-2909.103.3.411>
- Andras, F., & Macizo, P. (2026). Problem-size effect in 6 and 12-year-old children: From counting to memory retrieval. *Journal of Experimental Child Psychology*, 262, Article 106394. <https://doi.org/10.1016/j.jecp.2025.106394>
- Ashcraft, M. H., & Fierman, B. A. (1982). Mental addition in third, fourth, and sixth graders. *Journal of Experimental Child Psychology*, 33(2), 216–234. [https://doi.org/10.1016/0022-0965\(82\)90017-0](https://doi.org/10.1016/0022-0965(82)90017-0)
- Atit, K., Power, J.R., Pigott, T., Lee, J., Geer, E. A., Uttal, D. H., Ganley, C. M., & Sorby, S. A. (2022). Examining the relations between spatial skills and mathematical performance: A meta-analysis. *Psychonomic Bulletin & Review* 29, 699–720. <https://doi.org/10.3758/s13423-021-02012-w>

References

- Aubrey, C., Godfrey, R. & Dahl, S. (2006). Early mathematics development and later achievement: Further evidence. *Mathematics Education Research Journal*, 18(1), 27–46. <https://doi.org/10.1007/BF03217428>
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412. <https://doi.org/10.1016/j.jml.2007.12.005>
- Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. A. Bower (Ed.), *The psychology of learning and motivation* (pp. 47–89). Academic Press.
- Bailey, D. H. (2017). Causal inference and the spatial–math link in early childhood. *Monographs of the Society for Research in Child Development*, 82(1), 127–136. <https://doi.org/10.1111/mono.12288>
- Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51(6), 1173–1182. <https://doi.org/10.1037/0022-3514.51.6.1173>
- Baroody, A. J. (1983). The development of procedural knowledge: An alternative explanation for chronometric trends of mental arithmetic. *Developmental Review*, 3(2), 225–230. [https://doi.org/10.1016/0273-2297\(83\)90031-X](https://doi.org/10.1016/0273-2297(83)90031-X)
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Barros, R. (2020). *Factors contributing to children's understanding of fractions: quantitative reasoning and arithmetic* [Doctoral dissertation, University of Oxford]. Oxford University Research Archive.
- Barrouillet, P., Mignon, M., & Thevenot, C. (2008). Strategies in subtraction problem solving in children. *Journal of Experimental Child Psychology*, 99(4), 233–251. <https://doi.org/10.1016/j.jecp.2007.12.001>
- Barrouillet, P., & Thevenot, C. (2013). On the problem-size effect in small additions: Can we really discard any counting-based account? *Cognition*, 128(1), 35–44. <https://doi.org/10.1016/j.cognition.2013.02.018>
- 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>

References

- Bates, K. E., Gilligan-Lee, K., & Farran, E. K. (2021). Reimagining mathematics: The role of mental imagery in explaining mathematical calculation skills in childhood. *Mind, Brain, and Education*, 15(2), 189–198. <https://doi.org/10.1111/mbe.12281>
- Battista, M. T. (1990). Spatial visualization and gender differences in high school geometry. *Journal for Research in Mathematics Education*, 21(1), 47–60. <https://doi.org/10.2307/749456>
- Battista, M. T. (2003). Understanding students' thinking about area and volume measurement. In D. H. Clements & G. Brights (Eds.), *Learning and teaching measurement: 2003 yearbook* (pp. 122–142). National Council of Teachers of Mathematics.
- Bauer, D. J., Preacher, K. J., & Gil, K. M. (2006). Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: New procedures and recommendations. *Psychological Methods*, 11(2), 142–163. <https://doi.org/10.1037/1082-989X.11.2.142>
- Berch, D. B., Krikorian, R., & Huha, E. M. (1998). The Corsi block-tapping task: Methodological and theoretical considerations. *Brain and Cognition*, 38(3), 317–338. <https://doi.org/10.1006/brcg.1998.1039>
- Biesanz, J. C., Falk, C. F., & Savalei, V. (2010). Assessing mediational models: Testing and interval estimation for indirect effects. *Multivariate Behavioral Research*, 45(4), 661–701. <https://doi.org/10.1080/00273171.2010.498292>
- Blazhenkova, O., & Pechenkova, E. (2019). The two eyes of the blind mind: Object vs. spatial aphantasia? *The Russian Journal of Cognitive Science*, 6(4), 51–65. <http://dx.doi.org/10.47010/19.4.5>
- Boaler, J. (2016). *Mathematical mindsets: Unleashing students' potential through creative math, inspiring messages and innovative teaching*. Jossey-Bass/Wiley.
- Bollen, K. A. (1989). *Structural equations with latent variables*. Wiley. <https://doi.org/10.1002/9781118619179>
- Boonen, A. J. H., de Koning, B. B., Jolles, J., & van der Schoot, M. (2016). Word Problem Solving in Contemporary Math Education: A Plea for Reading Comprehension Skills Training. *Frontiers in Psychology*, 7, 1–10. <https://doi.org/10.3389/fpsyg.2016.00191>
- Boonen, A. J. H., van der Schoot, M., van Wesel, F., de Vries, M. H., & Jolles, J. (2013). What underlies successful word problem solving? A path analysis in sixth grade students. *Contemporary Educational Psychology*, 38(3), 271–279. <https://doi.org/10.1016/j.cedpsych.2013.05.001>

References

- Boonen, A. J. H., van Wesel, F., Jolles, J., & van der Schoot, M. (2014). The role of visual representation type, spatial ability, and reading comprehension in word problem solving: An item-level analysis in elementary school children. *International Journal of Educational Research*, 68, 15–26. <https://doi.org/10.1016/j.ijer.2014.08.001>
- British Educational Research Association (2018). *Ethical guidelines for educational research* (4th ed.). London.
- Bryant, P. E., & Kopytynska, H. (1976). Spontaneous measurement by young children. *Nature*, 260(5554), 773. <https://doi.org/10.1038/260773a0>
- Campos, I. S., Almeida, L. S., Ferreira A. I., et al. (2013). Cognitive processes and mathematical performance: A study with children at third grade of basic education. *European Journal of Psychology of Education*, 28(2), 421–436. <https://doi.org/10.1007/s10212-012-0121-x>
- Carpenter, T. P., Fennema, E., Franke, M. L., Levi, L., & Empson, S. B. (1999). *Children's mathematics: Cognitively guided instruction*. Heinemann.
- Carpenter, T. P., & Moser, J. M. (1982). The development of addition and subtraction problem-solving skills. In T. P. Carpenter, J. M. Moser & T. A. Romberg (Eds.), *Addition and subtraction: A cognitive perspective* (pp. 9–24). Erlbaum.
- Carpenter, T. P., & Moser, J. M. (1984). The acquisition of addition and subtraction concepts in grades one through three. *Journal for Research in Mathematics Education*, 15(3), 179–202. <https://doi.org/10.2307/748348>
- Carr, M., & Alexeev, N. (2011). Fluency, accuracy, and gender predict developmental trajectories of arithmetic strategies. *Journal of Educational Psychology*, 103(3), 617–631. <https://doi.org/10.1037/a0023864>
- Carr, M., Steiner, H. H., Kyser, B., & Biddlecomb, B. (2008). A comparison of predictors of early emerging gender differences in mathematics competency. *Learning and Individual Differences*, 18(1), 61–75. <https://doi.org/10.1016/j.lindif.2007.04.005>
- Casey, B. M., Dearing, E., Vasilyeva, M., Ganley, C. M., & Tine, M. (2011). Spatial and numerical predictors of measurement performance: The moderating effects of community income and gender. *Journal of Educational Psychology*, 103(2), 296–311. <https://doi.org/10.1037/a0022516>
- Casey, B. M., & Fell, H. (2018). Spatial reasoning: A critical problem-solving tool in children's mathematics strategy tool-kit. In K. Mix & M. Battista (Eds.), *Visualizing mathematics: The role of spatial reasoning in mathematical thought* (pp. 47–75). Springer Nature Publishing. https://doi.org/10.1007/978-3-319-98767-5_3

References

- Casey, B. M., & Ganley, C. M. (2021). An examination of gender differences in spatial skills and math attitudes in relation to mathematics success: A bio-psycho-social model. *Developmental Review*, 60, Article 100963. <https://doi.org/10.1016/j.dr.2021.100963>
- Casey, B. M., Lombardi, C. M., Pollock, A., Fineman, B., & Pezaris, E. (2017). Girls' spatial skills and arithmetic strategies in first grade as predictors of fifth-grade analytical math reasoning. *Journal of Cognition and Development*, 18(5), 530–555. <https://doi.org/10.1080/15248372.2017.1363044>
- Casey, M. B., Nuttal, R. L., Pezaris, E., & Benbow, C. P. (1995). The influence of spatial ability on gender differences in mathematics college entrance test scores across diverse samples. *Developmental Psychology*, 31(4), 697–705. <https://doi.org/10.1037/0012-1649.31.4.697>
- Casey, B. M., Pezaris, E., Fineman, B., Pollock, A., Demers, L., & Dearing, E. (2015). A longitudinal analysis of early spatial skills compared to arithmetic and verbal skills as predictors of fifth-grade girls' math reasoning. *Learning and Individual Differences*, 40, 90–100. <https://doi.org/10.1016/j.lindif.2015.03.028>
- Central University Research Ethics Committee. (n.d.). *Research ethics and human participants*. <https://researchsupport.admin.ox.ac.uk/files/researchethicswebpdf>
- Central University Research Ethics Committee. (2022). *Non-invasive methods with children as participants in institutional and non-institutional settings*. <https://researchsupport.admin.ox.ac.uk/files/ap25curecapprovedprocedureidrecpdf>
- Chan, W. W. L., & Wong, T. T.-Y. (2019). Visuospatial pathways to mathematical achievement. *Learning and Instruction*, 62, 11–19. <https://doi.org/10.1016/j.learninstruc.2019.03.001>
- Chen, D., & Fritz, M. S. (2021). Comparing alternative corrections for bias in the bias-corrected bootstrap test of mediation. *Evaluation & the Health Professions*, 44(4), 416–427. <https://doi.org/10.1177/01632787211024356>
- Cheng, Y.-L., & Mix, K. S. (2014). Spatial training improves children's mathematics ability. *Journal of Cognition and Development*, 15(1), 2–11. <https://doi.org/10.1080/15248372.2012.725186>
- Cheung, C.-N., Sung, J. Y., & Lourenco, S. F. (2020). Does training mental rotation transfer to gains in mathematical competence? Assessment of an at-home visuospatial intervention. *Psychological Research*, 84(7), 2000–2017. <https://doi.org/10.1007/s00426-019-01202-5>

References

- Ching, B. H.-H., & Nunes, T. (2017). The importance of additive reasoning in children's mathematical achievement: A longitudinal study. *Journal of Educational Psychology*, 109(4), 477–508. <https://doi.org/10.1037/edu0000154>
- Cho, S., Ryali, S., Geary, D. C., & Menon, V. (2011). How does a child solve $7 + 8$? Decoding brain activity patterns associated with counting and retrieval strategies. *Developmental Science*, 14(5), 989–1001. <https://doi.org/10.1111/j.1467-7687.2011.01055.x>
- Cipora, K., Schroeder, P. A., Soltanlou, M., & Nuerk, H.-C. (2018). More space, better mathematics: Is space a powerful tool or a cornerstone for understanding arithmetic? In K. S. Mix & M. T. Battista (Eds.), *Visualizing mathematics: The role of spatial reasoning in mathematical thought* (pp. 77–116). Springer Nature Publishing. https://doi.org/10.1007/978-3-319-98767-5_4
- Claessens, A., & Engel, M. (2013). How important is where you start? Early mathematics knowledge and later school success. *Teachers College Record*, 115(6), 1–29. <https://doi.org/10.1177/016146811311500603>
- Clements, D. H. (1999). Subitizing: What is it? Why teach it? *Teaching Children Mathematics*, 5(7), 400–405. <https://doi.org/10.5951/TCM.5.7.0400>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Lawrence Erlbaum Associates Publishers.
- Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (8th ed.). Routledge.
- Cole, D. A., & Maxwell, S. E. (2003). Testing Mediational Models With Longitudinal Data: Questions and Tips in the Use of Structural Equation Modeling. *Journal of Abnormal Psychology*, 112(4), 558–577. <https://doi.org/10.1037/0021-843X.112.4.558>
- Coolen, I. E. J. I., & Castronovo, J. (2023). How memory counts in mathematical development. *Journal of Cognition*, 6(1), 1–12. <https://doi.org/10.5334/joc.248>
- Corsi, P. M. (1972). *Human memory and the medial temporal region of the brain* (Publication No. NK14430). [Doctoral dissertation, McGill University]. ProQuest Dissertations Publishing. <https://escholarship.mcgill.ca/concern/theses/05741s554>
- Cui, X., & Guo, K. (2022). Supporting mathematics learning: A review of spatial abilities from research to practice. *Current Opinion in Behavioral Sciences*, 46, Article 101176. <https://doi.org/10.1016/j.cobeha.2022.101176>

References

- Cumming, G. (2008). Replication and p intervals: p values predict the future only vaguely, but confidence intervals do much better. *Perspectives on Psychological Science*, 3(4), 286–300. <https://doi.org/10.1111/j.1745-6924.2008.00079.x>
- Cummins, D. D., Kintsch, W., Reusser, K., & Weimer, R. (1988). The role of understanding in solving word problems. *Cognitive Psychology*, 20(4), 405–438. [https://doi.org/10.1016/0010-0285\(88\)90011-4](https://doi.org/10.1016/0010-0285(88)90011-4)
- Curran, P. J., & Bauer, D. J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. *Annual Review of Psychology*, 62, 583–619. <https://doi.org/10.1146/annurev.psych.093008.100356>
- Daroczy, G., Wolska, M., Meurers, W. D., & Nuerk, H.-C. (2015). Word problems: A review of linguistic and numerical factors contributing to their difficulty. *Frontiers in Psychology*, 6, Article 348. <https://doi.org/10.3389/fpsyg.2015.00348>
- Davis, B., Drefs, M., & Francis, K. (2015a). A history and analysis of current curriculum. In B. Davis (Ed.), *Spatial reasoning in the early years: Principles, assertions, and speculations* (pp. 47–63). Routledge. <https://doi.org/10.4324/9781315762371>
- Davis, B., Okamoto, Y., & Whiteley, W. (2015b). Spatializing school mathematics. In B. Davis (Ed.), *Spatial reasoning in the early years: Principles, assertions, and speculations* (pp. 139–150). Routledge. <https://doi.org/10.4324/9781315762371>
- Davydov, V. V. (1982). The psychological characteristics of the formation of elementary mathematical operations in children. In T. P. Carpenter, J. M. Moser & T. A. Romberg (Eds.), *Addition and subtraction: A cognitive perspective* (pp. 224–238). Erlbaum.
- DeBruine, L. M., & Barr, D. J. (2021). Understanding mixed-effects models through data simulation. *Advances in Methods and Practices in Psychological Science*, 4(1), Article 2515245920965119. <https://doi.org/10.1177/2515245920965119>
- Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*, 122(3), 371–396. <https://doi.org/10.1037/0096-3445.122.3.371>
- De Koning, B. B., Boonen, A. J. H., Jongerling, J., van Wesel, F., & van der Schoot, M. (2022). Model method drawing acts as a double-edged sword for solving inconsistent word problems. *Educational Studies in Mathematics*, 111(1), 29–45. <https://doi.org/10.1007/s10649-022-10150-8>
- Delgado, A. R., & Prieto, G. (2004). Cognitive mediators and sex-related differences in mathematics. *Intelligence*, 32(1), 25–32. [https://doi.org/10.1016/S0160-2896\(03\)00061-8](https://doi.org/10.1016/S0160-2896(03)00061-8)

References

- Della Sala, S., Gray, C., Baddeley, A., Allamano, N., & Wilson, L. (1999). Pattern span: a tool for unwinding visuo-spatial memory. *Neuropsychologia*, 37(10), 1189–1199. [https://doi.org/10.1016/S0028-3932\(98\)00159-6](https://doi.org/10.1016/S0028-3932(98)00159-6)
- Department for Education. (2024). *Academic year 2023/24: Schools, pupils and their characteristics*. Department for Education. <https://explore-education-statistics.service.gov.uk/find-statistics/school-pupils-and-their-characteristics>
- Department for Education. (2021). *Statutory framework for the early years foundation stage: Setting the standards for learning, development and care for children from birth to five*. Department for Education.
- Department for Education. (2013). *Mathematics programmes of study: Key stages 1 and 2*. Department for Education.
- De Pauw, S. S. W., Mervielde, I., De Clercq, B. J., De Fruyt, F., Tremmery, S., & Deboutte, D. (2009). Personality symptoms and self-esteem as correlates of psychopathology in child psychiatric patients: Evaluating multiple informant data. *Child Psychiatry and Human Development*, 40(4), 499–515. <https://doi.org/10.1007/s10578-009-0140-2>
- De Vaus, D. A. (2001). *Research design in social research*. SAGE Publications.
- De Vaus, D. A. (2014). *Surveys in social research* (6th ed.). Routledge.
- Dos Santos, C. F. (2022). Re-establishing the distinction between numerosity, numerousness, and number in numerical cognition. *Philosophical Psychology*, 35(8), 1152–1180. <https://doi.org/10.1080/09515089.2022.2029387>
- Duffy, G., Sorby, S., & Bowe, B. (2020). An investigation of the role of spatial ability in representing and solving word problems among engineering students. *Journal of Engineering Education*, 109(3), 424–442. <https://doi.org/10.1002/jee.20349>
- Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L. S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K., & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, 43(6), 1428–1446. <https://doi.org/10.1037/0012-1649.43.6.1428>
- Duque de Blas, G., Gómez-Veiga, I., & García-Madruga, J. A. (2021). Arithmetic word problems revisited: Cognitive processes and academic performance in secondary school. *Education Sciences*, 11(4). Article 155. <https://doi.org/10.3390/educsci11040155>
- Ebel, R. L., & Frisbie, D. A. (1991). *Essentials of educational measurement* (5th ed.). Prentice Hall.

References

- Efron, B., & Tibshirani, R. (1993). *An introduction to the bootstrap*. Chapman & Hall.
<https://doi.org/10.1007/S0033312300003379>
- Ekström, J. (2011). *The Phi-coefficient, the tetrachoric correlation coefficient, and the Pearson-Yule Debate*. UCLA: Department of Statistics, UCLA.
<https://escholarship.org/uc/item/7qp4604r>
- Elff, M. (2026). *mclogit: Multinomial logit models for categorical responses and discrete choices* (Version 0.9.17.1) [R package]. <http://melfff.github.io/mclogit/>
- El-Hamamsy, L., Zapata-Cáceres, M., Marcelino, P., Bruno, B., Dehler Zufferey, J., Martín-Barroso, E., & Román-González, M. (2022). Comparing the psychometric properties of two primary school Computational Thinking (CT) assessments for grades 3 and 4: The Beginners' CT test (BCTt) and the competent CT test (cCTt). *Frontiers in Psychology*, 13, 1082659. <https://doi.org/10.3389/fpsyg.2022.1082659>
- Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, 12(2), 121–138.
<https://doi.org/10.1037/1082-989X.12.2.121>
- Esmonde, I. (2009). Ideas and identities: Supporting equity in cooperative mathematics learning. *Review of Educational Research*, 79(2), 1008–1043.
<https://doi.org/10.3102/0034654309332562>
- Farah, M. J., Hammond, K. M., Levine, D. N., & Calvanio, R. (1988). Visual and spatial mental imagery: Dissociable systems of representation. *Cognitive Psychology*, 20, 439–462.
[https://doi.org/10.1016/0010-0285\(88\)90012-6](https://doi.org/10.1016/0010-0285(88)90012-6)
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>
- Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). SAGE Publications Limited.
- Fischer, M. H., & Shaki, S. (2014). Spatial associations in numerical cognition—From single digits to arithmetic. *Quarterly Journal of Experimental Psychology*, 67(8), 1461–1483.
<https://doi.org/10.1080/17470218.2014.927515>
- Foley, A. E., Vasilyeva, M., & Laski, E. V. (2017). Children's use of decomposition strategies mediates the visuospatial memory and arithmetic accuracy relation. *British Journal of Developmental Psychology*, 35(2), 303–309. <https://doi.org/10.1111/bjdp.12166>
- Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). Sage.
<https://www.john-fox.ca/Companion/>

References

- Frick, A. (2019). Spatial transformation abilities and their relation to later mathematics performance. *Psychological Research*, 83(7), 1465–1484. <https://doi.org/10.1007/s00426-018-1008-5>
- Fuchs, L. S., Compton, D. L., Fuchs, D., Paulsen, K., Bryant, J. D., & Hamlett, C. L. (2005). The prevention, identification, and cognitive determinants of math difficulty. *Journal of Educational Psychology*, 97(3), 493–513. <https://doi.org/10.1037/0022-0663.97.3.493>
- Fuchs, L. S., Fuchs, D., Compton, D. L., Powell, S. R., Seethaler, P. M., Capizzi, A. M., Schatschneider, C., & Fletcher, J. M. (2006). The cognitive correlates of third-grade skill in arithmetic, algorithmic computation, and arithmetic word problems. *Journal of Educational Psychology*, 98(1), 29–43. <https://doi.org/10.1037/0022-0663.98.1.29>
- Fuchs, L. S., Geary, D. C., Compton, D. L., Fuchs, D., Hamlett, C. L., Seethaler, P. M., Bryant, J. D., & Schatschneider, C. (2010). Do different types of school mathematics development depend on different constellations of numerical versus general cognitive abilities? *Developmental Psychology*, 46(6), 1731–1746. <https://doi.org/10.1037/a0020662>
- Fung, W. K., Chung, K. K. H., & Lam, C. B. (2020). Mathematics, executive functioning, and visual–spatial skills in Chinese kindergarten children: Examining the bidirectionality. *Journal of Experimental Child Psychology*, 199, Article 104923. <https://doi.org/10.1016/j.jecp.2020.104923>
- Fung, W., & Swanson, H. L. (2017). Working memory components that predict word problem solving: Is it merely a function of reading, calculation, and fluid intelligence? *Memory & Cognition*, 45(5), 804–823. <https://doi.org/10.3758/s13421-017-0697-0>
- Fuson, K. C. (1982). An analysis of the counting-on solution procedure in addition. In T. P. Carpenter, J. M. Moser & T. A. Romberg (Eds.), *Addition and subtraction: A cognitive perspective* (pp. 67–82). Routledge. <https://doi.org/10.4324/9781003046585>
- Gabler, L., & Ufer, S. (2024). Training flexibility in dealing with additive situations. *Learning and Instruction*, 92, Article 101902. <https://doi.org/10.1016/j.learninstruc.2024.101902>
- Ganley, G. M., & Vasilyeva, M. (2011). Sex differences in the relation between math performance, spatial skills, and attitudes. *Journal of Applied Developmental Psychology*, 32(4), 235–242. <https://doi.org/10.1016/j.appdev.2011.04.001>
- Gauthier, J., Wu, Q. V., & Gooley, T. A. (2020). Cubic splines to model relationships between continuous variables and outcomes: A guide for clinicians. *Bone Marrow Transplantation*, 55, 675–680. <https://doi.org/10.1038/s41409-019-0679-x>

References

- Geary, D. C. (2011). Cognitive predictors of achievement growth in mathematics: A 5-year longitudinal study. *Developmental Psychology*, 47(6), 1539–1552. <https://doi.org/10.1037/a0025510>
- Geer, E. A., Quinn, J. M., & Ganley, C. M. (2019). Relations between spatial skills and math performance in elementary school children: A longitudinal investigation. *Developmental Psychology*, 55(3), 637–652. <https://doi.org/10.1037/dev0000649>
- Gelfand, L. A., Mensinger, J. L., & Tenhave, T. (2009). Mediation analysis: A retrospective snapshot of practice and more recent directions. *Journal of General Psychology*, 136(2), 153–176. <https://doi.org/10.3200/GENP.136.2.153-178>
- Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.
- Gilligan, K. A., Flouri, E., & Farran, E. K. (2017). The contribution of spatial ability to mathematics achievement in middle childhood. *Journal of Experimental Child Psychology*, 163, 107–125. <https://doi.org/10.1016/j.jecp.2017.04.016>
- Gilligan-Lee, K. A., Hawes, Z. C. K., & Mix, K. S. (2022). Spatial thinking as the missing piece in mathematics curricula. *NPJ Science of Learning*, 7(1), Article 28. <https://doi.org/10.1038/s41539-022-00128-9>
- Gilligan, K. A., Thomas, M. S. C., & Farran, E. K. (2020). First demonstration of effective spatial training for near transfer to spatial performance and far transfer to a range of mathematics skills at 8 years. *Developmental Science*, 23(4), Article e12909. <https://doi.org/10.1111/desc.12909>
- Golding, J., Richardson, M., Isaacs, T., Barnes, I., Wilkinson, D., Swensson, C., Maris, R. (2024). *Trends in International Mathematics and Science Study (TIMSS) 2023: National report for England Volume 1*. UK Government.
- Goldsmith, L. T., Hetland, L., Hoyle, C., & Winner, E. (2016). Visual-spatial thinking in geometry and the visual arts. *Psychology of Aesthetics, Creativity, and the Arts*, 10(1), 56–71. <https://doi.org/10.1037/aca0000027>
- Green, C. T., Bunge, S. A., Chiongbian, V. B., Barrow, M., & Ferrer, E. (2017). Fluid reasoning predicts future mathematical performance among children and adolescents. *Journal of Experimental Child Psychology*, 157, 125–143. <https://doi.org/10.1016/j.jecp.2016.12.005>
- Guarnera, M., Pellerone, M., Commodari, E., Valenti, G. D., & Buccheri, S. L. (2019). Mental images and school learning: A longitudinal study on children. *Frontiers in Psychology*, 10, Article 2034. <https://doi.org/10.3389/fpsyg.2019.02034>

References

- Guedj, D. (1997). *Numbers: The universal language*. Harry N. Abrams.
- Gunderson, E. A., Ramirez, G., Beilock, S. L., & Levine, S. C. (2012). The relation between spatial skill and early number knowledge: The role of the linear number line. *Developmental Psychology*, 48(5), 1229–1241. <http://doi.org/10.1037/a0027433>
- Gutiérrez, R. (2007). (Re)defining equity: The importance of a critical perspective. In N. S. Nasir & P. Cobb (Eds.), *Improving access to mathematics: Diversity and equity in the classroom* (pp. 37–50). Teachers College Press.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Pearson.
- Harrell, F. E., Jr. (2015). *Regression modeling strategies: With applications to linear models, logistic and ordinal regression, and survival analysis* (2nd ed.). Springer. <https://doi.org/10.1007/978-3-319-19425-7>
- Hartig, F. (2024). *DHARMA: Residual diagnostics for hierarchical (multi-level / mixed) regression models* (Version 0.4.7) [R package]. <https://github.com/florianhartig/dharma>
- Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.
- Hawes, Z., & Ansari, D. (2020). What explains the relationship between spatial and mathematical skills? A review of evidence from brain and behavior. *Psychonomic Bulletin & Review*, 27(3), 465–482. <https://doi.org/10.3758/s13423-019-01694-7>
- Hawes, Z., Cain, M., Jones, S., Thomson, N., Bailey, C., Seo, J., Caswell, B., & Moss, J. (2019a). Effects of a teacher-designed and teacher-led numerical board game intervention: A randomized controlled study with 4- to 6-year-olds. *Mind, Brain, and Education*, 14(1), 71–80. <https://doi.org/10.1111/mbe.12215>
- Hawes, Z. C. K., Gilligan-Lee, K. A., & Mix, K. S. (2022). Effects of spatial training on mathematics performance: A meta-analysis. *Developmental Psychology*, 58(1), 112–137. <https://doi.org/10.1037/dev0001281>
- Hawes, Z., Moss, J., Caswell, B., Naqvi, S., & MacKinnon, S. (2017). Enhancing children's spatial and numerical skills through a dynamic spatial approach to early geometry instruction: Effects of a 32-week intervention. *Cognition and Instruction*, 35(3), 236–264. <https://doi.org/10.1080/07370008.2017.1323902>

References

- Hawes, Z., Moss, J., Caswell, B., & Poliszczuk, D. (2015). Effects of mental rotation training on children's spatial and mathematics performance: A randomized controlled study. *Trends in Neuroscience and Education*, 4, 60–68. <http://10.1016/j.tine.2015.05.001>
- Hawes, Z., Moss, J., Caswell, B., Seo, J., & Ansari, D. (2019b). Relations between numerical, spatial, and executive function skills and mathematics achievement: A latent-variable approach, *Cognitive Psychology*, 109, 68–90. <https://doi.org/10.1016/j.cogpsych.2018.12.002>
- Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (2nd ed.). Guilford Press.
- Hedeker, D. (2003). A mixed-effects multinomial logistic regression model. *Statistics in Medicine*, 22(9), 1433–1446. <https://doi.org/10.1002/sim.1522>
- Hegarty, M., & Kozhevnikov, M. (1999). Types of visual-spatial representations and mathematical problem solving. *Journal of Educational Psychology*, 91(4), 684–689. <https://doi.org/10.1037/0022-0663.91.4.684>
- Hegarty, M., Mayer, R. E., & Monk, C. A. (1995). Comprehension of arithmetic word problems: A comparison of successful and unsuccessful problem solvers. *Journal of Educational Psychology*, 87(1), 18–32. <https://doi.org/10.1037/0022-0663.87.1.18>
- Hegarty, M., & Waller, D. A. (2005). Individual differences in spatial abilities. In P. Shah & A. Miyake (Eds.), *The Cambridge handbook of visuospatial thinking* (pp. 121–169). Cambridge University Press. <https://doi.org/10.1017/CBO9780511610448.005>
- Hester, R. L., Kinsella, G. J., & Ong, B. (2004). Effect of age on forward and backward span tasks. *Journal of the International Neuropsychological Society*, 10(4), 475–481. <https://doi.org/10.1017/S1355617704104037>
- Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives*, 2(3), 172–177. <https://doi.org/10.1111/j.1750-8606.2008.00061.x>
- Holmes, J., & Adams, J. W. (2006). Working memory and children's mathematical skills: Implications for mathematical development and mathematics curricula. *Educational Psychology*, 26(3), 339–366. <https://doi.org/10.1080/01443410500341056>
- Hoyle, R. H., & Kenny, D. A. (1999). Statistical power and tests of mediation. In R. H. Hoyle (ed.), *Statistical strategies for small sample research* (pp. 195–222). Sage.
- Huang, F. L., & Li, X. (2022). Using cluster-robust standard errors when analyzing group-randomized trials with few clusters. *Behavior Research Methods*, 54(3), 1181–1199. <https://doi.org/10.3758/s13428-021-01627-0>

References

- Hughes, M. (1986). *Children and number: Difficulties in learning mathematics*. Wiley–Blackwell.
- Hume, S. (2019). *Pathspan* (Version 1.2). <https://hume.ca/ix/pathspan/>
- Imai, K., Keele, L., & Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological Methods*, 15(4), 309–334. <https://doi.org/10.1037/a0020761>
- Ingram, J., Stiff, J., Cadwallader, S., Lee, G., & Kayton, H. (2023). *PISA 2022 England national report*. UK Government.
- Isaacs, E. B., & Vargha-Khadem, F. (1989). Differential course of development of spatial and verbal memory span: A normative study. *British Journal of Developmental Psychology*, 7(4), 377–380. <https://doi.org/10.1111/j.2044-835X.1989.tb00814.x>
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59(4), 434–446. <https://doi.org/10.1016/j.jml.2007.11.007>
- Jaffe, J. B., & Bolger, D. J. (2023). Cognitive processes, linguistic factors, and arithmetic word problem success: A review of behavioral studies. *Educational Psychology Review*, 35, Article 105. <https://doi.org/10.1007/s10648-023-09821-6>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning: With applications in R* (2nd ed.). Springer. <https://doi.org/10.1007/978-1-0716-1418-1>
- Jeong, Y., Levine, S. C., & Huttenlocher, J. (2007). The development of proportional reasoning: Effect of continuous versus discrete quantities. *Journal of Cognition and Development*, 8(2), 237–256. <https://doi.org/10.1080/15248370701202471>
- Jitendra, A. K., Griffin, C. C., Haria, P., Leh, J., Adams, A., & Kaduvettoor, A. (2007). A comparison of single and multiple strategy instruction on third-grade students' mathematical problem solving. *Journal of Educational Psychology*, 99(1), 115–127. <https://doi.org/10.1037/0022-0663.99.1.115>
- Jitendra, A. K., & Woodward, J. (2019). The role of visual representations in mathematical word problems. In D. C. Geary, D. B. Berch, & K. M. Koepke (Eds.), *Cognitive foundations for improving mathematical learning* (pp. 269–294). Elsevier Academic Press. <https://doi.org/10.1016/B978-0-12-815952-1.00011-6>
- Johnson, R. B., & Christensen, L. (2019). *Educational research: Quantitative, qualitative, and mixed approaches* (7th ed.). SAGE Publications.

References

- Jordan, N. C., Kaplan, D., Ramineni, C., & Locuniak, M. N. (2008). Development of number combination skill in the early school years: When do fingers help? *Developmental Science*, 11(5), 662–668. <https://doi.org/10.1111/j.1467-7687.2008.00715.x>
- Kahl, T., Segerer, R., Grob, A., & Möhring, W. (2022). Bidirectional associations among executive functions, visual-spatial skills, and mathematical achievement in primary school students: Insights from a longitudinal study. *Cognitive Development*, 62, 1–14. <https://doi.org/10.1016/j.cogdev.2021.101149>
- Keith, T. Z., Fine, J. G., Taub, G. E., Reynolds, M. R., & Kranzler, J. H. (2006). Higher order, multisample, confirmatory factor analysis of the Wechsler Intelligence Scale for Children—Fourth Edition: What does it measure? *School Psychology Review*, 35(1), 108–127.
- Kessels, R. P. C., van den Berg, E., Ruis, C., & Brands, A. M. A. (2008). The backward span of the Corsi Block-Tapping Task and its association with the WAIS–III Digit Span. *Assessment*, 15(4), 426–434. <https://doi.org/10.1177/1073191108315611>
- Kilpatrick, J. (1967). *Analyzing the solution of word problems in mathematics: An exploratory study*. Stanford University.
- Kintsch, W., & Greeno, J. G. (1985). Understanding and solving word arithmetic problems. *Psychological Review*, 92(1), 109–129. <https://doi.org/10.1037/0033-295X.92.1.109>
- Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). The Guilford Press.
- Kozhevnikov, M., & Blazhenkova, O. (2013). Individual differences in object versus spatial imagery: From neural correlates to real-world applications. In S. Lacey & R. Lawson (Eds.), *Multisensory imagery* (pp. 299–318). Springer. https://doi.org/10.1007/978-1-4614-5879-1_16
- Kozhevnikov, M., Hegarty, M., & Mayer, R. E. (2002). Revising the visualizer-verbalizer dimension: Evidence for two types of visualizers. *Cognition and Instruction*, 20(1), 47–77. https://doi.org/10.1207/S1532690XCI2001_3
- Kozhevnikov, M., Kosslyn, S., & Shephard, J. (2005). Spatial versus object visualizers: A new characterization of visual cognitive style. *Memory & Cognition*, 33(4), 710–726. <https://doi.org/10.3758/BF03195337>
- Krawec J. L. (2014). Problem representation and mathematical problem solving of students of varying math ability. *Journal of Learning Disabilities*, 47(2), 103–115. <https://doi.org/10.1177/0022219412436976>

References

- Krenger, M., & Thevenot, C. (2025). The role of children's finger counting history on their addition skills. *Developmental Psychology*. Advance online publication. <https://doi.org/10.1037/dev0002099>
- Krutetskii, V. A. (1976). *The psychology of mathematical abilities in schoolchildren*. University of Chicago Press.
- Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, 28(5), 1–26. <https://doi.org/10.18637/jss.v028.i05>
- Kyttälä, M., Aunio, P., Lepola, J., & Hautamäki, J. (2014). The role of the working memory and language skills in the prediction of word problem solving in 4- to 7-year-old children. *Educational Psychology*, 34(6), 674–696. <https://doi.org/10.1080/01443410.2013.814192>
- Kyttälä, M., & Björn, P. M. (2014). The role of literacy skills in adolescents' mathematics word problem performance: Controlling for visuo-spatial ability and mathematics anxiety. *Learning and Individual Differences*, 29, 59–66. <https://doi.org/10.1016/j.lindif.2013.10.010>
- Kyttälä, M., & Lehto, J. E. (2008). Some factors underlying mathematical performance: The role of visuospatial working memory and non-verbal intelligence. *European Journal of Psychology of Education*, 23(1), 77–94. <https://doi.org/10.1007/BF03173141>
- Lamon, S. J. (2020). *Teaching fractions and ratios for understanding: Essential content knowledge and instructional strategies for teachers* (4th ed.). Routledge. <https://doi.org/10.4324/9781003008057>
- Langhorst, P., Ehler, A., & Fritz, A. (2012). Non-numerical and numerical understanding of the part-whole concept of children aged 4 to 8 in word problems. *Journal für Mathematik-Didaktik*, 33(2), 233–262. <http://doi.org/10.1007/s13138-012-0039-5>
- Laski, E. V., Casey, B. M., Yu, Q., Dulaney, A., Heyman, M., & Dearing, E. (2013). Spatial skills as a predictor of first grade girls' use of higher level arithmetic strategies. *Learning and Individual Differences*, 23, 123–130. <https://doi.org/10.1016/j.lindif.2012.08.001>
- Lê, M. L., Noël, M. P., & Thevenot, C. (2024). The efficacy of manipulatives versus fingers in supporting young children's addition skills. *Journal of Experimental Child Psychology*, 244, Article 105931. <https://doi.org/10.1016/j.jecp.2024.105931>
- LeFevre, J. A., Fast, L., Skwarchuk, S. L., Smith-Chant, B. L., Bisanz, J., Kamawar, D., & Penner-Wilger, M. (2010). Pathways to mathematics: Longitudinal predictors of

References

- performance. *Child Development*, 81(6), 1753–1767. <https://doi.org/10.1111/j.1467-8624.2010.01508.x>
- LeFevre, J.-A., Jimenez Lira, C., Sowinski, C., Cankaya, O., Kamawar, D., & Skwarchuk, S.-L. (2013). Charting the role of the number line in mathematical development. *Frontiers in Psychology*, 4, Article 641. <https://doi.org/10.3389/fpsyg.2013.00641>
- Lemaire, P., & Siegler, R. S. (1995). Four aspects of strategic change: Contributions to children's learning of multiplication. *Journal of Experimental Psychology: General*, 124(1), 83–97. <https://doi.org/10.1037/0096-3445.124.1.83>
- Leung, C. O.-Y., & Wong, T. T.-Y. (2023). How are spatial ability and math related? The mediating roles of numerical magnitude knowledge, understanding of arithmetic operations, and word-problem representation. *Journal of Educational Psychology*, 115(7), 969–984. <https://doi.org/10.1037/edu0000817>
- Levine, S. C., Ratliff, K. R., Huttenlocher, J., & Cannon, J. (2012). Early puzzle play: A predictor of preschoolers' spatial transformation skill. *Developmental Psychology*, 48(2), 530–542. <https://doi.org/10.1037/a0025913>
- Levine, D. N., Warach, J., and Farah, M. (1985). Two visual systems in mental imagery: Dissociation of “what” and “where” in imagery disorders due to bilateral posterior cerebral lesions. *Neurology*, 35, 1010–1018. <http://1212/wnl.35.7.1010>
- Li, Y., & Geary, D. C. (2013). Developmental gains in visuospatial memory predict gains in mathematics achievement. *PLoS One*, 8(7), Article e70160. <https://doi.org/10.1371/journal.pone.0070160>
- Li, Y., & Geary, D. C. (2017). Children's visuospatial memory predicts mathematics achievement through early adolescence. *PLoS One*, 12(2), Article e0172046. <https://doi.org/10.1371/journal.pone.0172046>
- Likhanov, M., Wang, F., Lyu, J., Wang, L., & Zhou, X. (2024). A special contribution from spatial ability to math word problem solving: Evidence from structural equation modelling and network analysis. *Intelligence*, 107, Article 101875. <https://doi.org/10.1016/j.intell.2024.101875>
- Lombardi, C. M., Casey, B. M., Pezaris, E., Shadmehr, M., & Jong, M. (2019). Longitudinal analysis of associations between 3-D mental rotation and mathematics reasoning skills during middle school: Across and within genders. *Journal of Cognition and Development*, 20(4), 487–509. <https://doi.org/10.1080/15248372.2019.1614592>
- Lourenco, S. F., Cheung, C.-N., & Aulet, L. S. (2018). Is visuospatial reasoning related to early mathematical development? A critical review. In A. Henik & W. Fias (Eds.),

References

- Heterogeneity of function in numerical cognition* (pp. 177–210). Academic Press.
<https://doi.org/10.1016/B978-0-12-811529-9.00010-8>
- Lowrie, T., & Diezmann, C. M. (2007). Solving graphics problems: Student performance in junior grades. *Journal of Educational Research*, 100(6), 369–377.
<https://doi.org/10.3200/JOER.100.6.369-378>
- Lowrie, T., Resnick, I., Harris, D., & Logan, T. (2020). In search of the mechanisms that enable transfer from spatial reasoning to mathematics understanding. *Mathematics Education Research Journal*, 32(2), 175–188. <https://doi.org/10.1007/s13394-020-00336-9>
- Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software*, 6(60), Article 3139. <https://doi.org/10.21105/joss.03139>
- Maas, C. J. M., & Hox, J. J. (2005). Sufficient Sample Sizes for Multilevel Modeling. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 1(3), 86–92. <https://doi.org/10.1027/1614-2241.1.3.86>
- MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. Taylor & Francis Group/Lawrence Erlbaum Associates.
- MacKinnon, D. P., Valente, M. J., & Gonzalez, O. (2020). The correspondence between causal and traditional mediation analysis: The link is the mediator by treatment interaction. *Prevention Science*, 21(2), 147–157. <https://doi.org/10.1007/s11121-019-01076-4>
- Marschark, M., Morrison, C., Lukomski, J., Borgna, G., & Convertino, C. (2013). Are deaf students visual learners? *Learning and Individual Differences*, 25(3), 156–162.
<https://doi.org/10.1016/j.lindif.2013.02.006>
- Mayer, R. E. (1985). Mathematical ability. In R. J. Sternberg (Ed.), *Human development: An information-processing approach* (pp. 127–150). W. H. Freeman and Company.
- Mayer, R. E. (1998). Cognitive, metacognitive, and motivational aspects of problem solving. *Instructional Science*, 26, 49–63.
- Mayer, R. E., & Hegarty, M. (1996). The process of understanding mathematical problems. In R. J. Sternberg & T. Ben-Zeev (Eds.), *The nature of mathematical thinking* (pp. 29–53). Lawrence Erlbaum.
- Mayer, R. E., & Wittrock, M. C. (1996). Problem-solving transfer. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 47–62). Macmillan Library Reference USA; Prentice Hall International.

References

- McGartland Rubio, D. (2005). Alpha reliability. In K. Kempf-Leonard (Ed.), *Encyclopedia of social measurement* (pp. 59–63). Elsevier. <https://doi.org/10.1016/B0-12-369398-5/00395-9>
- 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>
- McKenzie, B., Bull, R., & Gray, C. (2003). The effects of phonological and visual-spatial interference on children's arithmetical performance. *Educational and Child Psychology*, 20(3), 93–108. <https://doi.org/10.53841/bpsecp.2003.20.3.93>
- McNeil, N. M., Jordan, N. C., Viegut, A. A., & Ansari, D. (2025). What the science of learning teaches us about arithmetic fluency. *Psychological Science in the Public Interest*, 26(1), 10–57. <https://doi.org/10.1177/15291006241287726>
- McNeish, D., & Stapleton, L. M. (2016). Modeling clustered data with very few clusters. *Multivariate Behavioral Research*, 51(4), 495–518. <https://doi.org/10.1080/00273171.2016.1167008>
- Melby-Lervåg, M., Redick, T. S., & Hulme, C. (2016). Working memory training does not improve performance on measures of intelligence or other measures of “far transfer”: Evidence from a meta-analytic review. *Perspectives on Psychological Science*, 11(4), 512–534. <https://doi.org/10.1177/1745691616635612>
- Menard, S. (2010). *Logistic regression: From introductory to advanced concepts and applications*. Sage Publications. <https://doi.org/10.4135/9781483348964>
- Merlo, J., Chaix, B., Ohlsson, H., Beckman, A., Johnell, K., Hjerpe, P., Råstam, L., & Larsen, K. (2006). A brief conceptual tutorial of multilevel analysis in social epidemiology: Using measures of clustering in multilevel logistic regression to investigate contextual phenomena. *Journal of Epidemiology & Community Health*, 60(4), 290–297. <http://doi.org/10.1136/jech.2004.029454>
- Meyer, M. L., Salimpoor, V. N., Wu, S. S., Geary, D. C., & Menon, V. (2010). Differential contribution of specific working memory components to mathematical achievement in 2nd and 3rd graders. *Learning and Individual Differences*, 20(2), 101–109. <https://doi.org/10.1016/j.lindif.2009.08.004>
- Miller-Cotto, D., & Gordon, R. (2025). Revisiting working memory 50 years after Baddeley and Hitch: A review of field-specific conceptualisations, use and misuse, and paths forward for studying children. *Quarterly Journal of Experimental Psychology*, 78(2), 425–435. <https://doi.org/10.1177/17470218241301701>

References

- Mix, K. S. (2019). Why are spatial skill and mathematics related? *Child Development Perspectives*, 13(2), 121–126. <https://doi.org/10.1111/cdep.123233>
- Mix, K. S., & Cheng, Y.-L. (2012). The relation between space and math: developmental and educational implications. *Advances in Child Development and Behavior*, 42, 197–243. <https://doi.org/10.1016/b978-0-12-394388-0.00006-x>
- Mix, K. S., Huttenlocher, J., & Levine, S. C. (2002). *Quantitative development in infancy and early childhood*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195123005.001.0001>
- Mix, K. S., Levine, S. C., Cheng, Y.-L., Stockton, J. D., & Bower, C. (2021). Effects of spatial training on mathematics in first and sixth grade children. *Journal of Educational Psychology*, 113(2), 304–314. <https://doi.org/10.1037/edu0000494>
- Mix, K. S., Levine, S. C., Cheng, Y. L., Young, C. J., Hambrick, D. Z., & Konstantopoulos, S. (2017). The latent structure of spatial skills and mathematics: A replication of the two-factor model. *Journal of Cognition and Development*, 18(4), 465–492. <https://doi.org/10.1080/15248372.2017.1346658>
- Mix, K. S., Levine, S. C., Cheng, Y.-L., Young, C., Hambrick, D. Z., & Ping, R. (2016). Separate but correlated: The latent structure of space and mathematics across development. *Journal of Experimental Psychology: General*, 145(9), 1206–1227. <http://10.1037/xge0000182>
- 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>
- Möhring, W., Moll, L., & Szubielska, M. (2024). Unpacking associations among children's spatial skills, mathematics, and arithmetic strategies: Decomposition matters. *Psychological Research*, 88(5), 1550–1564. <https://doi.org/10.1007/s00426-024-01952-x>
- Moses, B. E. (1977). *The nature of spatial ability and its relationship to mathematical problem solving*. Indiana University.
- Moss, J., Hawes, Z., Naqvi, S., & Caswell, B. (2016). *Taking shape: Activities to develop geometric and spatial thinking*. Pearson Canada.
- Murnane, R. J., Willett, J. B., Braatz, M. J., & Duhaldeborde, Y. (2001). Do different dimensions of male high school students' skills predict labor market success a decade

References

- later? Evidence from the NLSY. *Economics of Education Review*, 20, 311–320. [https://doi.org/10.1016/S0272-7757\(00\)00056-X](https://doi.org/10.1016/S0272-7757(00)00056-X)
- Múñez, D., Orrantia, J., & Rosales, J. (2013). The effect of external representations on compare word problems: Supporting mental model construction. *Journal of Experimental Education*, 81(3), 337–355. <https://doi.org/10.1080/00220973.2012.715095>
- Nath, S., & Szücs, D. (2014). Construction play and cognitive skills associated with the development and mathematical abilities in 7-year-old children. *Learning and Instruction*, 32, 73–80. <https://doi.org/10.1016/j.learninstruc.2014.01.006>
- National Centre for Excellence in the Teaching of Mathematics. (2014). *Reasoning skills: Developing opportunities and ensuring progression in the development of reasoning skills*. <https://www.ncetm.org.uk/classroom-resources/pm-reasoning-skills/>
- National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. National Council of Teachers of Mathematics.
- National Council of Teachers of Mathematics. (2014). *Principles to actions: Ensuring mathematical success for all*. NCTM.
- National Governors Association Center for Best Practices, & Council of Chief State School Officers. (2010). *Common Core State Standards for mathematics*. https://www.corestandards.org/wp-content/uploads/Math_Standards.pdf
- National Research Council. (2001). *Adding It Up: Helping Children Learn Mathematics*. National Academies Press.
- Nesher, P. (1982). Levels of description in the analysis of addition and subtraction word problems. In T. P. Carpenter, J. M. Moser & T. A. Romberg (Eds.), *Addition and subtraction: A cognitive perspective* (pp. 25–38). Erlbaum.
- Newcombe, N. (2017). *OECD education working papers, no. 161: Harnessing spatial thinking to support STEM learning*. OECD Publishing.
- Newcombe, N. S., Booth, J. L., & Gunderson, E. A. (2019). Spatial skills, reasoning, and mathematics. In *The Cambridge handbook of cognition and education* (pp. 100–123). Cambridge University Press. <https://doi.org/10.1017/9781108235631.006>
- Newcombe, N. S., Levine, S. C., & Mix, K. S. (2015). Thinking about quantity: The intertwined development of spatial and numerical cognition. *WIREs Cognitive Science*, 6(6), 491–505. <https://doi.org/10.1002/wcs.1369>
- Newcombe, N. S., Möhring, W., & Frick, A. (2018). How big is many? Development of spatial and numerical magnitude understanding. In A. Henik & W. Fias (Eds.), *Heterogeneity*

References

- of function in numerical cognition* (pp. 157–176). Elsevier Academic Press.
<https://doi.org/10.1016/B978-0-12-811529-9.00009-1>
- Newcombe, N. S., & Shipley, T. F. (2015). Thinking about spatial thinking: New typology, new assessments. In J. S. Gero (Ed.) *Studying visual and spatial reasoning for design creativity* (pp. 179–192). Springer Netherlands. http://doi.org/10.1007/978-94-017-9297-4_10
- Newsom, J. T. (2023). *SEM with categorical variables*. Psy 523/623 Structural Equation Modeling. https://web.pdx.edu/~newsomj/semclass/ho_categorical.pdf
- Ng, S. F., & Lee, K. (2009). The model method: Singapore children's tool for representing and solving algebraic word problems. *Journal for Research in Mathematics Education*, 40(3), 282–313. <https://doi.org/10.5951/jresmetheduc.40.3.0282>
- Nuerk, H. C., Moeller, K., Klein, E., Willmes, K., & Fischer, M. H. (2015). Extending the mental number line: A review of multi-digit number processing. *Zeitschrift Für Psychologie*, 219(1), 3–22. <https://doi.org/10.1027/2151-2604/a000041>
- Nunes, T., & Bryant, P. (1996). *Children doing mathematics*. Blackwell Publishing.
- Nunes, T., & Bryant, P. (2015). The development of mathematical reasoning. In L. S. Liben, U. Müller, & R. M. Lerner (Eds.), *Handbook of child psychology and developmental science: Cognitive processes* (7th ed., pp. 715–762). John Wiley & Sons, Inc. <https://doi.org/10.1002/9781118963418.childpsy217>
- Nunes, T., Bryant, P., Barros, R., & Sylva, K. (2012). The relative importance of two different mathematical abilities to mathematical achievement. *British Journal of Educational Psychology*, 82(1), 136–156. <https://doi.org/10.1111/j.2044-8279.2011.02033.x>
- Nunes, T., Bryant, P., Evans, D., & Barros, R. (2015). Assessing quantitative reasoning in young children. *Mathematical Thinking and Learning*, 17(2–3), 178–196. <https://doi.org/10.1080/10986065.2015.1016815>
- Nunes, T., Bryant, P., Sylva, K., & Barros, R. (2009a). *Development of maths capabilities and confidence in primary school*. Department for Children, Schools and Families.
- Nunes, T., Bryant, P., & Watson, A. (2009b). *Key understanding in mathematics learning*. Nuffield Foundation.
- Nunes, T., Dorneles, B., Lin, P., & Rathgeb-Schnierer, E. (2016). *Teaching and learning about whole numbers in primary school*. In ICME-13 Topical Surveys. Springer, Cham. https://doi.org/10.1007/978-3-319-45113-8_1
- OECD (2023). *PISA 2022 Assessment and analytical framework*. OECD Publishing. <https://doi.org/10.1787/dfe0bf9c-en>

References

- Oostermeijer, M., Boonen, A. J., & Jolles, J. (2014). The relation between children's constructive play activities, spatial ability, and mathematical word problem-solving performance: a mediation analysis in sixth-grade students. *Frontiers in Psychology*, 5, Article 782. <https://doi.org/10.3389/fpsyg.2014.00782>
- Opfer, J. E., & Siegler, R. S. (2007). Representational change and children's numerical estimation. *Cognitive Psychology*, 55(3), 169–195. <https://doi.org/10.1016/j.cogpsych.2006.09.002>
- Passolunghi, M. C., De Blas, G. D., Carretti, B., Gomez-Veiga, I., Doz, E., & Garcia-Madruga, J. A. (2022). The role of working memory updating, inhibition, fluid intelligence, and reading comprehension in explaining differences between consistent and inconsistent arithmetic word-problem-solving performance. *Journal of Experimental Child Psychology*, 224, 1–17. <https://doi.org/10.1016/j.jecp.2022.105512>
- Peng, P., Lin, X., Ünal, Z. E., Lee, K., Namkung, J., Chow, J., & Sales, A. (2020). Examining the mutual relations between language and mathematics: A meta-analysis. *Psychological Bulletin*, 146(7), 595–634. <https://doi.org/10.1037/bul0000231>
- Peng, P., Wang, T., Wang, C., & Lin, X. (2019). A meta-analysis on the relation between fluid intelligence and reading/mathematics: Effects of tasks, age, and social economics status. *Psychological Bulletin*, 145(2), 189–236. <https://doi.org/10.1037/bul0000182>
- Pina, V., Fuentes, L. J., Castillo, A., & Diamantopoulou, S. (2014). Disentangling the effects of working memory, language, parental education, and non-verbal intelligence on children's mathematical abilities. *Frontiers in Psychology*, 5, Article 415. <https://doi.org/10.3389/fpsyg.2014.00415>
- Poletti, C., Krenger, M., Dupont-Boime, J., & Thevenot, C. (2022). The evolution of finger counting between kindergarten and grade 2. *Children*, 9(2), Article 132. <http://doi.org/10.3390/children9020132>
- Posit Team. (2024). *RStudio: Integrated development environment for R* (Version 2023.12.1.402) [Software]. RStudio, PBC. <https://www.rstudio.com/>
- Powell, M. J. D. (2009). *The BOBYQA algorithm for bound constrained optimization without derivatives*. Report DAMTP 2009/NA06. University of Cambridge.
- Preacher, K. J. (2015). Advances in mediation analysis: A survey and synthesis of new developments. *Annual Reviews of Psychology*, 66, 825–852. <https://doi.org/10.1146/annurev-psych-010814-015258>

References

- Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, 40(3), 879–891. <https://doi.org/10.3758/BRM.40.3.879>
- Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods*, 15(3), 209–233. <https://doi.org/10.1037/a0020141>
- Punch, K. F., & Oancea, A. (2014). *The introduction to research methods in education* (2nd ed.). SAGE Publications.
- R Core Team. (2024). *R: A language and environment for statistical computing* (Version 4.4.0) [Software]. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Sage Publications.
- Reinhold, F., Hofer, S. I., Berkowitz, M., Strohmaier, A., Scheuerer, S., Loch, F., Vogel-Heuser, B., & Reiss, K. (2020). The role of spatial, verbal, numerical, and general reasoning abilities in complex word problem solving for young female and male adults. *Mathematics Education Research Journal*, 32(2), 189–211. <https://doi.org/10.1007/s13394-020-00331-0>
- Resnick, L. B., Nesher, P., Leonard, F., Magone, M., Omanson, S., & Peled, I. (1989). Conceptual bases of arithmetic errors: The case of decimal fractions. *Journal for Research in Mathematics Education*, 20(1), 8–27. <https://doi.org/10.2307/749095>
- Reuhkala, M. (2001). Mathematical skills in ninth-graders: Relationship with visuo-spatial abilities and working memory. *Educational Psychology*, 21(4), 387–399. <https://doi.org/10.1080/01443410120090786>
- Revelle, W. (2026). *psych: Procedures for psychological, psychometric, and personality research* (Version 2.6.3) [Computer software]. Northwestern University. <https://CRAN.R-project.org/package=psych>
- Reynolds, M. R., & Keith, T. Z. (2017). Multi-group and hierarchical confirmatory factor analysis of the Wechsler Intelligence Scale for Children—Fifth edition: What does it measure? *Intelligence*, 62, 31–47. <https://doi.org/10.1016/j.intell.2017.02.005>
- Rhode, T. E., & Thompson, L. A. (2007). Predicting academic achievement with cognitive ability. *Intelligence*, 35(1), 83–92. <https://doi.org/10.1016/j.intell.2006.05.004>
- Rijnhart, J. J. M., Lamp, S. J., Valente, M. J., MacKinnon, D. P., Twisk, J. W. R., & Heymans, M. W. (2021). Mediation analysis methods used in observational research: A scoping

References

- review and recommendations. *BMC Medical Research Methodology*, 21, Article 226. <https://doi.org/10.1186/s12874-021-01426-3>
- Riley, M. S., & Greeno, J. G. (1988). Developmental analysis of understanding language about quantities and of solving problems. *Cognition and Instruction*, 5(1), 49–101. https://doi.org/10.1207/s1532690xci0501_2
- Riley, M. S., Greeno, J. G., & Heller, J. I. (1983). Development of children's problem-solving ability in arithmetic. In H. P. Ginsburg (Ed.), *The development of mathematical thinking* (pp. 153–196). Academic Press.
- Rittle-Johnson, B., Zippert, E. L., & Boice, K. L. (2019). The roles of patterning and spatial skills in early mathematics development. *Early Childhood Research Quarterly*, 46, 166–178. <https://doi.org/10.1016/j.ecresq.2018.03.006>
- Robinson, K. M. (2001). The validity of verbal reports in children's subtraction. *Journal of Educational Psychology*, 93(1), 211–222. <https://doi.org/10.1037/0022-0663.93.1.211>
- Rodán, A., Gimeno, P., Elosúa, M. R., Montoro, P. R., & Contreras, M. J. (2019). Boys and girls gain in spatial, but not in mathematical ability after mental rotation training in primary education. *Learning and Individual Differences*, 70, 1–11. <https://doi.org/10.1016/j.lindif.2019.01.001>
- Rohrer, J. M., Hünermund, P., Arslan, R. C., & Elson, M. (2022). That's a lot to process! Pitfalls of popular path models. *Advances in Methods and Practices in Psychological Science*, 5(2), 1–14. <https://doi.org/10.1177/25152459221095827>
- Rosenthal, D. J. A., & Resnick, L. B. (1974). Children's solution processes in arithmetic word problems. *Journal of Educational Psychology*, 66(6), 817–825. <https://doi.org/10.1037/h0021523>
- Rosseel Y (2012). Llavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36. <https://doi.org/10.18637/jss.v048.i02>
- Sattler, J. M. (2001). *Assessment of children: Cognitive applications* (4th ed.). Jerome M. Sattler, Publisher, Inc.
- Sattler, J. M. & Dumont, R. (2004). *Assessment of children: WISC-IV and WPPSI-III supplement*. Jerome M. Sattler, Publisher, Inc.
- Schenck, K. E., & Nathan, M. J. (2024). Navigating spatial ability for mathematics education: A review and roadmap. *Educational Psychology Review*, 36(3), Article 90. <https://doi.org/10.1007/s10648-024-09935-5>

References

- Schiffman, J., & Laski, E. V. (2018). Materials count: Linear-spatial materials improve young children's addition strategies and accuracy, irregular arrays don't. *PLoS ONE*, 13(12), Article e0208832. <https://doi.org/10.1371/journal.pone.0208832>
- Schneider, M., Beeres, K., Coban, L., Merz, S., Susan Schmidt, S., Stricker, J., & De Smedt, B. (2017). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: a meta-analysis. *Developmental Science*, 20(3), e12372. <https://doi.org/10.1111/desc.12372>
- Schneider, M., Merz, S., Stricker, J., De Smedt, B., Torbeyns, J., Verschaffel, L., & Luwel, K. (2018). Associations of number line estimation with mathematical competence: A meta-analysis. *Child Development*, 89(5), 1467–1484. <https://doi.org/10.1111/cdev.13068>
- Schneider, W. J., & McGrew, K. S. (2012). The Cattell-Horn-Carroll model of intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 99–144). The Guilford Press.
- Schoenfeld, A. H. (1985). *Mathematical problem solving*. Academic Press.
- Schwaighofer, M., Fischer, F., & Bühner, M. (2015). Does working memory training transfer? A meta-analysis including training conditions as moderators. *Educational Psychologist*, 50(2), 138–166. <https://doi.org/10.1080/00461520.2015.1036274>
- Seah, R., & Horne, M. (2019). A learning progression for geometric reasoning. In D. Siemon, T. Barkatsas & R. Seah (Eds.), *Researching and using progressions (trajectories) in mathematics education* (Vol. 3, pp. 157–180). Brill Sense Publishers.
- Sella, F., Berteletti, I., Lucangeli, D., & Zorzi, M. (2017). Preschool children use space, rather than counting, to infer the numerical magnitude of digits: Evidence for a spatial mapping principle. *Cognition*, 158, 56–67. <https://doi.org/10.1016/j.cognition.2016.10.010>
- Seo, J., Stylianides, G. J., & Evangelou, M. (in press). Children's problem-solving strategies as a mediator between spatial visualization and arithmetic word problem-solving accuracy. In *proceedings of the 49th conference of the international group for the Psychology of Mathematics Education* (PME 49). University of Helsinki.
- Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General*, 125(1), 4–27. <https://doi.org/10.1037/0096-3445.125.1.4>

References

- Shea, D. L., Lubinski, D., & Benbow, C. P. (2001). Importance of assessing spatial ability in intellectually talented young adolescents: A 20-year longitudinal study. *Journal of Educational Psychology*, 93(3), 604–614. <https://doi.org/10.1037/0022-0663.93.3.604>
- Siegler, R. S. (1987). The perils of averaging data over strategies: An example from children's addition. *Journal of Experimental Psychology: General*, 116(3), 250–264. <https://doi.org/10.1037/0096-3445.116.3.250>
- Siegler, R. S. (2005). Children's Learning. *American Psychologist*, 60(8), 769–778. <https://doi.org/10.1037/0003-066X.60.8.769>
- Siegler, R. S., & Booth, J. L. (2004). Development of numerical estimation in young children. *Child Development*, 75(2), 428–444. <https://doi.org/10.1111/j.1467-8624.2004.00684.x>
- Siegel, S., & Castellan, N. J., Jr. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). McGraw-Hill Book Company.
- Siegler, R. S., & Jenkins, E. (1989). *How children discover new strategies*. Lawrence Erlbaum Associates, Inc.
- Siegler, R. S., & Ramani, G. B. (2008). Playing linear numerical board games promotes low-income children's numerical development. *Developmental Science*, 11(5), 655–661. <https://doi.org/10.1111/j.1467-7687.2008.00714.x>
- Silverman, S., & Ashkenazi, S. (2022). The unique role of spatial working memory for mathematics performance. *Journal of Numerical Cognition*, 8(1), 226–243. <https://doi.org/10.5964/jnc.7159>
- Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). Sage Publications.
- Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological Methodology*, 13, 290–321. <https://doi.org/10.2307/270723>
- Sokolowski, H. M., Hawes, Z., & Lyons, I. M. (2019). What explains sex differences in math anxiety? A closer look at the role of spatial processing. *Cognition*, 182, 193–212. <https://doi.org/10.1016/j.cognition.2018.10.005>
- Song, M. K., Lin, F. C., Ward, S. E., & Fine, J. P. (2013). Composite variables: when and how. *Nursing Research*, 62(1), 45–49. <https://doi.org/10.1097/NNR.0b013e3182741948>
- Soper, D.S. (2025). *A-priori sample size calculator for structural equation models* [Software]. <https://www.danielsoper.com/statcalc>

References

- Sophian, C. (2008). *The origins of mathematical knowledge in childhood* (1st ed.). Routledge.
<https://doi.org/10.4324/9781315085654>
- Sophian, C., & McCorgray, P. (1994). Part-whole knowledge and early arithmetic problem solving. *Cognition and Instruction*, 12(1), 3–33.
https://doi.org/10.1207/s1532690xcil201_1
- Sorby, S. A., & Panther, G. C. (2020). Is the key to better PISA math scores improving spatial skills? *Mathematics Education Research Journal*, 32(2), 213–233.
<http://10.1007/s13394-020-00328-9>
- Spencer, M., Fuchs, L. S., Geary, D. C., & Fuchs, D. (2022). Connections between mathematics and reading development: Numerical cognition mediates relations between foundational competencies and later academic outcomes. *Journal of Educational Psychology*, 114(2), 273–288. <https://doi.org/10.1037/edu0000670>
- Steffe, L. P., Cobb, P., & Von Glasersfeld, E. (1988). *Construction of arithmetical meanings and strategies*. Springer-Verlag Publishing. <https://doi.org/10.1007/978-1-4612-3844-7>
- Steffe, L. P., Thompson, P. W., & Richards, J. (1982). Children's counting in arithmetical problem solving. In T. P. Carpenter, J. M. Moser & T. A. Romberg (Eds.), *Addition and subtraction: A cognitive perspective* (pp. 9–24). Erlbaum.
<https://doi.org/10.4324/9781003046585>
- Stern, E., & Lehrndorfer, A. (1992). The role of situational context in solving word problems. *Cognitive Development*, 7(2), 259–268. [https://doi.org/10.1016/0885-2014\(92\)90014-I](https://doi.org/10.1016/0885-2014(92)90014-I)
- Strohmaier, A. R., Reinhold, F., Hofer, S., Berkowitz, M., Vogel-Heuser, B., & Reiss, K. (2022). Different complex word problems require different combinations of cognitive skills. *Educational Studies in Mathematics*, 109(1), 89–114.
<https://doi.org/10.1007/s10649-021-10079-4>
- Swanson, H. L., & Fung, W. (2016). Working memory components and problem-solving accuracy: Are there multiple pathways? *Journal of Educational Psychology*, 108(8), 1153–1177. <https://doi.org/10.1037/edu0000116>
- Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Pearson.
- Tam, Y. P., Wong, T. T.-Y., & Chan, W. W. L. (2019). The relation between spatial skills and mathematical abilities: The mediating role of mental number line representation. *Contemporary Educational Psychology*, 56, 14–24.
<https://doi.org/10.1016/j.cedpsych.2018.10.007>

References

- Thevenot, C. (2010). Arithmetic word problem solving: Evidence for the construction of a mental model. *Acta Psychologica*, 133(1), 90–95. <https://doi.org/10.1016/j.actpsy.2009.10.004>
- Thevenot, C., & Barrouillet, P. (2015). Arithmetic word problem solving and mental representations. In R. C. Kadosh & A. Dowker (Eds.), *The Oxford handbook of numerical cognition* (pp. 158–179). Oxford University Press.
- Thevenot, C., & Prado, J. (2026). Patterns of solution times for simple addition problems in 12-year-old children are not compatible with retrieval models: A rebuttal to Andras and Macizo (2025). *Journal of Experimental Child Psychology*, 265, Article 106457. <https://doi.org/10.1016/j.jecp.2026.106457>
- Thompson, P. W. (1993). Quantitative reasoning, complexity, and additive structures. *Educational Studies in Mathematics*, 25(3), 165–208. <https://doi.org/10.1007/BF01273861>
- Tillotson, M. L. (1984). *The effect of instruction in spatial visualization on spatial abilities and mathematical problem solving*. University of Florida.
- Tolar, T. D., Lederberg, A. R., & Fletcher, J. M. (2009). A structural model of algebra achievement: Computational fluency and spatial visualization as mediators of the effect of working memory on algebra achievement. *Educational Psychology*, 29(2), 239–266. <https://doi.org/10.1080/01443410802708903>
- Tzur, R., Johnson, H. L., Norton, A., Davis, A., Wang, X., Ferrara, M., Harrington, C., & Hodkowski, N. M. (2021). Children's spontaneous additive strategy relates to multiplicative reasoning. *Cognition and Instruction*, 39(4), 451–476. <https://doi.org/10.1080/07370008.2021.1896521>
- Uttal, D. H., & Cohen, C. A. (2012). Spatial thinking and STEM education: When, why, and how? *Psychology of Learning and Motivation*, 57, 147–181. <https://doi.org/10.1016/B978-0-12-394293-7.00004-2>
- 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>
- Valcan, D. S., Davis, H. L., Pino-Pasternak, D., & Malpique, A. A. (2020). Executive functioning as a predictor of children's mathematics, reading and writing. *Journal of Applied Developmental Psychology*, 70, Article 101196. <https://doi.org/10.1016/j.appdev.2020.101196>

References

- VanderWeele, T. J. (2016). Mediation analysis: A practitioner's guide. *Annual Reviews of Public Health*, 37, 17–32. <http://doi.org/10.1146/annurev-publhealth-032315-021402>
- Vandierendonck, A., Kemps, E., Fastame, M. C., & Szmalec, A. (2004). Working memory components of the Corsi blocks task. *British Journal of Psychology*, 95(1), 57–79. <https://doi.org/10.1348/000712604322779460>
- Van Garderen, D. (2006). Spatial visualization, visual imagery, and mathematical problem solving of students with varying abilities. *Journal of Learning Disabilities*, 39(6), 496–506. <https://doi.org/10.1177/00222194060390060201>
- Van Garderen, D., & Montague, M. (2003). Visual-spatial representation, mathematical problem solving, and students of varying abilities. *Learning Disabilities Research & Practice*, 18(4), 246–254. <https://doi.org/10.1111/1540-5826.00079>
- Vasilyeva, M., Casey, B. M., Dearing, E., & Ganley, C. M. (2009). Measurement skills in low-income elementary school students: Exploring the nature of gender differences. *Cognition and Instruction*, 27(4), 401–428. <https://doi.org/10.1080/07370000903221809>
- Vasilyeva, M., Laski, E. V., Casey, B., Konstantopoulos, S., Lu, L., Ban, J., Betar, S., Cho, H. Y., & Wang, M. (2025). Recruiting spatial-numerical representations to increase arithmetic fluency in low-income students. *Developmental Psychology*, 61(6), 1152–1165. <https://doi.org/10.1037/dev0001921>
- Vergnaud, G. (1997). The nature of mathematical concepts. In T. Nunes & P. Bryant (Eds.), *Learning and teaching mathematics: An international perspective* (pp. 5–28). Psychology Press/Erlbaum (UK) Taylor & Francis.
- Vergnaud, G. (1982). A classification of cognitive tasks and operations of thought involved in addition and subtraction problems. In T. P. Carpenter, J. M. Moser & T. A. Romberg (Eds.), *Addition and subtraction: A cognitive perspective* (pp. 39–59). Erlbaum.
- Verschaffel, L., & De Corte, E. (1997). Word problems: A vehicle for promoting authentic mathematical understanding and problem solving in the primary school? In T. Nunes & P. Bryant (Eds.), *Learning and teaching mathematics: An international perspective* (pp. 69–97). Psychology Press/Erlbaum (UK) Taylor & Francis.
- Verschaffel, L., Greer, B., & De Corte, E. (2000). *Making sense of word problems*. Swets & Zeitlinger.
- Verschaffel, L., Schukajlow, S., Star, J., & Van Dooren, W. (2020). Word problems in mathematics education: A survey. *ZDM - Mathematics Education*, 52(1), 1–16. <https://doi.org/10.1007/s11858-020-01130-4>

References

- Vessonen, T., Dahlberg, M., Hellstrand, H., Widlund, A., Korhonen, J., Aunio, P., & Laine, A. (2024). Task characteristics associated with mathematical word problem-solving performance among elementary school-aged children: A systematic review and meta-analysis. *Educational Psychology Review*, 36(4), Article 117. <https://doi.org/10.1007/s10648-024-09954-2>
- Vicente, S., Verschaffel, L., Sánchez, R., & Múñez, D. (2022). Arithmetic word problem solving. Analysis of Singaporean and Spanish textbooks. *Educational Studies in Mathematics*, 111, 375–397. <https://doi.org/10.1007/s10649-022-10169-x>
- Vlahović-Štetić, V., Rován, D., & Mendek, Ž. (2004). The role of students' age, problem type and situational context in solving mathematical word problems. *Review of Psychology*, 11(1–2), 25–33.
- Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology*, 101(4), 817–835. <https://doi.org/10.1037/a0016127>
- Wang, L., Cao, C., Zhou, X., & Qi, C. (2022). Spatial abilities associated with open math problem solving. *Applied Cognitive Psychology*, 36(2), 306–317. <https://doi.org/10.1002/acp.3919>
- Wang, A. Y., Fuchs, L. S., & Fuchs, D. (2016). Cognitive and linguistic predictors of mathematical word problems with and without irrelevant information. *Learning and Individual Differences*, 52, 79–87. <https://doi.org/10.1016/j.lindif.2016.10.015>
- Watkins, M. W. (2018). Exploratory factor analysis: A guide to best practice. *Journal of Black Psychology*, 44(3), 219–246. <https://doi.org/10.1177/0095798418771807>
- Watts, T. W., Duncan, G. J., Siegler, R. S., & Davis-Kean, P. E. (2014). What's past is prologue: Relations between early mathematics knowledge and high school achievement. *Educational Researcher*, 43(7), 352–360. <https://doi.org/10.3102/0013189X14553660>
- Wechsler, D. (2011). *Wechsler abbreviated scale of intelligence—Second edition* (WASI–II). Pearson Clinical Assessment UK.
- Wechsler, D. (2005). *Wechsler individual achievement test—Second UK edition* (WIAT–II^{UK}). Pearson Clinical Assessment UK.
- Whitehead, H. L., & Hawes, Z. (2023). Cognitive Foundations of Early Mathematics: Investigating the Unique Contributions of Numerical, Executive Function, and Spatial Skills. *Journal of Intelligence*, 11(12), 1–18. <https://doi.org/10.3390/jintelligence11120221>

References

- Wong, T.-Y. T. (2021). Components of mathematical competence in middle childhood. *Child Development Perspectives*, 15(1), 18–23. <https://doi.org/10.1111/cdep.12394>
- Wong, T. T.-Y., & Ho, C. S.-H. (2017). Component processes in arithmetic word-problem solving and their correlates. *Journal of Educational Psychology*, 109(4), 520–531. <https://doi.org/10.1037/edu0000149>
- Wong, T. T.-Y., & Yip, E. S.-K. (2023). What is the unknown? The ability to identify the semantic role of the unknown from word problems longitudinally predicts mathematical problem solving performance. *Contemporary Educational Psychology*, 73, 1–11. <https://doi.org/10.1016/j.cedpsych.2023.102183>
- World Economic Forum. (2025). *The future of jobs report 2025*. World Economic Forum. https://reports.weforum.org/docs/WEF_Future_of_Jobs_Report_2025.pdf
- Xie, F., Zhang, L., Chen, X., & Xin, Z. (2020). Is spatial ability related to mathematical ability: A meta-analysis. *Educational Psychology Review*, 32(1), 113–155. <https://doi.org/10.1007/s10648-019-09496-y>
- Yang, W., Liu, H., Chen, N., Xu, P., & Lin, X. (2020). Is early spatial skills training effective? A meta-analysis. *Frontiers in Psychology*, 11, Article 1938. <https://doi.org/10.3389/fpsyg.2020.01938>
- Yip, E. S.-K., Wong, T. T.-Y., Cheung, S.-H., & Chan, K. K.-W. (2020). Do children with mathematics learning disability in Hong Kong perceive word problems differently? *Learning and Instruction*, 68, Article 101352. <https://doi.org/10.1016/j.learninstruc.2020.101352>
- Young, C. J., Levine, S. C., & Mix, K. S. (2018). The connection between spatial and mathematical ability across development. *Frontiers in Psychology*, 9, Article 755. <https://doi.org/10.3389/fpsyg.2018.00755>
- Yzerbyt, V., Muller, D., Batailler, C., & Judd, C. M. (2018). New recommendations for testing indirect effects in mediational models: The need to report and test component paths. *Journal of Personality and Social Psychology: Attitudes and Social Cognition*, 115(6), 929–943. <http://doi.org/10.1037/pspa0000132>
- Zhang, X., Koponen, T., Räsänen, P., Aunola, K., Lerkkanen, M. K., & Nurmi, J. E. (2014). Linguistic and spatial skills predict early arithmetic development via counting sequence knowledge. *Child Development*, 85(3), 1091–1107. <https://doi.org/10.1111/cdev.12173>
- Zhang, X., & Lin, D. (2017). Does growth rate in spatial ability matter in predicting early arithmetic competence? *Learning and Instruction*, 49, 232–241. <https://doi.org/10.1016/j.learninstruc.2017.02.003>

References

- Zhang, X., & Lin, D. (2015). Pathways to arithmetic: The role of visual-spatial and language skills in written arithmetic, arithmetic word problems, and nonsymbolic arithmetic. *Contemporary Educational Psychology*, 41, 188–197. <https://doi.org/10.1016/j.cedpsych.2015.01.005>
- Zhao, X., Lynch, J. G., Jr., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *Journal of Consumer Research*, 37(2), 197–206. <https://doi.org/10.1086/651257>
- Zheng, X., Swanson, H. L., & Marcoulides, G. A. (2011). Working memory components as predictors of children's mathematical word problem solving. *Journal of Experimental Child Psychology*, 110(4), 481–498. <https://doi.org/10.1016/j.jecp.2011.06.001>

Appendix 2A: Headteacher Initial Recruitment Email

Dear headteacher,

I hope this email finds you well.

My name is Jisoo Seo, and I'm an Ontario certified elementary school teacher who is now pursuing a DPhil in the Department of Education at the University of Oxford.

I'm writing to ask if you could please kindly consider allowing me to conduct my research study with your Year 2 students during school hours.

Under the supervision of Professor Gabriel Stylianides and Professor Maria Evangelou, I am studying the extent to which young children's spatial ability and representations (e.g., mental images, drawings, etc.) facilitate arithmetic word problem-solving performance. The study has been reviewed and received approval from Oxford's Central University Research Ethics Committee (CUREC) (Approval Reference: [CIA-2223-013]).

As part of this study, your students will complete a one-on-one testing with me. The testing will take about an hour in total (split into two sessions lasting approximately 30 minutes each, spaced at least 30 minutes apart) and involves various spatial and mathematics tasks that are interesting, fun, and appropriately challenging. Furthermore, these activities are expected to enrich your student's mathematics learning experience.

I will provide the school with a summary of the research findings once I analyse the data. I will also provide the participating classroom teachers with collated group assessment data for the students who will take part in the study, which could be useful in the classroom teacher's future planning. In addition, participating classroom teachers will receive research-based mathematics lesson plans developed by a research team at the University of Toronto.

I have attached a detailed information sheet for your consideration. I would be happy to discuss my study further with you and answer any questions you may have.

I sincerely appreciate your time and consideration, and I look forward to hearing from you.

With thanks and best wishes,

Jisoo Seo

Appendix 2B: Headteacher Follow-Up Recruitment Email

Dear Headteacher,

I hope this email finds you well.

I'm writing to follow up on an email I had sent earlier about the possibility of having your Year 2 students participate in my research study, which looks at the extent to which young children's spatial ability and representations (e.g., mental images, drawings, etc.) facilitate arithmetic word problem-solving performance.

As part of this study, your students will complete a one-on-one testing session with me. The testing will take about an hour in total, split into two sessions lasting approximately 30 minutes each, spaced at least 30 minutes apart. It involves various spatial and mathematical tasks that are interesting, fun, and appropriately challenging. Furthermore, these activities are expected to enrich your student's mathematics learning experience, especially as Year 2 students prepare for the National Curriculum Assessments. Finally, I can assure that this study will be minimal work for the school and the teachers.

I will provide the school with a summary of the research findings once I analyse the data. I will also provide the participating classroom teachers with collated group assessment data for the students who will take part in the study, which could be useful in the classroom teacher's future planning.

Furthermore, from a safeguarding perspective, I was a primary school teacher in Toronto before coming to Oxford. I have completed an enhanced DBS check, and I am also a curriculum tutor for the PGCE Mathematics Programme at the University of Oxford. Additionally, this study has been reviewed and approved by Oxford's Central University Research Ethics Committee. Both opt-in and opt-out procedures have been approved, but I am happy for the school to decide what would work best for their parent population.

I have attached a detailed information sheet for your consideration. I would be happy to discuss my study further with you and answer any questions you may have.

Thank you for taking the time to consider my request. I sincerely appreciate your time and consideration, and I look forward to hearing from you. If you require any further information, please do not hesitate to contact me.

With thanks and best wishes,

Jisoo Seo

Appendix 2C: Headteacher information Sheet and Consent Form

UNIVERSITY OF OXFORD

DEPARTMENT OF EDUCATION

15 Norham Gardens, Oxford OX2 6PY

Tel: +44(0)1865 274024 Fax: +44(0)1865 274027

general.enquiries@education.ox.ac.uk www.education.ox.ac.uk

Dr Gabriel Stylianides

email: gabriel.stylianides@education.ox.ac.uk

Jisoo Seo, DPhil student, jisoo.seo@education.ox.ac.uk



[The Role of Spatial Ability and Representation in Early Childhood Arithmetic Word Problem Solving]

HEADTEACHER PARTICIPANT INFORMATION SHEET

Central University Research Ethics Committee (CUREC) Approval Reference: [CIA-2223-013]

1. Why is this research being conducted?

This study aims to investigate (1) to what extent spatial ability predicts arithmetic word problem solving performance, (2) the representations young children produce to solve arithmetic word problems, and (3) whether the problem representations facilitate arithmetic word problem solving performance.

This study's unique foci could lend insights into how to leverage spatial ability within the context of early years mathematics classrooms. The findings of the study could shed light on where spatial ability is (and is not) helpful for arithmetic word problem solving. It is hoped that this study will have implications for efficiently and pragmatically improving mathematics teaching and learning through *the spatialisation of mathematics curriculum* (Newcombe, 2017).

2. Why have I been invited to take part?

You have been invited because this study focuses on students in Year 2 (age range: 6 - 7 years old).

3. Do I have to take part?

No. You can ask questions about the research before deciding whether to take part. If you do agree to take part, you may withdraw yourself from the study, without giving a reason, and without any negative consequences, by advising us of this decision. The deadline by which you can withdraw any information you have contributed to the research is within two months of giving consent. The data collected until the point of withdrawal will be destroyed immediately and will not be shared, used or published in any way.

4. What will happen to me if I take part in the research?

The school participating in this study will:

- choose between an 'opt-in' and an 'opt-out' consent approach for their parent/guardian population
 - [if 'opt-out' consent procedure is chosen,] send targeted email to parents of the involved classes to inform them about the research, including parent information sheet and opt-out form as attachments
- give permission for the researcher, Jisoo Seo, to enter the school property to work one-on-one with the children whose parent/guardian has given consent

UNIVERSITY OF OXFORD

DEPARTMENT OF EDUCATION

15 Norham Gardens, Oxford OX2 6PY

Tel: +44(0)1865 274024 Fax: +44(0)1865 274027

general.enquiries@education.ox.ac.uk www.education.ox.ac.uk

Dr Gabriel Stylianides

email: gabriel.stylianides@education.ox.ac.uk

Jisoo Seo, DPhil student, jisoo.seo@education.ox.ac.uk



The teacher participating in this study will:

- disseminate the parent/guardian information sheet and consent / opt-out form; collect the returned consent / opt-out forms and store them in a safe place until Jisoo comes to collect them
- Provide Jisoo with participating students' birth year and month, and sex
- introduce Jisoo to the class and that she is here to play some maths games with some students
- provide a quiet place within the classroom (or in the hallway) where Jisoo can work one-on-one with each student
- select dates and times for Jisoo to complete two testing sessions, each lasting about 30 minutes

The children participating in this study will:

- complete a series of individual spatial and mathematical assessments over a span of two testing sessions lasting about 30 minutes each (the two testing sessions will be at least 30 minutes apart to minimise the cognitive strain on the students)

If the 'opt-out' consent approach is used, no video or audio recordings of any kind will be made. No photographs will be taken.

If the 'opt-in' consent approach is used, with parent consent and child assent, an audio recording of the child will be made **only** during one part of the session that involves arithmetic word problem solving. This is to have an accurate record of the children's thinking process. No photographs will be taken.

5. Are there any potential risks in taking part?

- To reduce the risk of COVID-19 transmission, the researcher, Jisoo Seo will always wear a mask. Students will be asked to sanitise his/her hands before and after the one-on-one testing and touching an instrument (e.g., iPad, pencil). Jisoo will also sanitise her hands repeatedly and constantly during and in between interactions with the students.
- Students' instruction time may be disrupted. To reduce this, Jisoo will always agree the testing times with the teacher.
- To reduce safety concerns, the one-on-one testing will be done in an open area where another adult familiar to the student is nearby (e.g., a quiet space within the classroom or in the hallway, just outside of the classroom). In addition, Jisoo has completed her safeguarding clearance check (i.e., Enhanced DBS check).
- To reduce the risk of breach of confidentiality, the personal data will be pseudonymised such that you, your school, and your students are not identified in publication. All data collected for this study will be kept in a locked filing cabinet. Any electronic files will be encrypted and housed in password protected computer files.

UNIVERSITY OF OXFORD

DEPARTMENT OF EDUCATION

15 Norham Gardens, Oxford OX2 6PY

Tel: +44(0)1865 274024 Fax: +44(0)1865 274027

general.enquiries@education.ox.ac.uk www.education.ox.ac.uk

Dr Gabriel Stylianides

email: gabriel.stylianides@education.ox.ac.uk

Jisoo Seo, DPhil student, jisoo.seo@education.ox.ac.uk



6. *Are there any benefits in taking part?*

You will be provided with a summary of the research findings. Participating classroom teachers will receive collated group assessment data for the students who will take part in the study. This assessment data could be useful in the classroom teacher's future planning. Furthermore, they will receive research-based mathematics resources that were developed by a research team at the University of Toronto.

In addition, your students will benefit from the one-on-one instruction time with a certified teacher. The activities chosen are interesting, fun and appropriately challenging and are likely to enrich the children's spatial and mathematics learning experience. Particularly, your students will have an opportunity to practice a fundamental early mathematics skill, arithmetic word problem solving. This is a key skill outlined in the national curriculum in England, which dictates that "the principal focus of mathematics teaching in key stage 1 [5- to 7-year-olds] ... involve working with numerals, words and the four operations" (DfE, 2013, p. 5).

Finally, by agreeing to participate you are helping to contribute to an emerging field of research that looks at the role of spatial reasoning in the early years arithmetic word problem solving.

7. *What happens to the data provided?*

The information you provide during the study is the **research data**. Any research data from which you can be identified (e.g., your name, your school's name, your teachers' names, and your students' names, birth year and month, sex and audio recording) is known as **personal data**.

Personal / sensitive data will be anonymised such that you, and your school, teachers and students are not identifiable. Personal / sensitive data will be stored in a securely held locked filing cabinet for three years after publication or public release of the research.

Only the researchers will have access to the research data. Responsible members of the University of Oxford may be given access to data for monitoring and/or audit of the research.

8. *Will the research be published?*

The University of Oxford is committed to the dissemination of its research for the benefit of society and the economy and, in support of this commitment, has established an online archive of research materials. This archive includes digital copies of student theses successfully submitted as part of a University of Oxford postgraduate degree programme. Holding the archive online gives easy access for researchers to the full text of freely available theses, thereby increasing the likely impact and use of that research.

UNIVERSITY OF OXFORD

DEPARTMENT OF EDUCATION

15 Norham Gardens, Oxford OX2 6PY

Tel: +44(0)1865 274024 Fax: +44(0)1865 274027

general.enquiries@education.ox.ac.uk www.education.ox.ac.uk

Dr Gabriel Stylianides

email: gabriel.stylianides@education.ox.ac.uk

Jisoo Seo, DPhil student, jisoo.seo@education.ox.ac.uk



The research will be written up as a student's thesis. On successful submission of the thesis, it may be deposited both in print and online in the University archives to facilitate its use in future research. If so, the thesis will be openly accessible.

The research may also be published in academic publications, conference presentations, publicly available report and website publications.

9. *Who has reviewed this study?*

This study has been reviewed by, and received ethics clearance through, the University of Oxford Central University Research Ethics Committee (Reference number: CIA-2223-013).

10. *Who do I contact if I have a concern about the study or I wish to complain?*

If you have a concern about any aspect of this study, please contact

- Jisoo Seo at jisoo.seo@education.ox.ac.uk or
- Professor Gabriel Stylianides at gabriel.stylianides@education.ox.ac.uk,

and we will do our best to answer your query. We will acknowledge your concern within 10 working days and give you an indication of how it will be dealt with.

If you remain unhappy or wish to make a formal complaint, please contact the Chair of the Research Ethics Committee at the University of Oxford who will seek to resolve the matter as soon as possible:

Chair, **Social Sciences & Humanities Inter-Divisional Research Ethics Committee;**

Email: ethics@socsci.ox.ac.uk;

Address: Research Services, University of Oxford, Wellington Square, Oxford OX1 2JD

11. *Data Protection*

The University of Oxford is the data controller with respect to your personal data, and as such will determine how your personal data is used in the study. The University will process your personal data for the purpose of the research outlined above. Research is a task that is performed in the public interest. Further information about your rights with respect to your personal data is available from <http://www.admin.ox.ac.uk/councilsec/compliance/gdpr/individualrights/>.

12. *Further Information and Contact Details*

If you would like to discuss the research with someone beforehand (or if you have questions afterwards), please contact:

Professor Gabriel Stylianides

Department of Education

University of Oxford, 15 Norham Gardens, Oxford OX2 6PY, United Kingdom

University tel: +44 1865 611 002

University email: gabriel.stylianides@education.ox.ac.uk

UNIVERSITY OF OXFORD

DEPARTMENT OF EDUCATION

15 Norham Gardens, Oxford OX2 6PY

Tel: +44(0)1865 274024 Fax: +44(0)1865 274027

general.enquiries@education.ox.ac.uk www.education.ox.ac.uk

Dr Gabriel Stylianides

email: gabriel.stylianides@education.ox.ac.uk

Jisoo Seo, DPhil student, jisoo.seo@education.ox.ac.uk



HEADTEACHER PARTICIPANT CONSENT FORM (1/2)

Central University Research Ethics Committee (CUREC) Approval Reference: **[CIA-2223-013]**

**[The Role of Spatial Ability and Representation
in Early Childhood Arithmetic Word Problem Solving]**

This study seeks to investigate the relationship between young children's spatial ability and their representations of and performance on arithmetic word problems. This information could lend insights into how to leverage spatial ability within the context of early years mathematics classrooms.

Please initial each box

- 1 I confirm that I have read and understand the information sheet for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily. ☐
- 2 I understand that my participation is voluntary and that I am free to withdraw within two months of giving consent, without giving any reason, and without any adverse consequences or penalty. ☐
- 3 I understand who will have access to personal data provided, how the data will be stored and what will happen to the data at the end of the project. ☐
- 4 I understand that I will not be identifiable from any publications or presentations. ☐
- 5 I understand how to raise a concern or make a complaint. ☐
- 6 I agree to take part. ☐

[Select one]

My school will use the 'opt-in' consent procedure. ☐

Or

My school can and will use the 'opt-out' consent procedure. ☐

UNIVERSITY OF OXFORD

DEPARTMENT OF EDUCATION

15 Norham Gardens, Oxford OX2 6PY

Tel: +44(0)1865 274024 Fax: +44(0)1865 274027

general.enquiries@education.ox.ac.uk www.education.ox.ac.uk

Dr Gabriel Stylianides

email: gabriel.stylianides@education.ox.ac.uk

Jisoo Seo, DPhil student, jisoo.seo@education.ox.ac.uk



HEADTEACHER PARTICIPANT CONSENT FORM (2/2)

Central University Research Ethics Committee (CUREC) Approval Reference: **[CIA-2223-013]**

**[The Role of Spatial Ability and Representation
in Early Childhood Arithmetic Word Problem Solving]**

This study seeks to investigate the relationship between young children's spatial ability and their representations of and performance on arithmetic word problems. This information could lend insights into how to leverage spatial ability within the context of early years mathematics classrooms.

Name of Headteacher

Name of School

dd / mm / yyyy

Date

Signature

Jisoo Seo

Name of person taking consent

dd / mm / yyyy

Date

Signature

Appendix 2D: Teacher Information Sheet and Consent Form

UNIVERSITY OF OXFORD

DEPARTMENT OF EDUCATION

15 Norham Gardens, Oxford OX2 6PY

Tel: +44(0)1865 274024 Fax: +44(0)1865 274027

general.enquiries@education.ox.ac.uk www.education.ox.ac.uk

Dr Gabriel Stylianides

email: gabriel.stylianides@education.ox.ac.uk

Jisoo Seo, DPhil student, jisoo.seo@education.ox.ac.uk



[The Role of Spatial Ability and Representation in Early Childhood Arithmetic Word Problem Solving]

TEACHER PARTICIPANT INFORMATION SHEET

Central University Research Ethics Committee (CUREC) Approval Reference: **[CIA-2223-013]**

1. Why is this research being conducted?

This study aims to investigate (1) to what extent spatial ability predicts arithmetic word problem solving performance, (2) the representations young children produce to solve arithmetic word problems, and (3) whether the problem representations facilitate arithmetic word problem solving performance.

This study's unique foci could lend insights into how to leverage spatial ability within the context of early years mathematics classrooms. The findings of the study could shed light on where spatial ability is (and is not) helpful for arithmetic word problem solving. It is hoped that this study will have implications for efficiently and pragmatically improving mathematics teaching and learning through *the spatialisation of mathematics curriculum* (Newcombe, 2017).

2. Why have I been invited to take part?

You have been invited because this study focuses on students in Year 2 (age range: 6 - 7 years old).

3. Do I have to take part?

No. You can ask questions about the research before deciding whether to take part. If you do agree to take part, you may withdraw yourself from the study, without giving a reason, and without any negative consequences, by advising us of this decision. The deadline by which you can withdraw any information you have contributed to the research is within two months of giving consent. The data collected until the point of withdrawal will be destroyed immediately and will not be shared, used or published in any way.

4. What will happen to me if I take part in the research?

If you are happy to take part in the research, you will be invited to:

- disseminate the parent/guardian information sheet and consent / opt-out form; collect the returned consent / opt-out forms and store them in a safe place until the researcher, Jisoo Seo, comes to collect them
- Provide Jisoo with participating students' birth year and month, and sex
- Introduce Jisoo to the class and that she is here to play some maths games with some students
- provide a quiet place within the classroom (or in the hallway) where Jisoo can work one-on-one with each student
- select dates and times for Jisoo to complete two testing sessions, each lasting about 30 minutes

UNIVERSITY OF OXFORD

DEPARTMENT OF EDUCATION

15 Norham Gardens, Oxford OX2 6PY

Tel: +44(0)1865 274024 Fax: +44(0)1865 274027

general.enquiries@education.ox.ac.uk www.education.ox.ac.uk

Dr Gabriel Stylianides

email: gabriel.stylianides@education.ox.ac.uk

Jisoo Seo, DPhil student, jisoo.seo@education.ox.ac.uk



The children participating in the study will:

- complete a series of individual spatial and mathematical assessments over a span of two testing sessions lasting about 30 minutes each (the two testing sessions will be at least 30 minutes apart to minimise the cognitive strain on the students)

If the 'opt-out' consent approach is used, no video or audio recordings of any kind will be made. No photographs will be taken.

If the 'opt-in' consent approach is used, with parent consent and child assent, an audio recording of the child will be made **only** during one part of the session that involves arithmetic word problem solving. This is to have an accurate record of the children's thinking process. No photographs will be taken.

5. Are there any potential risks in taking part?

- Your instruction time may be disrupted. To reduce this, the researcher, Jisoo Seo, will always agree the testing times with you.
- To reduce the risk of COVID-19 transmission, Jisoo will always wear a mask. Students will be asked to sanitise his/her hands before and after the one-on-one testing and touching an instrument (e.g., iPad, pencil). Jisoo will also sanitise her hands repeatedly and constantly during and in between interactions with the students.
- To reduce safety concerns, the one-on-one testing will be done in an open area where another adult familiar to the student is nearby (e.g., a quiet space within the classroom or in the hallway, just outside of the classroom). In addition, Jisoo has completed her safeguarding clearance check (i.e., Enhanced DBS check).
- To reduce the risk of breach of confidentiality, the personal data will be pseudonymised such that you, your school, and your students are not identified in publication. All data collected for this study will be kept in a locked filing cabinet. Any electronic files will be encrypted and housed in password protected computer files.

6. Are there any benefits in taking part?

You will receive collated group assessment data for the students who will take part in the study. This assessment data could be useful in your future planning. You will also be provided with a summary of the research findings. Furthermore, you will receive research-based mathematics resources that were developed by a research team at the University of Toronto.

In addition, your students will benefit from the one-on-one instruction time with a certified teacher. The activities chosen are interesting, fun and appropriately challenging and are likely to enrich your children's spatial and mathematics learning experience. Particularly, your student will have an opportunity to practice a fundamental early mathematics skill, arithmetic word problem solving. This is a

UNIVERSITY OF OXFORD

DEPARTMENT OF EDUCATION

15 Norham Gardens, Oxford OX2 6PY

Tel: +44(0)1865 274024 Fax: +44(0)1865 274027

general.enquiries@education.ox.ac.uk www.education.ox.ac.uk

Dr Gabriel Stylianides

email: gabriel.stylianides@education.ox.ac.uk

Jisoo Seo, DPhil student, jisoo.seo@education.ox.ac.uk



key skill outlined in the national curriculum in England, which dictates that “the principal focus of mathematics teaching in key stage 1 [5- to 7-year-olds] ... involve working with numerals, words and the four operations” (DfE, 2013, p. 5).

Finally, by agreeing to participate you are helping to contribute to an emerging field of research that looks at the role of spatial reasoning in the early years arithmetic word problem solving.

7. What happens to the data provided?

The information you provide during the study is the **research data**. Any research data from which you can be identified (your name, your school’s name, and your students’ names, birth year and month, sex and audio recording) is known as **personal data**.

Personal / sensitive data will be anonymised such that you, your school and your students are not identifiable. Personal / sensitive data will be stored in a securely held locked filing cabinet for three years after publication or public release of the research.

Only the researchers will have access to the research data. Responsible members of the University of Oxford may be given access to data for monitoring and/or audit of the research.

8. Will the research be published?

The University of Oxford is committed to the dissemination of its research for the benefit of society and the economy and, in support of this commitment, has established an online archive of research materials. This archive includes digital copies of student theses successfully submitted as part of a University of Oxford postgraduate degree programme. Holding the archive online gives easy access for researchers to the full text of freely available theses, thereby increasing the likely impact and use of that research.

The research will be written up as a student’s thesis. On successful submission of the thesis, it may be deposited both in print and online in the University archives to facilitate its use in future research. If so, the thesis will be openly accessible.

The research may also be published in academic publications, conference presentations, publicly available report and website publications.

9. Who has reviewed this study?

This study has been reviewed by, and received ethics clearance through, the University of Oxford Central University Research Ethics Committee (Reference number: **CIA-2223-013**).

UNIVERSITY OF OXFORD

DEPARTMENT OF EDUCATION

15 Norham Gardens, Oxford OX2 6PY

Tel: +44(0)1865 274024 Fax: +44(0)1865 274027

general.enquiries@education.ox.ac.uk www.education.ox.ac.uk

Dr Gabriel Stylianides

email: gabriel.stylianides@education.ox.ac.uk

Jisoo Seo, DPhil student, jisoo.seo@education.ox.ac.uk



10. Who do I contact if I have a concern about the study or I wish to complain?

If you have a concern about any aspect of this study, please contact

- Jisoo Seo at jisoo.seo@education.ox.ac.uk or
- Professor Gabriel Stylianides at gabriel.stylianides@education.ox.ac.uk,

and we will do our best to answer your query. We will acknowledge your concern within 10 working days and give you an indication of how it will be dealt with.

If you remain unhappy or wish to make a formal complaint, please contact the Chair of the Research Ethics Committee at the University of Oxford who will seek to resolve the matter as soon as possible: Chair, **Social Sciences & Humanities Inter-Divisional Research Ethics Committee**;

Email: ethics@socsci.ox.ac.uk;

Address: Research Services, University of Oxford, Wellington Square, Oxford OX1 2JD

11. Data Protection

The University of Oxford is the data controller with respect to your personal data, and as such will determine how your personal data is used in the study. The University will process your personal data for the purpose of the research outlined above. Research is a task that is performed in the public interest. Further information about your rights with respect to your personal data is available from <http://www.admin.ox.ac.uk/councilsec/compliance/gdpr/individualrights/>.

12. Further Information and Contact Details

If you would like to discuss the research with someone beforehand (or if you have questions afterwards), please contact:

Professor Gabriel Stylianides

Department of Education

University of Oxford, 15 Norham Gardens, Oxford OX2 6PY, United Kingdom

University tel: +44 1865 611 002

University email: gabriel.stylianides@education.ox.ac.uk

UNIVERSITY OF OXFORD

DEPARTMENT OF EDUCATION

15 Norham Gardens, Oxford OX2 6PY

Tel: +44(0)1865 274024 Fax: +44(0)1865 274027

general.enquiries@education.ox.ac.uk www.education.ox.ac.uk

Dr Gabriel Stylianides

email: gabriel.stylianides@education.ox.ac.uk

Jisoo Seo, DPhil student, jisoo.seo@education.ox.ac.uk



TEACHER PARTICIPANT CONSENT FORM

Central University Research Ethics Committee (CUREC) Approval Reference: **[CIA-2223-013]**

**[The Role of Spatial Ability and Representation
in Early Childhood Arithmetic Word Problem Solving]**

This study seeks to investigate the relationship between young children's spatial ability and their representations of and performance on arithmetic word problems. This information could lend insights into how to leverage spatial ability within the context of early years mathematics classrooms.

Please initial each box

- | | |
|--|--------------------------|
| 1 I confirm that I have read and understand the information sheet for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily. | ☐ |
| 2 I understand that my participation is voluntary and that I am free to withdraw within two months of giving consent, without giving any reason, and without any adverse consequences or penalty. | ☐ |
| 3 I understand who will have access to personal data provided, how the data will be stored and what will happen to the data at the end of the project. | ☐ |
| 4 I understand that I will not be identifiable from any publications or presentations. | ☐ |
| 5 I understand how to raise a concern or make a complaint. | ☐ |
| 6 I agree to take part. | ☐ |

Name of Teacher

Name of School

dd / mm / yyyy

Date

Signature

Jisoo Seo
Name of person taking consent

dd / mm / yyyy

Date

Signature

Appendix 2E: Parent Information Sheet and Opt-Out Consent Form (Extended Version)

UNIVERSITY OF OXFORD

DEPARTMENT OF EDUCATION

15 Norham Gardens, Oxford OX2 6PY

Tel: +44(0)1865 274024 Fax: +44(0)1865 274027

general.enquiries@education.ox.ac.uk www.education.ox.ac.uk

Dr Gabriel Stylianides

email: gabriel.stylianides@education.ox.ac.uk

Jisoo Seo, DPhil student, jisoo.seo@education.ox.ac.uk



[The Role of Spatial Ability and Representation in Early Childhood Arithmetic Word Problem Solving]

PARENT/GUARDIAN PARTICIPANT INFORMATION SHEET

Central University Research Ethics Committee (CUREC) Approval Reference: [CIA-2223-013]

1. What are we trying to find out?

This research aims to learn more about how young children think when solving arithmetic word problems. It investigates whether children's spatial skills are important for problem solving. This is an important research topic as arithmetic word problem solving is a fundamental early mathematics skill.

2. Why has my child been invited to take part?

Your child has been invited because this study focuses on children in Year 2 and between the ages of 6 and 7.

3. Does my child have to take part?

No. You can ask questions about the research before deciding whether to allow your child to participate. If you do agree to participation, you may withdraw your child from the study, without giving a reason, and without any negative consequences, by advising us of this decision. The deadline by which you can withdraw any information your child has contributed to the research is xxx. The data collected until the point of withdrawal will be destroyed immediately and will not be shared, used or published in any way.

4. What will happen if my child takes part?

Your child will complete a series of spatial and mathematical activities one-on-one with the researcher, Jisoo Seo, who is an Ontario (Canada) certified primary school teacher.

The activities selected are similar to activities that children do in the classroom on a daily basis (e.g., counting, adding, iPad game) and most children generally find the kind of tasks selected interesting, fun and appropriately challenging. As well, most children enjoy and benefit from the one-on-one instruction time with a mathematics teacher.

The activities are split into two sessions lasting about 30 minutes each. The two sessions will be at least 30 minutes apart, so your child does not feel tired. Furthermore, the one-on-one activity will be conducted where an adult familiar to your child will be nearby (e.g., a quiet space within the classroom or in the hallway, just outside of the classroom).

No video or audio recordings of any kind will be made. No photographs will be taken.

UNIVERSITY OF OXFORD

DEPARTMENT OF EDUCATION

15 Norham Gardens, Oxford OX2 6PY

Tel: +44(0)1865 274024 Fax: +44(0)1865 274027

general.enquiries@education.ox.ac.uk www.education.ox.ac.uk

Dr Gabriel Stylianides

email: gabriel.stylianides@education.ox.ac.uk

Jisoo Seo, DPhil student, jisoo.seo@education.ox.ac.uk



5. *Are there any potential risks in taking part?*

There is no serious risk to your child in taking part. You may wish to know that:

- to reduce the risk of COVID-19 transmission, the researcher, Jisoo Seo, will always wear a mask. Your child will be asked to sanitise his/her hands before and after the one-on-one testing and touching an instrument (e.g., iPad, pencil). The researcher will also sanitise her hands repeatedly and constantly during and in between interactions with children.
- Jisoo will always agree the testing times with the teacher so that the interruptions to your child's classroom routines and instruction/play time are minimised.
- to reduce safety concerns, the one-on-one testing will be done in an open area where another adult familiar to your child is nearby (e.g., a quiet space within the classroom or in the hallway, just outside of the classroom).
- efforts will be made to ensure that your child is comfortable and enjoying him/herself. However, if at any point during the assessment process your child feels uncomfortable, the assessment will be stopped immediately, without any negative consequences.
- to reduce the risk of breach of confidentiality, the personal data will be pseudonymised such that you, your child, and your child's school and teacher are not identified in publication. Your child's name will not be attached to any assessment data that leave the classroom. All data collected for this study will be kept in a locked filing cabinet.
- your child's individual assessment scores will be kept confidential. The classroom teacher will receive data on the average performance of all students in his/her class who participate in the study. These scores will be held in confidence by him/her in the same way that he/she keeps other data on your child confidential. The assessment data will not be used for evaluative purposes and will not become part of your child's official student record. This information may be used by the classroom teacher to help him/her plan suitable activities for the class and your child.

6. *Are there any benefits in taking part?*

Your child will benefit from the one-on-one instruction time with a certified teacher. The activities chosen are interesting, fun and appropriately challenging and are likely to enrich your child's spatial and mathematics learning experience. As well, most children enjoy and benefit from the one-on-one instruction time with a mathematics teacher.

In addition, your child will have an opportunity to practice a fundamental early mathematics skill, arithmetic word problem solving. This is a key skill outlined in the national curriculum in England, which dictates that "the principal focus of mathematics teaching in key stage 1 [5- to 7-year-olds] ... involve working with numerals, words and the four operations" (DfE, 2013, p. 5).

UNIVERSITY OF OXFORD

DEPARTMENT OF EDUCATION

15 Norham Gardens, Oxford OX2 6PY

Tel: +44(0)1865 274024 Fax: +44(0)1865 274027

general.enquiries@education.ox.ac.uk www.education.ox.ac.uk

Dr Gabriel Stylianides

email: gabriel.stylianides@education.ox.ac.uk

Jisoo Seo, DPhil student, jisoo.seo@education.ox.ac.uk



Furthermore, by agreeing to participate you are helping contribute to an emerging field of research that looks at the role of spatial reasoning in the early years arithmetic word problem solving.

7. What happens to the data provided?

The information you provide during the study is the **research data**. Any research data from which you can be identified (your name, your child's name, birth year and month and sex, and your child's school and teacher) is known as **personal data**.

Personal / sensitive data will be anonymised such that you, your child, and your child's school and teacher are not identifiable. Personal / sensitive data will be stored in a securely held locked filing cabinet for three years after publication or public release of the research.

Only the researchers will have access to the research data. Responsible members of the University of Oxford may be given access to data for monitoring and/or audit of the research.

Opt-out forms will be retained by the school for the duration of the study, and for as long as the school determines appropriate after research activities have concluded at the school.

8. Will the research be published?

The University of Oxford is committed to the dissemination of its research for the benefit of society and the economy and, in support of this commitment, has established an online archive of research materials. This archive includes digital copies of student theses successfully submitted as part of a University of Oxford postgraduate degree programme. Holding the archive online gives easy access for researchers to the full text of freely available theses, thereby increasing the likely impact and use of that research.

The research will be written up as a student's thesis. On successful submission of the thesis, it may be deposited both in print and online in the University archives to facilitate its use in future research. If so, the thesis will be openly accessible.

The research may also be published in academic publications, conference presentations, publicly available report and website publications.

9. Who has reviewed this study?

This study has been reviewed by, and received ethics clearance through, the University of Oxford Central University Research Ethics Committee (Reference number: CIA-2223-013).

UNIVERSITY OF OXFORD

DEPARTMENT OF EDUCATION

15 Norham Gardens, Oxford OX2 6PY

Tel: +44(0)1865 274024 Fax: +44(0)1865 274027

general.enquiries@education.ox.ac.uk www.education.ox.ac.uk

Dr Gabriel Stylianides

email: gabriel.stylianides@education.ox.ac.uk

Jisoo Seo, DPhil student, jisoo.seo@education.ox.ac.uk



10. Who do I contact if I have a concern about the study or I wish to complain?

If you have a concern about any aspect of this study, please contact

- Jisoo Seo at jisoo.seo@education.ox.ac.uk or
- Professor Gabriel Stylianides at gabriel.stylianides@education.ox.ac.uk,

and we will do our best to answer your query. We will acknowledge your concern within 10 working days and give you an indication of how it will be dealt with.

If you remain unhappy or wish to make a formal complaint, please contact the Chair of the Research Ethics Committee at the University of Oxford who will seek to resolve the matter as soon as possible:

Chair, **Social Sciences & Humanities Inter-Divisional Research Ethics Committee**;

Email: ethics@socsci.ox.ac.uk;

Address: Research Services, University of Oxford, Wellington Square, Oxford OX1 2JD

11. Data Protection

The University of Oxford is the data controller with respect to your personal data, and as such will determine how your personal data is used in the study. The University will process your personal data for the purpose of the research outlined above. Research is a task that is performed in the public interest. Further information about your rights with respect to your personal data is available from <http://www.admin.ox.ac.uk/councilsec/compliance/gdpr/individualrights/>.

12. Your Child's Assent

To help your child decide whether he/she would like to participate in this study, the researcher, Jisoo Seo, will read the following assent script to your child before each activity session.

I am a teacher and I want to know more about how children like you think about maths. If you decide to help, I will ask you to imagine making some figures and to answer some questions about numbers. Some activities will be fun and easy, and some might be a little bit hard. It is okay if you do not know the answers. I am mostly interested in your thinking.

I would like to use your ideas in a paper that I am writing. When I write about your ideas, I will not use your name, so no one will know it is you. Instead, I will think of a made-up name to give you in the paper.

If at any time you don't want to do any more activities with me, you can just say you want to stop, and we will. No one will be cross with you if you want to stop. It will be okay if you want to stop at any point.

Thank you.

UNIVERSITY OF OXFORD

DEPARTMENT OF EDUCATION

15 Norham Gardens, Oxford OX2 6PY

Tel: +44(0)1865 274024 Fax: +44(0)1865 274027

general.enquiries@education.ox.ac.uk www.education.ox.ac.uk

Dr Gabriel Stylianides

email: gabriel.stylianides@education.ox.ac.uk

Jisoo Seo, DPhil student, jisoo.seo@education.ox.ac.uk



13. What should I do next?

Please fill in the enclosed form and return it to your child's class teacher by xxx if you would not like your child to take part in this study. Please remember that you may withdraw your child by xxx, without any penalty and without giving a reason, by notifying the researcher.

14. Further Information and Contact Details

If you would like to discuss the research with someone beforehand (or if you have questions afterwards), please contact:

Professor Gabriel Stylianides

Department of Education

University of Oxford, 15 Norham Gardens, Oxford OX2 6PY, United Kingdom

University tel: +44 1865 611 002

University email: gabriel.stylianides@education.ox.ac.uk

UNIVERSITY OF OXFORD

DEPARTMENT OF EDUCATION

15 Norham Gardens, Oxford OX2 6PY

Tel: +44(0)1865 274024 Fax: +44(0)1865 274027

general.enquiries@education.ox.ac.uk www.education.ox.ac.uk

Dr Gabriel Stylianides

email: gabriel.stylianides@education.ox.ac.uk

Jisoo Seo, DPhil student, jisoo.seo@education.ox.ac.uk



PARENT/GUARDIAN OPT-OUT FORM

Central University Research Ethics Committee (CUREC) Approval Reference: **[CIA-2223-013]**

**[The Role of Spatial Ability and Representation
in Early Childhood Arithmetic Word Problem Solving]**

If you **DO NOT** want your child to participate in the above-named research study, please fill out the form below and return it to the school by [xxx].

If we do not receive an opt-out form from you by this date, your child may be invited to take part in this study, as described in the accompanying information sheet.

I, the undersigned, hereby DO NOT give permission for my child to take part in the study titled, The Role of Spatial Ability and Representation in Early Childhood Arithmetic Word Problem Solving.

Name of child: _____

Name of parent/guardian: _____

Signature: _____ Date: _____

Name of researcher: _____

Appendix 2F: Parent Information Sheet and Opt-Out Consent Form (Simplified Version)

UNIVERSITY OF OXFORD

DEPARTMENT OF EDUCATION

15 Norham Gardens, Oxford OX2 6PY
Tel: +44(0)1865 274024 Fax: +44(0)1865 274027
general.enquiries@education.ox.ac.uk

Dr Gabriel Stylianides

gabriel.stylianides@education.ox.ac.uk

Jisoo Seo, DPhil student

jisoo.seo@education.ox.ac.uk



Dear Parent/Carer,

An opportunity has arisen for the year 2 children to take part in a maths research project with Jisoo Seo, an Oxford University PhD student. The project will be looking at the relationship between children's spatial thinking and their ability to solve maths problems.

Your child will work on-on-one with Jisoo for one hour during the school day, with a break in the middle. The activities involve doing puzzles and addition and subtraction, which will be interesting, fun and appropriately challenging for your child.

The data will be anonymized so that your child, and your child's school and teacher are not identifiable in any published materials. No recordings or photos of any kind will be made.

If you DO NOT want your child to participate in this study, please fill out the below slip and return to Reception or your child's teacher by xxx. If we do not receive an opt-out form from you by xxx, your child may be invited to take part in this study.

If you do agree to participation, you may withdraw your child from the study, without giving a reason, and without any negative consequences, by advising us of this decision. The deadline by which you can withdraw any information your child has contributed to the research is xxx.

If you have any questions or concerns about the study, please email Jisoo at jisoo.seo@education.ox.ac.uk.

With thanks and best wishes,

Jisoo Seo

☐

I Do NOT give permission for my child to take part in the maths project.

Signed _____ Name of Child _____

Appendix 2G: Parent Information Sheet and Opt-In Consent Form (Extended Version)

UNIVERSITY OF OXFORD

DEPARTMENT OF EDUCATION

15 Norham Gardens, Oxford OX2 6PY

Tel: +44(0)1865 274024 Fax: +44(0)1865 274027

general.enquiries@education.ox.ac.uk www.education.ox.ac.uk

Dr Gabriel Stylianides

email: gabriel.stylianides@education.ox.ac.uk

Jisoo Seo, DPhil student, jisoo.seo@education.ox.ac.uk



[The Role of Spatial Ability and Representation in Early Childhood Arithmetic Word Problem Solving]

PARENT/GUARDIAN PARTICIPANT INFORMATION SHEET

Central University Research Ethics Committee (CUREC) Approval Reference: [CIA-2223-013]

1. What are we trying to find out?

This research aims to learn more about how young children think when solving arithmetic word problems. It investigates whether children's spatial skills are important for problem solving. This is an important research topic as arithmetic word problem solving is a fundamental early mathematics skill.

2. Why has my child been invited to take part?

Your child has been invited because this study focuses on children in Year 2 and between the ages of 6 and 7.

3. Does my child have to take part?

No. You can ask questions about the research before deciding whether to allow your child to participate. If you do agree to participation, you may withdraw your child from the study, without giving a reason, and without any negative consequences, by advising us of this decision. The deadline by which you can withdraw any information your child has contributed to the research is within two months of giving consent. The data collected until the point of withdrawal will be destroyed immediately and will not be shared, used or published in any way.

4. What will happen if my child takes part?

Your child will complete a series of spatial and mathematical activities one-on-one with the researcher, Jisoo Seo, who is an Ontario (Canada) certified primary school teacher.

The activities selected are similar to activities that children do in the classroom on a daily basis (e.g., counting, adding, iPad game) and most children generally find the kind of tasks selected interesting, fun and appropriately challenging. As well, most children enjoy and benefit from the one-on-one instruction time with a mathematics teacher.

The activities are split into two sessions lasting about 30 minutes each. The two sessions will be at least 30 minutes apart, so your child does not feel tired. Furthermore, the one-on-one activity will be conducted where an adult familiar to your child will be nearby (e.g., a quiet space within the classroom or in the hallway, just outside of the classroom).

With your consent and your child's assent, an audio recording of your child will be made **only** during one part of the session that involves arithmetic word problem solving. This is to have an accurate record of your child's thinking process.

UNIVERSITY OF OXFORD

DEPARTMENT OF EDUCATION

15 Norham Gardens, Oxford OX2 6PY

Tel: +44(0)1865 274024 Fax: +44(0)1865 274027

general.enquiries@education.ox.ac.uk www.education.ox.ac.uk

Dr Gabriel Stylianides

email: gabriel.stylianides@education.ox.ac.uk

Jisoo Seo, DPhil student, jisoo.seo@education.ox.ac.uk



5. Are there any potential risks in taking part?

There is no serious risk to your child in taking part. You may wish to know that:

- to reduce the risk of COVID-19 transmission, the researcher, Jisoo Seo, will always wear a mask. Your child will be asked to sanitise his/her hands before and after the one-on-one testing and touching an instrument (e.g., iPad, pencil). The researcher will also sanitise her hands repeatedly and constantly during and in between interactions with children.
- Jisoo will always agree the testing times with the teacher so that the interruptions to your child's classroom routines and instruction/play time are minimised.
- to reduce safety concerns, the one-on-one testing will be done in an open area where another adult familiar to your child is nearby (e.g., a quiet space within the classroom or in the hallway, just outside of the classroom).
- efforts will be made to ensure that your child is comfortable and enjoying him/herself. However, if at any point during the assessment process your child feels uncomfortable, the assessment will be stopped immediately, without any negative consequences.
- to reduce the risk of breach of confidentiality, the personal data will be pseudonymised such that you, your child, and your child's school and teacher are not identified in publication. Your child's name will not be attached to any assessment data that leave the classroom. All data collected for this study will be kept in a locked filing cabinet.
- your child's individual assessment scores will be kept confidential. The classroom teacher will receive data on the average performance of all students in his/her class who participate in the study. These scores will be held in confidence by him/her in the same way that he/she keeps other data on your child confidential. The assessment data will not be used for evaluative purposes and will not become part of your child's official student record. This information may be used by the classroom teacher to help him/her plan suitable activities for the class and your child.

6. Are there any benefits in taking part?

Your child will benefit from the one-on-one instruction time with a certified teacher. The activities chosen are interesting, fun and appropriately challenging and are likely to enrich your child's spatial and mathematics learning experience. As well, most children enjoy and benefit from the one-on-one instruction time with a mathematics teacher.

In addition, your child will have an opportunity to practice a fundamental early mathematics skill, arithmetic word problem solving. This is a key skill outlined in the national curriculum in England, which dictates that "the principal focus of mathematics teaching in key stage 1 [5- to 7-year-olds] ... involve working with numerals, words and the four operations" (DfE, 2013, p. 5).

UNIVERSITY OF OXFORD

DEPARTMENT OF EDUCATION

15 Norham Gardens, Oxford OX2 6PY

Tel: +44(0)1865 274024 Fax: +44(0)1865 274027

general.enquiries@education.ox.ac.uk www.education.ox.ac.uk

Dr Gabriel Stylianides

email: gabriel.stylianides@education.ox.ac.uk

Jisoo Seo, DPhil student, jisoo.seo@education.ox.ac.uk



Furthermore, by agreeing to participate you are helping contribute to an emerging field of research that looks at the role of spatial reasoning in the early years arithmetic word problem solving.

7. What happens to the data provided?

The information you provide during the study is the **research data**. Any research data from which you can be identified (your name, your child's name, date of birth, sex and audio recording, and your child's school and teacher) is known as **personal data**.

Personal / sensitive data will be anonymised such that you, your child, and your child's school and teacher are not identifiable. Personal / sensitive data will be stored in a securely held locked filing cabinet for three years after publication or public release of the research.

Only the researchers will have access to the research data. Responsible members of the University of Oxford may be given access to data for monitoring and/or audit of the research.

8. Will the research be published?

The University of Oxford is committed to the dissemination of its research for the benefit of society and the economy and, in support of this commitment, has established an online archive of research materials. This archive includes digital copies of student theses successfully submitted as part of a University of Oxford postgraduate degree programme. Holding the archive online gives easy access for researchers to the full text of freely available theses, thereby increasing the likely impact and use of that research.

The research will be written up as a student's thesis. On successful submission of the thesis, it may be deposited both in print and online in the University archives to facilitate its use in future research. If so, the thesis will be openly accessible.

The research may also be published in academic publications, conference presentations, publicly available report and website publications.

9. Who has reviewed this study?

This study has been reviewed by, and received ethics clearance through, the University of Oxford Central University Research Ethics Committee (Reference number: CIA-2223-013).

10. Who do I contact if I have a concern about the study or I wish to complain?

If you have a concern about any aspect of this study, please contact

- Jisoo Seo at jisoo.seo@education.ox.ac.uk or
- Professor Gabriel Stylianides at gabriel.stylianides@education.ox.ac.uk,

and we will do our best to answer your query. We will acknowledge your concern within 10 working days and give you an indication of how it will be dealt with.

UNIVERSITY OF OXFORD

DEPARTMENT OF EDUCATION

15 Norham Gardens, Oxford OX2 6PY

Tel: +44(0)1865 274024 Fax: +44(0)1865 274027

general.enquiries@education.ox.ac.uk www.education.ox.ac.uk

Dr Gabriel Stylianides

email: gabriel.stylianides@education.ox.ac.uk

Jisoo Seo, DPhil student, jisoo.seo@education.ox.ac.uk



If you remain unhappy or wish to make a formal complaint, please contact the Chair of the Research Ethics Committee at the University of Oxford who will seek to resolve the matter as soon as possible:

Chair, **Social Sciences & Humanities Inter-Divisional Research Ethics Committee;**

Email: ethics@socsci.ox.ac.uk;

Address: Research Services, University of Oxford, Wellington Square, Oxford OX1 2JD

11. Data Protection

The University of Oxford is the data controller with respect to your personal data, and as such will determine how your personal data is used in the study. The University will process your personal data for the purpose of the research outlined above. Research is a task that is performed in the public interest. Further information about your rights with respect to your personal data is available from <http://www.admin.ox.ac.uk/councilsec/compliance/gdpr/individualrights/>.

12. Your Child's Assent

To help your child decide whether he/she would like to participate in this study, we ask that you discuss the study with him/her. You can use text from below, which will also be read to your child before each activity session.

I am a teacher and I want to know more about how children like you think about maths. If you decide to help, I will ask you to imagine making some figures and to answer some questions about numbers. Some activities will be fun and easy, and some might be a little bit hard. It is okay if you do not know the answers. I am mostly interested in your thinking.

Also, I would like to audio record your voice while you solve some maths problems. But if you don't want to be audio recorded, you can say no, and we can still do maths together.

I would like to use your ideas in a paper that I am writing. When I write about your ideas, I will not use your name, so no one will know it is you. Instead, I will think of a made-up name to give you in the paper.

If at any time you don't want to do any more activities with me, you can just say you want to stop, and we will. No one will be cross with you if you want to stop. It will be okay if you want to stop at any point.

Thank you.

From Jisoo Seo

UNIVERSITY OF OXFORD

DEPARTMENT OF EDUCATION

15 Norham Gardens, Oxford OX2 6PY

Tel: +44(0)1865 274024 Fax: +44(0)1865 274027

general.enquiries@education.ox.ac.uk www.education.ox.ac.uk

Dr Gabriel Stylianides

email: gabriel.stylianides@education.ox.ac.uk

Jisoo Seo, DPhil student, jisoo.seo@education.ox.ac.uk



13. What should I do next?

Please fill in the enclosed form and return it to your child's class teacher if you would like your child to take part in this study. Please remember that you may withdraw your child within two months of giving consent, without any penalty and without giving a reason, by notifying the researcher.

14. Further Information and Contact Details

If you would like to discuss the research with someone beforehand (or if you have questions afterwards), please contact:

Professor Gabriel Stylianides

Department of Education

University of Oxford, 15 Norham Gardens, Oxford OX2 6PY, United Kingdom

University tel: +44 1865 611 002

University email: gabriel.stylianides@education.ox.ac.uk

UNIVERSITY OF OXFORD

DEPARTMENT OF EDUCATION

15 Norham Gardens, Oxford OX2 6PY

Tel: +44(0)1865 274024 Fax: +44(0)1865 274027

general.enquiries@education.ox.ac.uk www.education.ox.ac.uk

Dr Gabriel Stylianides

email: gabriel.stylianides@education.ox.ac.uk

Jisoo Seo, DPhil student, jisoo.seo@education.ox.ac.uk



PARENT/GUARDIAN CONSENT FORM (1/2)

Central University Research Ethics Committee (CUREC) Approval Reference: **[CIA-2223-013]**

**[The Role of Spatial Ability and Representation
in Early Childhood Arithmetic Word Problem Solving]**

This study investigates how young children think when solving arithmetic word problems.

Please initial each box

- | | | |
|---|--|----------------------|
| 1 | I confirm that I have read and understand the information sheet for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily. | <input type="text"/> |
| 2 | I understand that my child's participation is voluntary and that I am free to withdraw my child within two months of giving consent, without giving any reason, and without any adverse consequences or penalty. | <input type="text"/> |
| 3 | I understand who will have access to personal data provided, how the data will be stored and what will happen to the data at the end of the project. | <input type="text"/> |
| 4 | I understand that my child and I will not be identifiable from any publications or presentations. | <input type="text"/> |
| 5 | I understand how to raise a concern or make a complaint. | <input type="text"/> |
| 6 | I agree to take part. | <input type="text"/> |
| 7 | I have discussed the study with my child. My child gives his/her assent to taking part in the study. My child understands the project and that it's okay to stop taking part at any time. | <input type="text"/> |

UNIVERSITY OF OXFORD

DEPARTMENT OF EDUCATION

15 Norham Gardens, Oxford OX2 6PY

Tel: +44(0)1865 274024 Fax: +44(0)1865 274027

general.enquiries@education.ox.ac.uk www.education.ox.ac.uk

Dr Gabriel Stylianides

email: gabriel.stylianides@education.ox.ac.uk

Jisoo Seo, DPhil student, jisoo.seo@education.ox.ac.uk



PARENT/GUARDIAN CONSENT FORM (1/2)

Central University Research Ethics Committee (CUREC) Approval Reference: **[CIA-2223-013]**

**[The Role of Spatial Ability and Representation
in Early Childhood Arithmetic Word Problem Solving]**

This study investigates how young children think when solving arithmetic word problems.

[OPTIONAL]

I consent to my child being audio recorded, and my child assents to being audio recorded.

☐

If so, select one of the following options:

A) I do not wish for my child to be quoted even though he/she will not be identifiable

☐

or

I agree to the use of my child's quotations in research outputs if he/she is not identifiable

☐

Name of Child

dd / mm / yyyy
Child's date of birth

Girl / Boy
Child's Sex (circle one)

Name of Child's Teacher

Name of Child's School

Name of Parent/Guardian

dd / mm / yyyy
Today's Date

Parent/Guardian's Signature

Jisoo Seo
Name of person taking consent

dd / mm / yyyy
Date

Signature

Appendix 2H: Parent Information Sheet and Opt-In Consent Form (Simplified Version)

UNIVERSITY OF OXFORD

DEPARTMENT OF EDUCATION

15 Norham Gardens, Oxford OX2 6PY
Tel: +44(0)1865 274024 Fax: +44(0)1865 274027
general.enquiries@education.ox.ac.uk

Dr Gabriel Stylianides

gabriel.stylianides@education.ox.ac.uk

Jisoo Seo, DPhil student

jisoo.seo@education.ox.ac.uk



Dear parent/Carer,

An opportunity has arisen for the year 2 children to take part in a maths research project with Jisoo Seo, an Oxford University PhD student. The project will be looking at the relationship between children's spatial thinking and their ability to solve maths problems.

Your child will work on-on-one with Jisoo for one hour during the school day, with a break in the middle. The spatial and mathematical activities, such as puzzles and addition, are interesting, fun and appropriately challenging.

The data will be anonymized so that your child, and your child's school and teacher are not identifiable in any published materials. No recordings or photos of any kind will be made.

If you would like for child to participate in this study, please return the signed consent form back to the school by xxx. If you do agree to participation, you may withdraw your child from the study, without giving a reason, and without any negative consequences, by advising us of this decision. The deadline by which you can withdraw any information your child has contributed to the research is xxx.

Please fill out the slip below and return it to xxx by xxx.

If you have any questions or concerns about the study, please email Jisoo at jisoo.seo@education.ox.ac.uk.

With thanks and best wishes,

Jisoo Seo

Please tick one:

☐

I give permission for my child to take part in the maths project.

☐

I do not give permission for my child to take part in the maths project.

Signed _____ Name of Child _____

Appendix 2I: Item Difficulty and Discrimination Indices

Table 2I-1

Item Difficulty and Discrimination Indices

	Form A		Form B		Form C		Form D		Average	
	Difficulty (SD)	Discrimination (SD)	Difficulty (SD)	Discrimination (SD)	Difficulty (SD)	Discrimination (SD)	Difficulty (SD)	Discrimination (SD)	Difficulty (SD)	Discrimination (SD)
Item 1	.68 (0.60)	.54 (0.60)	.62 (0.55)	.55 (0.55)	.59 (0.66)	.66 (0.66)	.59 (0.56)	.72 (0.56)	.62 (0.04)	.62 (0.09)
Item 2	.65 (0.80)	.52 (0.80)	.47 (0.83)	.45 (0.83)	.38 (0.55)	.74 (0.55)	.50 (0.56)	.77 (0.56)	.50 (0.11)	.62 (0.16)
Item 3	.48 (0.85)	.51 (0.85)	.62 (1.04)	.32 (1.04)	.32 (0.53)	.51 (0.53)	.18 (0.46)	.34 (0.46)	.40 (0.19)	.42 (0.10)
Item 4	.42 (0.67)	.59 (0.67)	.59 (0.86)	.49 (0.86)	.29 (0.52)	.41 (0.52)	.32 (0.59)	.62 (0.59)	.41 (0.13)	.53 (0.10)
Item 5	.58 (0.81)	.50 (0.81)	.47 (0.79)	.26 (0.79)	.50 (0.56)	.80 (0.56)	.47 (0.51)	.73 (0.51)	.51 (0.05)	.57 (0.25)
Item 6	.35 (0.55)	.41 (0.55)	.59 (0.99)	.43 (0.99)	.44 (0.79)	.69 (0.79)	.29 (0.46)	.49 (0.46)	.42 (0.13)	.50 (0.13)
Item 7	.84 (0.45)	.42 (0.45)	.62 (0.49)	.49 (0.49)	.76 (0.43)	.51 (0.43)	.76 (0.50)	.51 (0.50)	.75 (0.09)	.48 (0.04)
Item 8	.84 (0.45)	.42 (0.45)	.71 (0.52)	.66 (0.52)	.88 (0.69)	.38 (0.69)	.79 (0.41)	.60 (0.41)	.81 (0.08)	.52 (0.14)
Item 9	.61 (0.84)	.45 (0.84)	.29 (0.63)	.47 (0.63)	.56 (0.79)	.51 (0.79)	.50 (0.56)	.56 (0.56)	.49 (0.14)	.50 (0.05)
Item 10	.71 (0.78)	.54 (0.78)	.21 (0.41)	.40 (0.41)	.41 (0.50)	.65 (0.50)	.50 (0.62)	.45 (0.62)	.46 (0.21)	.51 (0.11)
Item 11	.35 (0.55)	.55 (0.55)	.38 (0.78)	.48 (0.78)	.32 (0.47)	.69 (0.47)	.53 (0.86)	.55 (0.86)	.40 (0.09)	.57 (0.09)
Item 12	.48 (0.51)	.61 (0.51)	.71 (0.68)	.69 (0.68)	.71 (0.63)	.57 (0.63)	.74 (0.67)	.39 (0.67)	.66 (0.12)	.57 (0.13)
Item 13	.68 (0.83)	.51 (0.83)	.65 (1.01)	.48 (1.01)	.50 (1.05)	.57 (1.05)	.79 (0.23)	.44 (0.23)	.65 (0.12)	.50 (0.06)
Item 14	.29 (0.78)	.35 (0.78)	.32 (0.64)	.59 (0.64)	.38 (0.78)	.63 (0.78)	.21 (0.41)	.71 (0.41)	.30 (0.07)	.57 (0.15)
Item 15	.19 (0.40)	.54 (0.40)	.65 (1.12)	.32 (1.12)	.50 (1.02)	.29 (1.02)	.44 (0.08)	.32 (0.08)	.45 (0.19)	.37 (0.11)
Item 16	.42 (0.50)	.65 (0.50)	.56 (0.70)	.69 (0.70)	.50 (1.05)	.65 (1.05)	.44 (0.50)	.71 (0.50)	.48 (0.06)	.68 (0.03)
Item 17	.35 (0.49)	.55 (0.49)	.29 (0.63)	.56 (0.63)	.29 (0.46)	.66 (0.46)	.24 (0.43)	.67 (0.43)	.29 (0.05)	.61 (0.06)
Item 18	.35 (0.80)	.29 (0.80)	.59 (0.99)	.48 (0.99)	.41 (0.78)	.43 (0.78)	.32 (0.47)	.70 (0.47)	.42 (0.12)	.47 (0.17)
Item 19	.61 (0.50)	.30 (0.50)	.68 (0.91)	.50 (0.91)	.68 (0.84)	.51 (0.84)	.44 (0.50)	.58 (0.50)	.60 (0.11)	.47 (0.12)
Item 20	.58 (0.56)	.57 (0.56)	.32 (0.47)	.63 (0.47)	.29 (0.46)	.43 (0.46)	.32 (0.47)	.63 (0.47)	.38 (0.13)	.57 (0.09)
Item 21	.52 (1.03)	.42 (1.03)	.41 (0.61)	.62 (0.61)	.26 (0.45)	.59 (0.45)	.32 (0.64)	.55 (0.64)	.38 (0.11)	.55 (0.09)
Item 22	.16 (0.37)	.18 (0.37)	.26 (0.62)	.43 (0.62)	.09 (0.29)	.14 (0.29)	.29 (0.52)	.36 (0.52)	.20 (0.09)	.28 (0.14)
Item 23	.45 (0.62)	.62 (0.62)	.38 (0.55)	.48 (0.55)	.62 (0.99)	.52 (0.99)	.68 (0.07)	.35 (0.07)	.53 (0.14)	.49 (0.11)
Item 24	.39 (0.56)	.63 (0.56)	.47 (0.90)	.54 (0.90)	.50 (0.79)	.57 (0.79)	.47 (0.83)	.57 (0.83)	.46 (0.05)	.58 (0.04)
Average	.50 (0.18)	.49 (0.12)	.49 (0.16)	.50 (0.11)	.47 (0.18)	.55 (0.15)	.46 (0.19)	.55 (0.14)	.48 (0.17)	.52 (0.13)

Appendix 2J: Standardized Factor Loadings of Confirmatory Factor Analyses

Table 2J-1

M1 (One-Factor Model) Factor Loadings

	Factor 1 (β)	SE	z	p-value	Lower CI	Upper CI
Item1	.79	0.07	10.73	< .001	.65	.94
Item2	.77	0.08	9.71	< .001	.61	.92
Item3	.74	0.06	12.00	< .001	.62	.86
Item4	.70	0.07	10.04	< .001	.56	.83
Item5	.92	0.03	26.71	< .001	.85	.98
Item6	.72	0.07	9.90	< .001	.58	.86
Item7	.92	0.04	21.19	< .001	.84	1.01
Item8	.89	0.04	21.73	< .001	.81	.97
Item9	.64	0.08	8.24	< .001	.49	.80
Item10	.75	0.06	11.66	< .001	.62	.88
Item11	.89	0.04	20.81	< .001	.80	.97
Item12	.80	0.06	14.32	< .001	.69	.90
Item13	.55	0.09	6.34	< .001	.38	.72
Item14	.76	0.06	12.66	< .001	.64	.88
Item15	.73	0.07	11.05	< .001	.60	.86
Item16	.40	0.10	3.98	< .001	.20	.60
Item17	.81	0.05	14.99	< .001	.70	.91
Item18	.83	0.05	17.89	< .001	.74	.93
Item19	.76	0.06	12.89	< .001	.64	.87
Item20	.78	0.06	14.14	< .001	.67	.89
Item21	.70	0.07	9.97	< .001	.56	.83
Item22	.84	0.05	17.58	< .001	.75	.93
Item23	.89	0.04	25.50	< .001	.82	.96
Item24	.84	0.05	18.00	< .001	.74	.93

Note. Factor 1 = One-step additive word problems.

Table 2J-2

M2 (Two-Factors Model: Quantity) Factor Loadings

	Factor 1 (β)	Factor 2 (β)	<i>SE</i>	<i>z</i>	<i>p</i> -value	Lower CI	Upper CI
Item1	.80		0.07	10.78	< .001	.65	.94
Item2	.77		0.08	9.82	< .001	.62	.93
Item3	.75		0.06	12.22	< .001	.63	.87
Item4	.70		0.07	10.16	< .001	.57	.84
Item5	.93		0.03	27.16	< .001	.86	.99
Item6	.72		0.07	9.92	< .001	.58	.87
Item7	.93		0.04	21.77	< .001	.85	1.02
Item8	.90		0.04	22.00	< .001	.82	.98
Item9	.65		0.08	8.28	< .001	.50	.81
Item10	.76		0.07	11.72	< .001	.63	.89
Item11	.90		0.04	21.01	< .001	.81	.98
Item12	.81		0.06	14.27	< .001	.69	.92
Item13		.56	0.09	6.32	< .001	.39	.73
Item14		.77	0.06	12.79	< .001	.65	.89
Item15		.74	0.07	11.15	< .001	.61	.87
Item16		.41	0.10	3.96	< .001	.21	.61
Item17		.82	0.05	15.21	< .001	.71	.92
Item18		.85	0.05	18.08	< .001	.76	.94
Item19		.77	0.06	13.06	< .001	.65	.88
Item20		.79	0.06	14.23	< .001	.68	.90
Item21		.71	0.07	10.02	< .001	.57	.84
Item22		.86	0.05	17.49	< .001	.76	.95
Item23		.91	0.03	26.40	< .001	.84	.97
Item24		.85	0.05	18.39	< .001	.76	.94

Note. Factor 1 = Discrete-Quantity problems; Factor 2 = Continuous-Quantity problems.

Table 2J-3

M3 (Two-Factors Model: Quantitative Relation) Factor Loadings

	Factor 1 (β)	Factor 2 (β)	SE	z	p-value	Lower CI	Upper CI
Item1	.81		0.08	10.72	< .001	.66	.96
Item2	.78		0.08	9.78	< .001	.63	.94
Item3	.76		0.06	12.25	< .001	.64	.88
Item4	.71		0.07	10.26	< .001	.58	.85
Item5	.95		0.04	27.02	< .001	.88	1.01
Item6	.73		0.07	10.04	< .001	.59	.88
Item13	.56		0.09	6.41	< .001	.39	.74
Item14	.78		0.06	12.92	< .001	.66	.90
Item15	.74		0.07	11.24	< .001	.61	.87
Item16	.41		0.10	3.96	< .001	.21	.61
Item17	.82		0.05	15.30	< .001	.72	.93
Item18	.85		0.05	18.50	< .001	.76	.94
Item7		.94	0.04	21.27	< .001	.85	1.02
Item8		.90	0.04	21.64	< .001	.82	.98
Item9		.65	0.08	8.24	< .001	.50	.81
Item10		.76	0.07	11.74	< .001	.63	.89
Item11		.90	0.04	20.84	< .001	.82	.98
Item12		.81	0.06	14.57	< .001	.70	.91
Item19		.77	0.06	12.99	< .001	.65	.88
Item20		.79	0.06	14.31	< .001	.68	.90
Item21		.71	0.07	10.00	< .001	.57	.84
Item22		.85	0.05	17.91	< .001	.76	.94
Item23		.90	0.04	26.03	< .001	.83	.97
Item24		.84	0.05	18.19	< .001	.75	.94

Note. Factor 1 = Dynamic-Relation problems; Factor 2 = Static-Relation problems.

Table 2J-4

M4 (Four-Factors Model: Quantity $\times$ Quantitative Relation) Factor Loadings

	Factor 1 (β)	Factor 2 (β)	Factor 3 (β)	Factor 4 (β)	SE	z	p-value	Lower CI	Upper CI
Item1	.81				0.08	10.68	< .001	.66	.96
Item2	.79				0.08	10.00	< .001	.63	.94
Item3	.77				0.06	12.83	< .001	.65	.88
Item4	.72				0.07	10.53	< .001	.58	.85
Item5	.95				0.04	24.39	< .001	.88	1.03
Item6	.74				0.07	10.10	< .001	.60	.88
Item7		.93			0.04	22.04	< .001	.85	1.01
Item8		.89			0.04	21.70	< .001	.81	.98
Item9		.65			0.08	8.19	< .001	.49	.81
Item10		.76			0.07	11.67	< .001	.63	.88
Item11		.89			0.04	20.42	< .001	.81	.98
Item12		.80			0.06	14.12	< .001	.69	.91
Item13			.58		0.09	6.49	< .001	.40	.75
Item14			.79		0.06	13.11	< .001	.68	.91
Item15			.76		0.07	11.31	< .001	.63	.89
Item16			.42		0.11	3.94	< .001	.21	.63
Item17			.84		0.06	15.40	< .001	.73	.95
Item18			.87		0.05	18.61	< .001	.78	.97
Item19				.79	0.06	13.52	< .001	.68	.91
Item20				.81	0.06	14.54	< .001	.70	.92
Item21				.73	0.07	10.10	< .001	.59	.87
Item22				.88	0.05	18.10	< .001	.79	.98
Item23				.93	0.03	28.11	< .001	.86	.99
Item24				.87	0.05	18.64	< .001	.78	.96

Note. Factor 1 = [Discrete Quantity–Dynamic Relation] problems; Factor 2 = [Discrete Quantity–Static Relation] problems; Factor 3 = [Continuous Quantity–Dynamic Relation] problems; Factor 4 = [Continuous Quantity–Static Relation] problems.

Appendix 2K: Central University Research Ethics Committee Approval

UNIVERSITY OF OXFORD

DEPARTMENT OF EDUCATION

15 Norham Gardens, Oxford OX2 6PY

student.curec@education.ox.ac.uk; staff.curec@education.ox.ac.uk

DREC Chair: Professor Liam Gearon



CONFIDENTIAL

Jisoo Seo

November 9, 2022

Dear Jisoo,

Research Ethics Approval – CUREC 1a

Ethics Approval Reference: **CIA-2223-013**

Study Title: The Role of Spatial Ability and Representation in Early Childhood Arithmetic Word Problem Solving

The above application has been considered on behalf of the Departmental Research Ethics Committee (DREC) in accordance with the procedures laid down by the University for Ethical Approval of all research involving human participants.

I am pleased to inform you that, on the basis of the information provided to the DREC, the proposed research has been judged as meeting appropriate ethical standards, and accordingly, approval has been granted.

Please continue to follow *COVID-19: CUREC guidance on research involving human participants*
<https://researchsupport.admin.ox.ac.uk/governance/ethics/coronavirus>

If relevant, please also check the CUREC website for their **best practice guidance**:
<https://researchsupport.admin.ox.ac.uk/governance/ethics/resources/bpg>

Amendments

Should there be any subsequent changes to the study, you should submit details to the DREC for consideration and approval. Details of changes must be listed on an amendment form.

Please do not hesitate to contact me if you have any queries.

Good luck with your research study.

Yours sincerely,

Dr Pinar Kolancali
Post-doctoral Researcher
Departmental Research Ethics Committee Member
Email: pinar.kolancali@education.ox.ac.uk

Appendix 2L: Child Assent Script

Before the first testing session:

I am a teacher and I want to know more about how children like you think about maths. If you decide to help, I will ask you to imagine making some figures and to answer some questions about numbers. Some activities will be fun and easy, and some might be a little bit hard. It is okay if you do not know the answers. I am mostly interested in your thinking.

I would like to use your ideas in a paper that I am writing. When I write about your ideas, I will not use your name, so no one will know it is you. Instead, I will think of a made-up name to give you in the paper.

If at any time you don't want to do any more activities with me, you can just say you want to stop, and we will. No one will be cross with you if you want to stop. It will be okay if you want to stop at any point.

Thank you.

Before the second testing session:

Remember, if at any time you don't want to do any more activities with me, you can just say you want to stop, and we will. No one will be cross with you if you want to stop. It will be okay if you want to stop at any point.

Appendix 3A: Assumption Diagnostics for Linearity of the Logit

The linearity of the logit assumption for continuous predictors was assessed by refitting the models using natural cubic splines with 2–5 degrees of freedom for each continuous predictor (Gauthier et al., 2020; Harrell, 2015). Model improvement was evaluated using AIC and BIC to determine whether nonlinear terms were warranted. They are summarized Table 3B-1.

Table 3B-1

AIC and BIC Comparisons for Linear and Spline Specifications of Continuous Predictors

Variable	df	AIC	Δ AIC	BIC	Δ BIC
Composite Spatial Visualization	2	3041	1.85	3108	7.95
	3	3042	3.11	3115	15.31
	4	3040	1.35	3119	19.64
	5	3042	2.97	3127	27.36
Visuospatial Working Memory	2	3041	1.89	3108	7.99
	3	3037	−1.22	3111	10.97
	4	3039	0.63	3119	18.93
	5	3041	2.61	3127	27.00
Age	2	3041	1.96	3108	8.06
	3	3043	3.83	3116	16.03
	4	3045	5.83	3124	24.13
	5	3044	5.02	3129	29.41
Language Ability	2	3040	1.69	3107	7.79
	3	3042	3.02	3115	15.22
	4	3043	3.95	3122	22.25
	5	3045	6.16	3130	30.55
Numerical Ability	2	3035	−3.27	3103	2.83
	3	3032	−6.65	3105	5.55
	4	3026	−12.80	3105	5.49
	5	3026	−13.09	3111	11.30
Fluid Reasoning	2	3040	1.74	3108	7.84
	3	3039	0.03	3112	12.22
	4	3036	−2.85	3115	15.45
	5	3034	−4.77	3119	19.63

For composite spatial visualization, visuospatial working memory, language ability, fluid reasoning, and age, spline specifications yielded either negligible improvements or increases in AIC and substantially higher BIC values relative to the linear model, indicating limited evidence for meaningful nonlinearity. Predicted probability curves exhibited approximately monotonic trends across the observed range of values. Accordingly, linear specifications were retained for these predictors for parsimony (James et al., 2021).

For numerical ability, spline models produced a large reduction in AIC ($\Delta\text{AIC} = -12.80$ for $\text{df} = 4$, and $\Delta\text{AIC} = -13.09$ for $\text{df} = 5$), suggesting potential nonlinearity. However, BIC increased, reflecting the added model complexity. Visual inspection of predicted probability curves indicated that curvature was most pronounced in the extreme lower range of the distribution, where observations were sparse and confidence intervals were wide. Across the central range of observed values, the association between numerical ability and problem-solving performance appeared approximately linear. Given the mixed model fit indices, concentration of curvature in sparsely populated regions, and the principle of parsimony, the linear specification was retained.

Appendix 4A: Assumption Diagnostics of Mixed-Effects Multinomial Logistic Regression Models

The following assumptions of cross-classified multinomial logistic regression were evaluated: (1) linearity of the logit for continuous predictors, (2) absence of problematic multicollinearity among fixed effects, and (3) absence of influential clusters (Field, 2018; Hedeker, 2003; Gelman & Hill, 2007; Menard, 2010). Model assumptions were evaluated for the full model including interaction terms (i.e., RQ4).

Linearity of the Logit

The linearity of the logit assumption for continuous predictors was assessed by refitting the models using natural cubic splines with 2–5 degrees of freedom for each continuous predictor (Gauthier et al., 2020; Harrell, 2015). Model improvement was evaluated using AIC and BIC to determine whether nonlinear terms were warranted. As summarized in Table 4A-1, all predictors except for numerical ability were adequately modeled as linear.

For numerical ability, the spline specification ($df = 2$) resulted in a substantial reduction in AIC ($\Delta AIC = -28.6$) and BIC ($\Delta BIC = -4.22$), indicating evidence for nonlinearity. As illustrated in Figure 4A-1, visual inspection of the predicted probability curves further supported a nonlinear association between numerical ability and problem-solving strategy use. Higher numerical ability was associated with an accelerated increase in the probability of using ‘Quantitatively Correct Number Fact’ strategies and a nonlinear decline in ‘Non-Quantitative’ strategies, whereas some strategies (e.g., ‘Quantitatively Incorrect Counting’) exhibited relatively weak or approximately linear patterns. Given the substantial improvement in AIC and BIC and the meaningful curvature observed in the predicted probabilities, the spline specification ($df = 2$) was retained for numerical ability.

Table 4A-1*AIC and BIC Comparisons for Linear and Spline Specifications of Continuous Predictors*

Variable	df	AIC	Δ AIC	BIC	Δ BIC
Composite Spatial Visualization	2	6570	2.13	6912	26.5
	3	6574	6.09	6940	54.9
	4	6572	4.06	6962	77.2
	5	6578	10.4	6993	108
Visuospatial Working Memory	2	6570	2.09	6912	26.5
	3	6575	6.79	6941	55.6
	4	6580	12.3	6971	85.5
	5	6585	16.7	6999	114
Age	2	6574	5.71	6915	30.1
	3	6581	12.6	6947	61.4
	4	6577	8.6	6967	81.8
	5	6578	10.5	6993	108
Language Ability	2	6574	6.32	6916	30.7
	3	6578	9.75	6944	58.5
	4	6583	15.4	6974	88.6
	5	6589	21.3	7004	119
Numerical Ability	2	6539	-28.6	6881	-4.22
	3	6540	-28.2	6906	20.6
	4	6535	-32.8	6925	40.4
	5	6541	-27.3	6955	70.3
Fluid Reasoning	2	6567	-1.11	6908	23.3
	3	6568	0.199	6934	49
	4	6566	-2.45	6956	70.7
	5	6572	3.52	6986	101

Note. AIC of linear model = 6568; BIC of linear model = 6885. Each row replaces the listed variable with a restricted cubic spline of the indicated degrees of freedom, keeping other predictors linear. Δ AIC/ Δ BIC are relative to the linear model. The shaded row marks the smallest df with meaningful AIC and BIC improvement.

Note that the spline method was retained rather than applying mathematical transformations to numerical ability (e.g., squaring, cubing, or square-root transformations). Spline coefficients cannot be interpreted directly at the individual level in the same way as logistic regression coefficients (Harrell, 2015). Instead, splines generate multiple coefficients corresponding to basis functions that, when linearly combined, represent a smooth, piecewise polynomial relationship. Increasing the degrees of freedom allows for more curvature, which can better capture complex relationships; however, this also increases the risk of overfitting. Conversely, decreasing the degrees of freedom leads to a smoother and more generalized curve but may lead to fitting the data less precisely. Because two degrees of freedom were specified, two basis functions were generated, yielding two coefficients. It is the joint contribution of these coefficients that defines the specific non-linear functional form.

Because spline coefficients cannot be interpreted individually, their interpretation is less straightforward. Moreover, spline functions may be more difficult to replicate exactly across studies, as the estimated functional form is partly data-driven. However, this trade-off was considered acceptable because numerical ability was included as a covariate rather than as a primary variable of interest. As well, when appropriately constrained (e.g., by limiting the degrees of freedom, as done here with $df = 2$), splines can reduce the risk of overfitting relative to higher-order polynomial transformations, which impose rigid functional-form assumptions on the data, while still flexibly capturing non-linear relationships (Harrell, 2015).

Multicollinearity

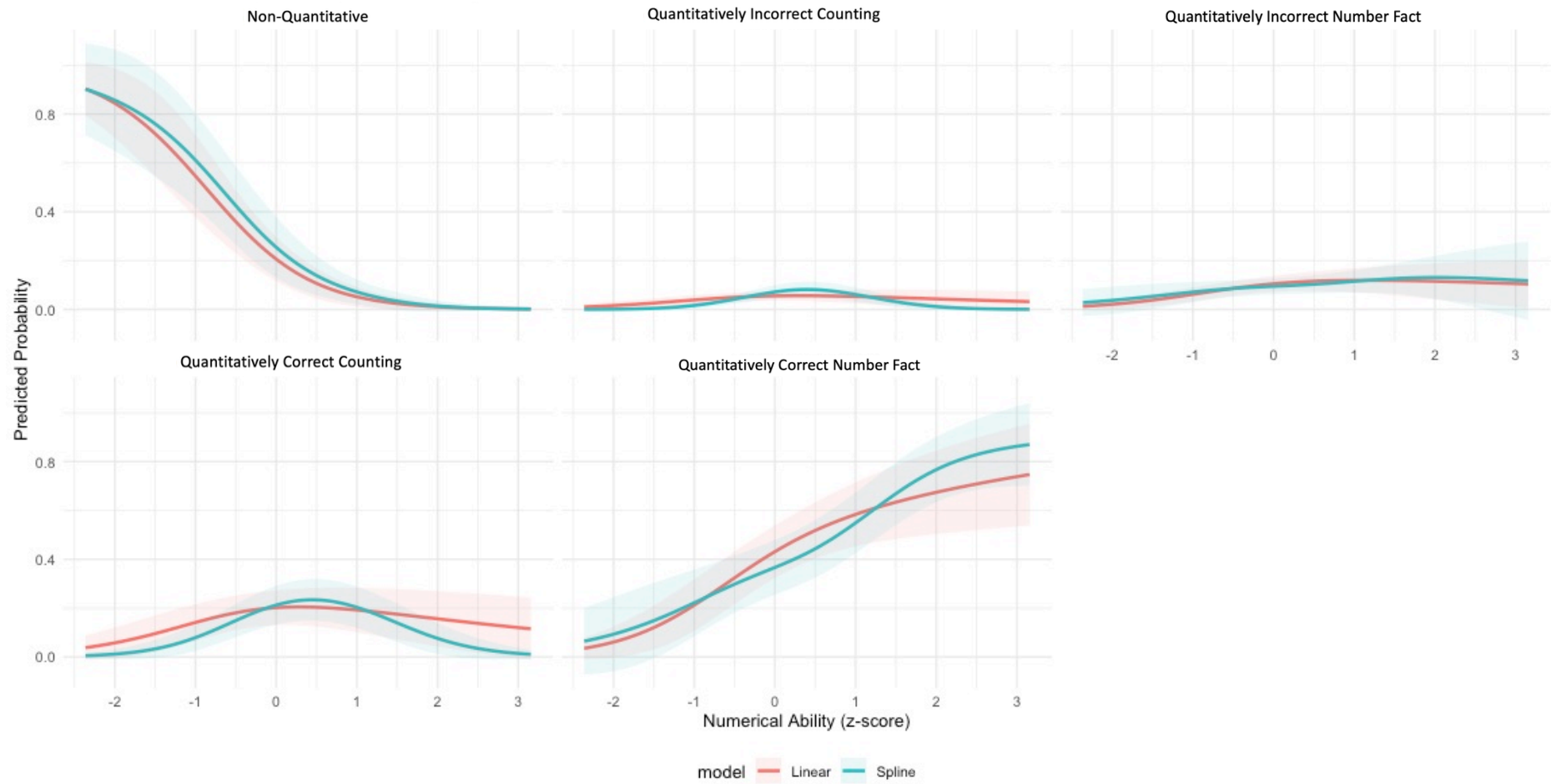
The assumption of absence of problematic multicollinearity was evaluated by first examining Pearson correlation coefficients among the continuous predictors. Correlations ranged from .01 to .57, all below the conventional threshold of .70 (Tabachnick & Fidell, 2013). Variance inflation factors (VIFs) were additionally computed from an auxiliary linear model including all predictors. All VIF values were below 2, indicating negligible multicollinearity.

Influential Clusters

Cluster influence was assessed using the Euclidean (L_2) distance between the observed and model-predicted response distributions for each participant and problem (Hedeker, 2003). For participants, L_2 values ranged from 0.02 to 0.16, and for problems, L_2 values ranged from 0.02 to 0.09. These values were small in magnitude and exhibited no extreme deviations, indicating that no participants or problem items exerted disproportionate influence on the model estimates. Thus, the model's conclusions were considered robust to potential cluster-level influence.

Figure 4A-1

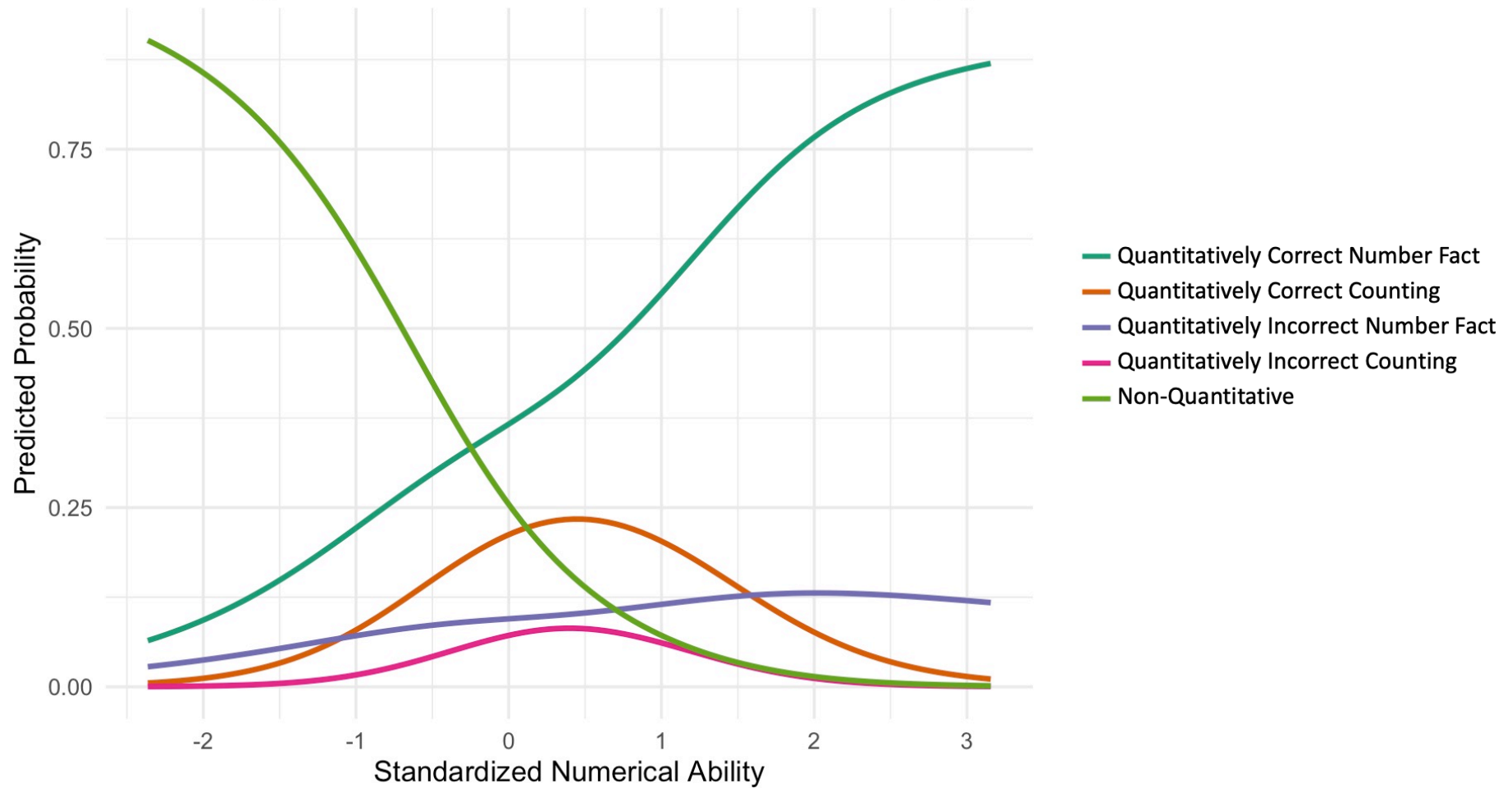
Predicted Probabilities of Strategy Use Across Numerical Ability: Linear and Spline Specifications



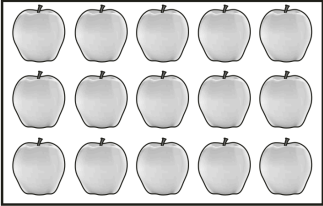
Appendix 4B: RQ3 Predicted Probabilities of Problem-Solving Strategies as Nonlinear Functions of Numerical Ability

Figure 4B-1

Nonlinear Relationship Between Numerical Ability and Predicted Problem-Solving Strategy Probabilities

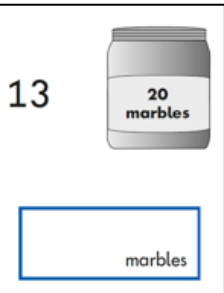
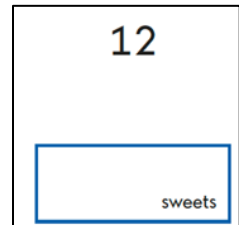


Appendix 4C: Discrete Quantity–Dynamic Relation Problems
the Key Stage 1 National Curriculum Assessments

Year	Problems
2018	-
2019 ^a	Ben has 63 beads. He gives 37 beads away. How many beads does Ben have left? beads
2022	-
2023	-
2024 ^b	There are 16 ducks on the pond. 5 ducks fly away. How many ducks stay on the pond? ducks Sam has 90 marbles. He gives 50 marbles to Ben. How many marbles does Sam have left? marbles Ajay had 18 strawberries. He picked some more. Ajay now has 24 strawberries. How many more strawberries did he pick? strawberries
2025 ^c	Kemi has 20 seeds. She gives some to Ben. Kemi has 8 seeds left. How many seeds does Kemi give to Ben? seeds A shopkeeper has 15 apples in a box. He sells 6 of the apples. How many apples does he have left?  apples

Note. ^{a-c}Contains material developed by the Standards and Testing Agency for 2019, 2024, and 2025 national curriculum assessments and licensed under Open Government Licence v3.0 (<https://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/>). The Key Stage 1 National Curriculum Assessments were not administered in 2020 and 2021 due to the COVID-19 pandemic.

**Appendix 4D: Discrete Quantity–Static Relation Problems from
the Key Stage 1 National Curriculum Assessments**

Year	Problems
2018 ^a	There are thirteen marbles in a jar. The jar can hold 20 marbles. How many more marbles can fit in the jar? 13 
2019 ^b	Ben has twelve sweets. Amy has seven more sweets than Ben. How many sweets does Amy have? 12 
2022	-
2023 ^c	Ben wants to build a tower with 15 blocks. He only has 9 blocks. How many more blocks does Ben need?  
2024	-
2025	-

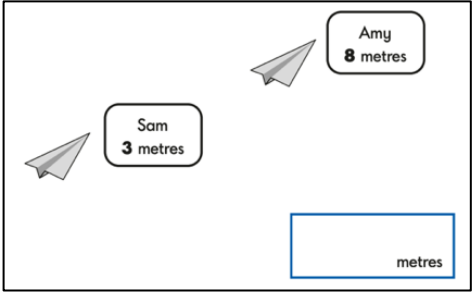
Note. ^{a-c}Contains material developed by the Standards and Testing Agency for 2018, 2019, and 2023 national curriculum assessments and licensed under Open Government Licence v3.0 (<https://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/>). The Key Stage 1 National Curriculum Assessments were not administered in 2020 and 2021 due to the COVID-19 pandemic.

**Appendix 4E: Continuous Quantity–Dynamic Relation problems from
the Key Stage 1 National Curriculum Assessments**

Year	Problems
2018 ^a	Ajay's plant was 11 centimetres tall. It grows 7 centimetres taller. How tall is the plant now?  cm
2019	-
2022	-
2023	-
2024	-
2025	-

Note. ^aContains material developed by the Standards and Testing Agency for 2018 national curriculum assessments and licensed under Open Government Licence v3.0 (<https://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/>). The Key Stage 1 National Curriculum Assessments were not administered in 2020 and 2021 due to the COVID-19 pandemic.

**Appendix 4F: Continuous Quantity–Static Relation problems from
the Key Stage 1 National Curriculum Assessments**

Year	Problems
2018	-
2019	-
2022	-
2023	-
2024 ^a	Amy and Sam are flying paper aeroplanes. Amy throws her aeroplane eight metres. Sam throws his aeroplane three metres. How much further did Amy’s aeroplane fly than Sam’s? 
2025	-

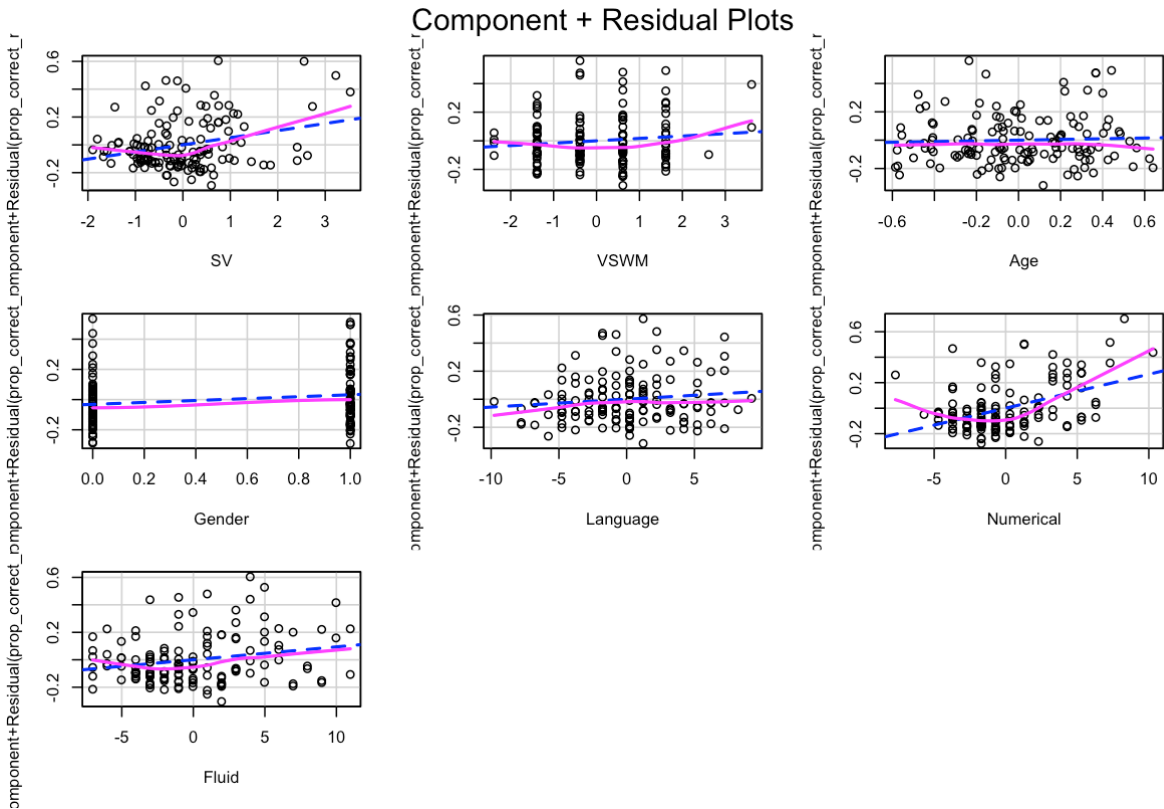
Note. ^{a-c}Contains material developed by the Standards and Testing Agency for 2024 national curriculum assessments and licensed under Open Government Licence v3.0 (<https://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/>). The Key Stage 1 National Curriculum Assessments were not administered in 2020 and 2021 due to the COVID-19 pandemic.

Appendix 5A: Assumption Diagnostics of Path Analysis

Linearity

Table 5A-1

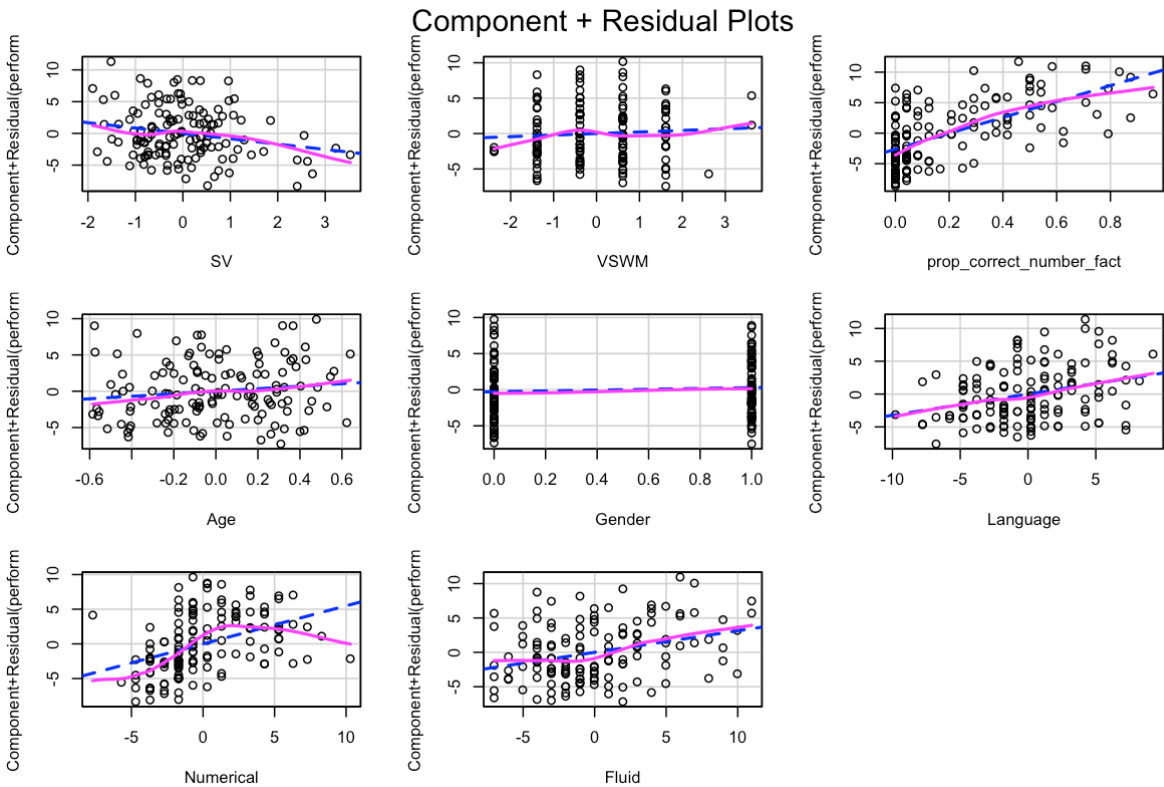
Component-Plus-Residual Plots for the Mediation Model



Note. Component-plus-residual plots showing the relationship between each predictor and the mediator, controlling for other variables. The dashed blue line represents the linear fit, and the solid pink line represents a LOESS smoother.

Table 5A-2

Component-Plus-Residual Plots for the Outcome Model



Note. Component-plus-residual plots showing the relationship between each predictor and the mediator, controlling for other variables. The dashed blue line represents the linear fit, and the solid pink line represents a LOESS smoother.

Appendix 5B: Path Coefficients (95% Percentile-Based CI) and Indirect Effects Across Problem Types

Table 5B-1

Path Coefficients (95% Percentile-Based CI) Predicting Quantitative Correct Number-Fact Strategy use and Problem-Solving Accuracy

	‘Quantitatively Correct Number Fact’ Strategy				Problem-Solving Accuracy			
	<i>B</i>	β	SE	95% Percentile-Based CI	<i>B</i>	β	SE	95% Percentile-Based CI
Spatial Visualization	0.05	.21	0.03	[−0.005, 0.10]	−0.84	−.12	0.49	[−1.80, 0.11]
Visuospatial Working Memory	0.02	.08	0.01	[−0.01, 0.04]	0.22	.04	0.31	[−0.42, 0.80]
Language Ability	0.01	.09	0.004	[−0.002, 0.01]	0.32	.18	0.11	[0.11, 0.53]
Numerical Ability	0.03	.35	0.01	[0.01, 0.04]	0.55	.25	0.20	[0.19, 0.98]
Fluid Reasoning	0.01	.16	0.01	[−0.0002, 0.02]	0.31	.19	0.12	[0.08, 0.55]
Age	0.03	.03	0.05	[−0.08, 0.13]	1.73	.07	1.32	[−0.91, 4.22]
Gender	0.06	.13	0.03	[0.002, 0.12]	0.57	.04	0.74	[−0.87, 2.01]
Quantitatively Correct Number-Fact Strategy	-	-	-	-	13.04	.45	2.27	[8.33, 17.29]

Note. *B* = unstandardized coefficient; β = standardized coefficient; SE = standard error; Strategy = ‘Quantitatively Correct Number Fact’ strategy. Confidence intervals (CIs) are percentile-based bootstrap intervals based on 5,000 resamples. Statistical significance was evaluated based on 95% percentile-based CIs and is indicated by boldface.

Table 5B-2

Indirect Effects (95% Percentile-Based CI) of Spatial Skills on Problem-Solving Accuracy

Indirect Effect	<i>B</i>	β	SE	95% Percentile- Based CI
Spatial Visualization				
→ ‘Quantitatively Correct Number Fact’ strategy	0.67	.09	0.39	[-0.07, 1.48]
→ Problem-Solving Accuracy				
Visuospatial Working Memory				
→ ‘Quantitatively Correct Number Fact’ strategy	0.22	.04	0.18	[-0.14, 0.58]
→ Problem-Solving Accuracy				

Note. *B* = unstandardized coefficient; β = standardized coefficient. SE = standard error. Confidence intervals (CIs) are percentile-based bootstrap intervals based on 5,000 resamples. Statistical significance was evaluated based on 95% percentile-based CIs.

Appendix 5C: Path Coefficients (95% BCa CI) and Indirect Effects by Problem Type

Table 5C-1

Path Coefficients (95% BCa CI) Predicting Quantitative Correct Number-Fact Strategy use and Problem-Solving Accuracy of Discrete Quantity–Dynamic Relation Problems

	‘Quantitatively Correct Number Fact’ Strategy				Problem-Solving Accuracy			
	<i>B</i>	β	SE	95% BCa CI	<i>B</i>	β	SE	95% BCa CI
Spatial Visualization	0.05	.15	0.04	[−0.02, 0.12]	0.0004	.0004	0.16	[−0.31, 0.30]
Visuospatial Working Memory	0.02	.05	0.02	[−0.02, 0.06]	0.04	.02	0.11	[−0.20, 0.25]
Language Ability	0.01	.10	0.01	[−0.004, 0.02]	0.06	.12	0.04	[−0.006, 0.13]
Numerical Ability	0.03	.30	0.01	[0.01, 0.05]	0.19	.32	0.05	[0.09, 0.30]
Fluid Reasoning	0.02	.23	0.01	[0.003, 0.03]	0.08	.18	0.04	[0.004, 0.15]
Age	0.03	.03	0.07	[−0.11, 0.17]	0.41	.06	0.45	[−0.49, 1.25]
Gender	0.08	.12	0.05	[−0.01, 0.17]	0.22	.06	0.24	[−0.25, 0.69]
Quantitatively Correct Number-Fact Strategy	-	-	-	-	1.56	.27	0.41	[0.75, 2.34]

Note. *B* = unstandardized coefficient; β = standardized coefficient; SE = standard error; Strategy = ‘Quantitatively Correct Number Fact’ strategy. Confidence intervals (CIs) are bias-corrected and accelerated (BCa) bootstrap intervals based on 5,000 resamples. Statistical significance was evaluated based on 95% BCa CIs and is indicated by boldface.

Table 5C-2

Indirect Effects (95% BCa CI) of Spatial Skills on Problem-Solving Accuracy of Discrete Quantity–Dynamic Relation Problems

Indirect Effect	<i>B</i>	β	SE	95% BCa CI
Spatial Visualization				
→ ‘Quantitatively Correct Number Fact’ strategy	0.08	.04	0.07	[–0.03, 0.24]
→ Problem-Solving Accuracy				
Visuospatial Working Memory				
→ ‘Quantitatively Correct Number Fact’ strategy	0.02	.01	0.03	[–0.03, 0.10]
→ Problem-Solving Accuracy				

Note. *B* = unstandardized coefficient; β = standardized coefficient. SE = standard error. Confidence intervals (CIs) are bias-corrected and accelerated (BCa) bootstrap intervals based on 5,000 resamples. Statistical significance was evaluated based on 95% BCa CIs.

Table 5C-3

Path Coefficients (95% BCa CI) Predicting Quantitative Correct Number-Fact Strategy use and Problem-Solving Accuracy of Discrete Quantity–Static Relation problems

	‘Quantitatively Correct Number Fact’ Strategy				Problem-Solving Accuracy			
	<i>B</i>	β	SE	95% BCa CI	<i>B</i>	β	SE	95% BCa CI
Spatial Visualization	0.07	.24	0.03	[0.01, 0.13]	−0.29	−.14	0.15	[−0.57, 0.03]
Visuospatial Working Memory	0.02	.07	0.02	[−0.02, 0.05]	0.03	.02	0.10	[−0.18, 0.22]
Language Ability	-1.57×10^{-6}	-2.19×10^{-5}	0.005	[−0.01, 0.01]	0.11	.21	0.04	[0.04, 0.19]
Numerical Ability	0.03	.31	0.01	[0.005, 0.05]	0.15	.23	0.06	[0.03, 0.27]
Fluid Reasoning	0.01	.21	0.01	[0.003, 0.03]	0.12	.24	0.04	[0.04, 0.20]
Age	0.003	.003	0.06	[−0.11, 0.12]	0.60	.09	0.41	[−0.24, 1.37]
Gender	0.04	.07	0.04	[−0.04, 0.11]	0.27	.06	0.25	[−0.22, 0.77]
Quantitatively Correct Number-Fact Strategy	-	-	-	-	2.99	.40	0.57	[1.85, 4.07]

Note. *B* = unstandardized coefficient; β = standardized coefficient; SE = standard error; Strategy = ‘Quantitatively Correct Number Fact’ strategy. Confidence intervals (CIs) are bias-corrected and accelerated (BCa) bootstrap intervals based on 5,000 resamples. Statistical significance was evaluated based on 95% BCa CIs and is indicated by boldface.

Table 5C-4

Indirect Effects (95% BCa CI) of Spatial Skills on Problem-Solving Accuracy of Discrete Quantity–Static Relation Problems

Indirect Effect	<i>B</i>	β	SE	95% BCa CI
Spatial Visualization				
→ ‘Quantitatively Correct Number Fact’ strategy	0.20	.10	0.10	[0.04, 0.45]
→ Problem-Solving Accuracy				
Visuospatial Working Memory				
→ ‘Quantitatively Correct Number Fact’ strategy	0.05	.03	0.05	[-0.04, 0.16]
→ Problem-Solving Accuracy				

Note. *B* = unstandardized coefficient; β = standardized coefficient. SE = standard error. Confidence intervals (CIs) are bias-corrected and accelerated (BCa) bootstrap intervals based on 5,000 resamples. Statistical significance was evaluated based on 95% BCa CIs and is indicated by boldface.

Table 5C-5

Path Coefficients (95% BCa CI) Predicting Quantitative Correct Number-Fact Strategy use and Problem-Solving Accuracy of Continuous Quantity–Dynamic Relation problems

	‘Quantitatively Correct Number Fact’ Strategy				Problem-Solving Accuracy			
	<i>B</i>	β	SE	95% BCa CI	<i>B</i>	β	SE	95% BCa CI
Spatial Visualization	0.07	.29	0.03	[0.01, 0.13]	−0.27	−.15	0.15	[−0.55, 0.02]
Visuospatial Working Memory	−0.001	−.01	0.02	[−0.03, 0.03]	0.15	.10	0.09	[−0.04, 0.33]
Language Ability	0.01	.09	0.01	[−0.004, 0.02]	0.10	.20	0.03	[0.04, 0.16]
Numerical Ability	0.02	.28	0.01	[0.001, 0.04]	0.10	.17	0.05	[0.001, 0.21]
Fluid Reasoning	0.003	.05	0.01	[−0.01, 0.02]	0.06	.13	0.04	[−0.01, 0.13]
Age	0.03	.03	0.06	[−0.09, 0.15]	0.20	.03	0.38	[−0.57, 0.91]
Gender	0.07	.14	0.04	[−0.001, 0.15]	0.08	.02	0.21	[−0.32, 0.51]
Quantitatively Correct Number-Fact Strategy	-	-	-	-	4.00	.55	0.55	[2.90, 5.07]

Note. *B* = unstandardized coefficient; β = standardized coefficient; SE = standard error; Strategy = ‘Quantitatively Correct Number Fact’ strategy. Confidence intervals (CIs) are bias-corrected and accelerated (BCa) bootstrap intervals based on 5,000 resamples. Statistical significance was evaluated based on 95% BCa CIs and is indicated by boldface.

Table 5C-6

Indirect Effects (95% BCa CI) of Spatial Skills on Problem-Solving Accuracy of Continuous Quantity–Dynamic Relation Problems

Indirect Effect	<i>B</i>	β	SE	95% BCa CI
Spatial Visualization				
→ ‘Quantitatively Correct Number Fact’ strategy	0.29	.16	0.13	[0.05, 0.57]
→ Problem-Solving Accuracy				
Visuospatial Working Memory				
→ ‘Quantitatively Correct Number Fact’ strategy	–0.01	–.004	0.07	[–0.14, 0.12]
→ Problem-Solving Accuracy				

Note. *B* = unstandardized coefficient; β = standardized coefficient. SE = standard error. Confidence intervals (CIs) are bias-corrected and accelerated (BCa) bootstrap intervals based on 5,000 resamples. Statistical significance was evaluated based on 95% BCa CIs and is indicated by boldface.

Table 5C-7

Path Coefficients (95% BCa CI) Predicting Quantitative Correct Number-Fact Strategy use and Problem-Solving Accuracy of Continuous Quantity–Static Relation problems

	‘Quantitatively Correct Number Fact’ Strategy				Problem-Solving Accuracy			
	<i>B</i>	β	SE	95% BCa CI	<i>B</i>	β	SE	95% BCa CI
Spatial Visualization	0.02	.06	0.03	[−0.04, 0.07]	−0.24	−.12	0.16	[−0.53, 0.09]
Visuospatial Working Memory	0.04	.18	0.02	[0.0003, 0.08]	0.01	.003	0.10	[−0.18, 0.20]
Language Ability	0.01	.12	0.01	[−0.004, 0.02]	0.06	.11	0.04	[−0.01, 0.13]
Numerical Ability	0.03	.35	0.01	[0.01, 0.05]	0.14	.22	0.07	[0.003, 0.28]
Fluid Reasoning	0.003	.04	0.01	[−0.01, 0.01]	0.09	.18	0.04	[0.02, 0.16]
Age	0.04	.05	0.06	[−0.07, 0.17]	0.54	.08	0.42	[−0.31, 1.33]
Gender	0.07	.13	0.03	[0.005, 0.14]	0.04	.01	0.26	[−0.44, 0.58]
Quantitatively Correct Number-Fact Strategy	-	-	-	-	3.89	.49	0.84	[2.27, 5.54]

Note. *B* = unstandardized coefficient; β = standardized coefficient; SE = standard error; Strategy = ‘Quantitatively Correct Number Fact’ strategy. Confidence intervals (CIs) are bias-corrected and accelerated (BCa) bootstrap intervals based on 5,000 resamples. Statistical significance was evaluated based on 95% BCa CIs and is indicated by boldface.

Table 5C-8

Indirect Effects (95% BCa CI) of Spatial Skills on Problem-Solving Accuracy of Continuous Quantity–Static Relation Problems

Indirect Effect	<i>B</i>	β	SE	95% BCa CI
Spatial Visualization				
→ ‘Quantitatively Correct Number Fact’ strategy	0.06	.03	0.12	[–0.16, 0.31]
→ Problem-Solving Accuracy				
Visuospatial Working Memory				
→ ‘Quantitatively Correct Number Fact’ strategy	0.15	.09	0.08	[0.01, 0.34]
→ Problem-Solving Accuracy				

Note. *B* = unstandardized coefficient; β = standardized coefficient. SE = standard error. Confidence intervals (CIs) are bias-corrected and accelerated (BCa) bootstrap intervals based on 5,000 resamples. Statistical significance was evaluated based on 95% BCa CIs and is indicated by boldface.

Appendix 5D: Path Coefficients (95% Percentile-Based CI) and Indirect Effects by Problem Type

Table 5D-1

Path Coefficients (95% Percentile-Based CI) Predicting Quantitative Correct Number-Fact Strategy use and Problem-Solving Accuracy of Discrete Quantity–Dynamic Relation Problems

	‘Quantitatively Correct Number Fact’ Strategy				Problem-Solving Accuracy			
	<i>B</i>	β	SE	95% Percentile- Based CI	<i>B</i>	β	SE	95% Percentile- Based CI
Spatial Visualization	0.05	.15	0.04	[−0.03, 0.12]	0.0004	.0004	0.16	[−0.31, 0.30]
Visuospatial Working Memory	0.02	.05	0.02	[−0.02, 0.05]	0.04	.02	0.11	[−0.19, 0.25]
Language Ability	0.01	.10	0.01	[−0.004, 0.02]	0.06	.12	0.04	[−0.009, 0.13]
Numerical Ability	0.03	.30	0.01	[0.01, 0.05]	0.19	.32	0.05	[0.10, 0.30]
Fluid Reasoning	0.02	.23	0.01	[0.003, 0.03]	0.08	.18	0.04	[0.01, 0.16]
Age	0.03	.03	0.07	[−0.11, 0.17]	0.41	.06	0.45	[−0.49, 1.25]
Gender	0.08	.12	0.05	[−0.01, 0.17]	0.22	.06	0.24	[−0.25, 0.69]
Quantitatively Correct Number-Fact Strategy	-	-	-	-	1.56	.27	0.41	[0.75, 2.33]

Note. *B* = unstandardized coefficient; β = standardized coefficient; SE = standard error; Strategy = ‘Quantitatively Correct Number Fact’ strategy. Confidence intervals (CIs) are percentile-based bootstrap intervals based on 5,000 resamples. Statistical significance was evaluated based on 95% percentile-based CIs and is indicated by boldface.

Table 5D-2

Indirect Effects (95% Percentile-Based CI) of Spatial Skills on Problem-Solving Accuracy of Discrete Quantity–Dynamic Relation Problems

Indirect Effect	<i>B</i>	β	SE	95% Percentile- Based CI
Spatial Visualization				
→ ‘Quantitatively Correct Number Fact’ strategy	0.08	.04	0.07	[–0.03, 0.22]
→ Problem-Solving Accuracy				
Visuospatial Working Memory				
→ ‘Quantitatively Correct Number Fact’ strategy	0.02	.01	0.03	[–0.04, 0.09]
→ Problem-Solving Accuracy				

Note. *B* = unstandardized coefficient; β = standardized coefficient. SE = standard error. Confidence intervals (CIs) are percentile-based bootstrap intervals based on 5,000 resamples. Statistical significance was evaluated based on 95% percentile-based CIs.

Table 5D-3

Path Coefficients (95% Percentile-Based CI) Predicting Quantitative Correct Number-Fact Strategy use and Problem-Solving Accuracy of Discrete Quantity–Static Relation problems

	‘Quantitatively Correct Number Fact’ Strategy				Problem-Solving Accuracy			
	<i>B</i>	β	SE	95% Percentile- Based CI	<i>B</i>	β	SE	95% Percentile- Based CI
Spatial Visualization	0.07	.24	0.03	[0.01, 0.13]	−0.29	−.14	0.15	[−0.58, 0.02]
Visuospatial Working Memory	0.02	.07	0.02	[−0.02, 0.05]	0.03	.02	0.10	[−0.17, 0.22]
Language Ability	-1.57×10^{-6}	-2.19×10^{-5}	0.005	[−0.01, 0.01]	0.11	.21	0.04	[0.04, 0.18]
Numerical Ability	0.03	.31	0.01	[0.005, 0.05]	0.15	.23	0.06	[0.04, 0.28]
Fluid Reasoning	0.01	.21	0.01	[0.002, 0.03]	0.12	.24	0.04	[0.04, 0.20]
Age	0.003	.003	0.06	[−0.11, 0.12]	0.60	.09	0.41	[−0.25, 1.36]
Gender	0.04	.07	0.04	[−0.04, 0.11]	0.27	.06	0.25	[−0.24, 0.76]
Quantitatively Correct Number-Fact Strategy	-	-	-	-	2.99	.40	0.57	[1.81, 4.04]

Note. *B* = unstandardized coefficient; β = standardized coefficient; SE = standard error; Strategy = ‘Quantitatively Correct Number Fact’ strategy. Confidence intervals (CIs) are percentile-based bootstrap intervals based on 5,000 resamples. Statistical significance was evaluated based on 95% percentile-based CIs and is indicated by boldface.

Table 5D-4

Indirect Effects (95% Percentile-Based CI) of Spatial Skills on Problem-Solving Accuracy of Discrete Quantity–Static Relation Problems

Indirect Effect	<i>B</i>	β	SE	95% Percentile- Based CI
Spatial Visualization				
→ ‘Quantitatively Correct Number Fact’ strategy	0.20	.10	0.10	[0.02, 0.42]
→ Problem-Solving Accuracy				
Visuospatial Working Memory				
→ ‘Quantitatively Correct Number Fact’ strategy	0.05	.03	0.05	[-0.04, 0.15]
→ Problem-Solving Accuracy				

Note. *B* = unstandardized coefficient; β = standardized coefficient. SE = standard error. Confidence intervals (CIs) are percentile-based bootstrap intervals based on 5,000 resamples. Statistical significance was evaluated based on 95% percentile-based CIs and is indicated by boldface.

Table 5D-5

Path Coefficients (95% Percentile-Based CI) Predicting Quantitative Correct Number-Fact Strategy use and Problem-Solving Accuracy of Continuous Quantity–Dynamic Relation problems

	‘Quantitatively Correct Number Fact’ Strategy				Problem-Solving Accuracy			
	<i>B</i>	β	SE	95% Percentile- Based CI	<i>B</i>	β	SE	95% Percentile- Based CI
Spatial Visualization	0.07	.29	0.03	[0.01, 0.13]	−0.27	−.15	0.15	[−0.55, 0.02]
Visuospatial Working Memory	−0.001	−.01	0.02	[−0.04, 0.03]	0.15	.10	0.09	[−0.03, 0.34]
Language Ability	0.01	.09	0.01	[−0.004, 0.02]	0.10	.20	0.03	[0.04, 0.15]
Numerical Ability	0.02	.28	0.01	[0.001, 0.04]	0.10	.17	0.05	[0.005, 0.21]
Fluid Reasoning	0.003	.05	0.01	[−0.01, 0.02]	0.06	.13	0.04	[−0.02, 0.13]
Age	0.03	.03	0.06	[−0.09, 0.15]	0.20	.03	0.38	[−0.54, 0.94]
Gender	0.07	.14	0.04	[−0.003, 0.14]	0.08	.02	0.21	[−0.33, 0.50]
Quantitatively Correct Number-Fact Strategy	-	-	-	-	4.00	.55	0.55	[2.89, 5.06]

Note. *B* = unstandardized coefficient; β = standardized coefficient; SE = standard error; Strategy = ‘Quantitatively Correct Number Fact’ strategy. Confidence intervals (CIs) are percentile-based bootstrap intervals based on 5,000 resamples. Statistical significance was evaluated based on 95% percentile-based CIs and is indicated by boldface.

Table 5D-6

Indirect Effects (95% Percentile-Based CI) of Spatial Skills on Problem-Solving Accuracy of Continuous Quantity–Dynamic Relation Problems

Indirect Effect	<i>B</i>	β	SE	95% Percentile- Based CI
Spatial Visualization				
→ ‘Quantitatively Correct Number Fact’ strategy	0.29	.16	0.13	[0.03, 0.54]
→ Problem-Solving Accuracy				
Visuospatial Working Memory				
→ ‘Quantitatively Correct Number Fact’ strategy	–0.01	–.004	0.07	[–0.15, 0.11]
→ Problem-Solving Accuracy				

Note. *B* = unstandardized coefficient; β = standardized coefficient. SE = standard error. Confidence intervals (CIs) are percentile-based bootstrap intervals based on 5,000 resamples. Statistical significance was evaluated based on 95% percentile-based CIs and is indicated by boldface.

Table 5D-7

Path Coefficients (95% Percentile-Based CI) Predicting Quantitative Correct Number-Fact Strategy use and Problem-Solving Accuracy of Continuous Quantity–Static Relation problems

	‘Quantitatively Correct Number Fact’ Strategy				Problem-Solving Accuracy			
	<i>B</i>	β	SE	95% Percentile- Based CI	<i>B</i>	β	SE	95% Percentile- Based CI
Spatial Visualization	0.02	.06	0.03	[−0.05, 0.07]	−0.24	−.12	0.16	[−0.56, 0.06]
Visuospatial Working Memory	0.04	.18	0.02	[−0.002, 0.08]	0.01	.003	0.10	[−0.19, 0.19]
Language Ability	0.01	.12	0.01	[−0.003, 0.02]	0.06	.11	0.04	[−0.01, 0.13]
Numerical Ability	0.03	.35	0.01	[0.01, 0.05]	0.14	.22	0.07	[0.01, 0.29]
Fluid Reasoning	0.003	.04	0.01	[−0.01, 0.02]	0.09	.18	0.04	[0.02, 0.16]
Age	0.04	.05	0.06	[−0.07, 0.17]	0.54	.08	0.42	[−0.27, 1.36]
Gender	0.07	.13	0.03	[0.003, 0.13]	0.04	.01	0.26	[−0.49, 0.53]
Quantitatively Correct Number-Fact Strategy	-	-	-	-	3.89	.49	0.84	[2.30, 5.59]

Note. *B* = unstandardized coefficient; β = standardized coefficient; SE = standard error; Strategy = ‘Quantitatively Correct Number Fact’ strategy. Confidence intervals (CIs) are percentile-based bootstrap intervals based on 5,000 resamples. Statistical significance was evaluated based on 95% percentile-based CIs and is indicated by boldface.

Table 5D-8

Indirect Effects (95% Percentile-Based CI) of Spatial Skills on Problem-Solving Accuracy of Continuous Quantity–Static Relation Problems

Indirect Effect	<i>B</i>	β	SE	95% Percentile- Based CI
Spatial Visualization				
→ ‘Quantitatively Correct Number Fact’ strategy	0.06	.03	0.12	[-0.18, 0.30]
→ Problem-Solving Accuracy				
Visuospatial Working Memory				
→ ‘Quantitatively Correct Number Fact’ strategy	0.15	.09	0.08	[-0.01, 0.32]
→ Problem-Solving Accuracy				

Note. *B* = unstandardized coefficient; β = standardized coefficient. SE = standard error. Confidence intervals (CIs) are percentile-based bootstrap intervals based on 5,000 resamples. Statistical significance was evaluated based on 95% percentile-based CIs.