Knowing what you know

A pedagogical model based on learners' metacognitive abilities

Maria Blancas Muñoz

TESI DOCTORAL UPF / YEAR 2020

Thesis supervisor

Dr. Paul Verschure,

DEPARTMENT OF INFORMATION AND COMMUNICATIONS TECHNOLOGIES

Universitat Pompeu Fabra Barcelona

“You should never be ashamed of what you don’t know
but of what you don’t want to learn”
— Fernando Blancas,
great domino player and crosswords lover,
who I had the luck to call “grandpa”

In appreciation

First of all, I would like to thank my supervisor, Professor Paul Verschure, and Dr. Anna Mura, for giving me this opportunity and for the enriching experience.

Thank you also to all the people I have collaborated with, for all the learning and memories. Special thanks to Dr. Brenda Jansen, for welcoming me in her lab and teaching me so much.

Thank you to my parents, my main cheerleaders, who have believed in me since I was merely breathing. Thank you for your love, your patience and for raising me with that integrity and values. *Os quiero mucho, bonitos*.

Thank you to my friends, the close ones and the ones from far, the ones I know since childhood and the newly discovered ones. You all have made possible this; with your support, with your time and kind words. Thanks for cheering me up and understanding me every time I said: “Sorry, I cannot join today. I need to work”.

Thank you to my colleagues, outstanding professionals who have become great friends. Thanks for the laughs, the *vermuts* and the late nights in the lab. I can say with no doubt that I have learnt something from each one of you. You guys are my fire.

Thank you to my lovely coincidence. Thank you for entering my life in strange times and deciding to stay, for your support, your hugs and your summer breeze in the cloudy days.

Thank you to all the participants and the people who collaborated in this research. Special thanks to the children, their parents, caregivers and teachers.

I would also like to thank Lydia Garcia, for all her help and kindness, and for being a sort of academic godmother fairy, helping grad students swim the bureaucratic seas.

I’d finally like to thank my family from supporting me from the distance; and special thanks again to my grandfather, wherever he is, for telling me I could become whatever I wanted (look, it seems I’m becoming a doctor!).

And thanks to you, whoever is reading this thesis, for your interest in this research.

Abstract in English

Students' metacognition, that is, the ability to accurately assess one's skill level and to update one's internal model of those skills, is a crucial educational factor, as it can help students self-regulate their learning. The purpose of this thesis is to assess how individuals' metacognitive abilities relate to learning processes and outcomes in STEM-related tasks. It also highlights how they relate to students' characteristics, as their gender. This thesis also presents the first steps towards a methodology to teach programming and robotics to elementary-school children and a study on digital technologies in museums to present historical content. Finally, it also explores prediction and collaborative behaviour in young adults in the autistic spectrum, and how this behaviour is self-perceived during a collaborative task with an artificial agent.

Abstract in Spanish

Las capacidades metacognitivas de los/as estudiantes (la habilidad de evaluar con precisión las capacidades de uno/a mismo/a y de actuar con nuestro modelo interno de estas capacidades) es un factor educativo crucial, ya que puede ayudar a los/as estudiantes a autorregular su proceso de aprendizaje. El objetivo de esta tesis es evaluar como las habilidades metacognitivas de los/as estudiantes se relacionan con su proceso de aprendizaje y resultados en tareas STEM (Ciencia, tecnología, ingeniería y matemáticas). También investiga cómo se relacionan estas capacidades con sus características, como el género. Esta tesis también presenta los primeros pasos para desarrollar una metodología para enseñar programación y robótica a niños/as de primaria y un estudio sobre tecnologías digitales en museos para mostrar contenido histórico. Finalmente, explora también procesos de predicción y colaboración en adultos jóvenes dentro del espectro autista y cómo perciben su comportamiento durante una actividad colaborativa con un agente sintético.

Abstract in Catalan

Les capacitats metacognitives dels/les estudiants (l'habilitat d'avaluar amb precisió els capacitats d'un mateix i d'actualitzar el nostre model intern d'aquestes capacitats) és un factor educatiu crucial, ja que pot ajudar els/les estudiants a autoregular el seu procés d'aprenentatge. L'objectiu d'aquesta tesi és avaluar com les habilitats metacognitives dels/les estudiants es relacionen amb el seu procés d'aprenentatge i resultat en tasques STEM (ciència, tecnologia, enginyeria i matemàtiques). També investiga com es relacionen amb les seves característiques, com el gènere. Aquesta tesi també presenta els primers passos per desenvolupar una metodologia per ensenyar programació i robòtica a nens/es de primària i un estudi sobre tecnologies digitals en museus per mostrar contingut històric. Finalment, també explora processos de predicció i col·laboració en adults joves dintre de l'espectre autista i com perceben seu comportament durant una activitat col·laborativa amb un agent sintètic.

Table of contents

Abstract in English .... ***x***

Abstract in Spanish .... ***x***

Abstract in Catalan .... ***xi***

List of figures .... ***xxii***

List of tables .... ***xxvii***

1 INTRODUCTION TO THIS THESIS ***3***

1.1 Problem statement **3**

a) Seven principles of human learning 4

1.2 Purpose of this thesis and Research Questions **6**

1.3 Methodology used **9**

a) Self-reported questionnaires 9

b) Computer logfiles 10

c) Monitoring judgment tasks 10

d) Interviews 10

1.4 Main publications **11**

1.5 Other publications **11**

1.6 Thesis outline **13**

2 GENERAL THEORETICAL FRAMEWORK ***15***

2.1. Introduction **15**

2.2. Learning Theories **15**

a) Piaget's Constructivism: Mental schemes 16

b) Vygotsky's Zone of Proximal Development 17

c) Dewey, Kolb and Experiential Learning 18

d) Inquiry-Based Learning 19

e) Constructionism 20

2.3. Metacognition **21**

a) Models of metacognition 22

b) Confidence 27

c) Metacognition and Self-Regulated Learning 29

d) Gender in Metacognition 30

e) Implications in pedagogy 33

2.4. Mathematical Learning **35**
a) Metacognition in Mathematics 35
b) Balance-Beam Task 36

2.5. Learning Technologies **41**
a) Individualized Learning and Intelligent Tutoring Systems 41

2.6. The Distributed Adaptive Control (DAC) Architecture **43**
a) DAC as a pedagogical model 45

3. STUDY 1: Students' difficulties in recognizing their competence level ***47***

3.1 Introduction **47**
a) The original study 47

Metacognition **48**
a) Confidence 49
b) The importance of assessing confidence in education 50
c) Metacognitive differences between genders 51
d) Hypotheses 51

3.2 Materials and Methods **52**
a) Overview 52
b) Participants 52
c) The test 53
d) Experimental Protocol 53
e) Measures 54

3.3 Results **54**
a) General results 55
b) Results on gender 56

3.4 Discussion **62**
a) General discussion 63
b) Discussion on gender differences 64

3.5 Conclusion **64**
a) Strengths and limitations of the study 65

4. STUDY 2: How confident are you? A study on feedback and gender ***69***

4.1 Introduction **69**
a) The role of confidence, feedback and gender in metacognition 70

4.2 Methods **74**
a) The Balance Scale task 75
b) Experimental Procedure 76
c) Measures 79

4.3 Results 80
a) Results on performance 81
b) Which is the relationship between participants' metacognitive skills and their performance? 83
c) How does metacognition and performance during the task (while receiving feedback) relate to initial metacognition and performance (when no feedback is provided)? 83
d) How does the relationship between confidence and performance differ between genders? 85
e) Confidence differences between correct and incorrect answers. Gender differences in confidence for wrong answers but not for correct ones 87

4.4 Discussion 89
a) Other variables to explore: Gender 90

4.5 Conclusions 92

5. STUDY 3: Metacognitive factors behind rule change in the Balance Scale task 93

5.1 Introduction 93
a) Metacognition in Mathematics 95
b) Exploration 95
c) Active/Passive learning 96

5.2 Metacognitive and motivational factors 97
a) Mathematical anxiety 97
b) Perceived competence 98
c) Goal orientation 99
d) Intolerance of uncertainty 100
e) Hypotheses 101

5.3 Materials and Methods 102
a) Participants 102
b) Conditions 102
c) Experimental Protocol 104
d) Measures and Tools 106

5.4 Results 107
a) Results on confidence 107
b) Results on gender differences 110
c) Results on exploration 113

6. STUDY 4: Virtual Reality for Historical and Cultural Learning 121

6.1 Introduction 121
a) Recommender systems for HCL 122
b) Virtual and Augmented Reality for HCL 122

c) Visitors' experience in HCL 123
d) Active Learning for HCL 124
e) The FutureMemory learning paradigms 125
e) The FutureMemory application 127

6.2 Materials and Methods 130
a) The Hollandsche Schouwburg Memorial 130
b) Participants 131
c) Experimental Protocol 132

6.3 Results 133
a) Questionnaire results 133
b) Content navigation results 140

6.4 Discussion 142

7. STUDY 5: A methodology to teach programming and robotics to children 145

7.1 Introduction 145
a) Educational Robotics 147
b) Computational Thinking 150
c) Metacognition in coding and educational robotics 152
d) Metacognition and collaboration 153

7.2 The Framework 154
a) Tools 154
b) ASSURE Methodology 155
c) "CREA" - A methodology to teach programming to children 157

7.3 Materials and Methods 158
a) Overview and Participants 158
b) Experimental Protocol 159

7.4 Results 160

7.5 Discussion and Conclusion 162

8. STUDY 6: Prediction and Collaboration in Autism Spectrum Disorder 165

8.1 Introduction 165
a) Autism 165
b) Metacognition in autism 166
c) Sensorimotor integration 167
d) Prediction and collaboration 169
e) Hypotheses 170

8.2 The Scenario 171
a) The Task 171

b) The Point of Social Subjective Equality 172

8.3 Materials and Methods 175

a) Participants 175
b) Experimental Protocol 176
c) Measures 177
d) Tools 178

8.4 Results 179

a) Participants in the autism spectrum showed more variable behaviour than neurotypicals 180
b) Participants in the autism spectrum showed slower adaptation to the artificial agent than neurotypical ones 181
c) Participants in the autism spectrum showed less adaptation to the other agent when the task became more uncertain 183
d) No differences in perception of the task between groups, only by perceiving the other agent as human or synthetic 184
e) Participants in the ASD group focused on movement to describe the other agent 187

8.5 Discussion 188

a) Discussion on differences in behaviour 189
b) Discussion on differences in perception 190
c) General discussion 191

9. IMPLICATIONS FOR A PEDAGOGICAL MODEL 193

9.1 Introduction 193

9.2 Existing theories 194

a) Hegel’s Dialectic model 194
b) Peirce’s theory of Abductive Reasoning 195
c) Piaget’s theory on Cognitive Disequilibrium 196
d) Buch’s Conscious Competence Theory 197
e) Zimmerman’s Self-Regulated Learning models............................ 197
f) Boekaerts’s Dual Processing Model 198

9.3 Phases of Learning 199

a) Resistance: Content and Self 200
b) Confusion 201
c) Abduction 202
d) Utilisation: Performance and Mastery 202

9.4 Behavioural outcomes 203

a) Expected behavioural outcomes 203
b) Examples in this thesis 205

9.5 Limitations and Further Steps 211

10. GENERAL DISCUSSION AND CONCLUSION 213

10.1 Introduction 213

10.2 How this thesis answers the research questions 214

10.3 How this thesis supports the presented learning principles ... 216

10.4 Conclusions on metacognition... 219
a) Implications 222
b) Strengths & Limitations 222
c) Further steps.. 223

10.5 Conclusions on individual differences in metacognition......... 225
a) Gender.. 225
b) Neurodivergence.. 227
c) Implications 228
d) Further steps.. 230

10.6 Conclusions on educational applications 230
a) Methodologies to teach programming and robotics.................... 231
b) Implications 232
c) Recommender Systems for Museums.. 234
d) Considering learners' metacognitive abilities when designing Intelligent Tutoring Systems 235
e) Further steps.. 236

10.7 Contribution of this thesis 236

10.8 General future research directions 238

10.9 General Conclusions 239

APPENDIX 241

Test from study in Chapter 3 242

Tests from study in Chapter 4 251
Pre-test 251
Post-test 259

Scales and tests used in Chapter 5 266
The intolerance of uncertainty scale for children 267
Modified Abbreviated Math Anxiety Scale 275
2 x 2 Achievement Goal Framework.. 276
Perceived Competence Scale for Children.. 277
Tempo Test Automatizeren (TTA) 280

Questionnaires from study in Chapter 6 283
Pre-test 283
Post-test 285

Tests and Interview from study in Chapter 7................................. 290

Pre-test 290
Post-test 295
Interview 296

Questionnaires and Interview from study in Chapter 8 297
Questionnaire before the task 297
Questionnaire during the task 298
Questionnaire after the task 299
Interview after the task 300

BIBLIOGRAPHY 302

List of figures

Figure 1. Kolb's experiential learning cycle 19

Figure 2. Rules explaining the behaviour during the task............. 40

Figure 3. DAC architecture 44

Figure 4. Comparison between performance and confidence. 56

Figure 5. Differences between genders in performance and confidence, overall and divided by performance quartiles. 57

Figure 6. Differences between genders in performance by year and its relation to the class' gender ratio. 58

Figure 7. Differences in confidence reports between genders and correct and wrong answers. 59

Figure 8. Participants' performance and confidence by question type. 61

Figure 9. Representation of the task rules, example and experimental protocol.. 77

Figure 10. General results on performance: Improvement between pre- and pos-tests and error distribution among trials. 82

Figure 11. Bias evolution during the experiment and its relation to confidence 84

Figure 12. Differences between genders in performance, confidence and bias.. 86

Figure 13. Differences in confidence reports between genders and between correct and wrong answers 88

Figure 14. Other measures of metacognition............................... 89

Figure 15. Experimental protocol. 105

Figure 16. Blocks of the computerized Balance Scale task......... 106

Figure 17. Correlation between performance and confidence in pre- and post-tests. 108

Figure 18. Differences between genders in the correlation between performance and confidence in pre- and post-tests. 109

Figure 19. Correlations between bias in pre-test and the four dimensions of the AGF scale. 110

Figure 20. Differences between genders in mathematical anxiety and perceived competence in cognitive, self, and mathematics.......... 111

Figure 21. Differences between genders in the avoidant profiles of the AGF. 112

Figure 22. Proportion between the three groups in choosing a lower (blue), the same (orange) or one rule more (green) than the one they used in the pre-test. 114

Figure 23. Result of the 2-class model representing the probability patterns in the pre-test. 116

Figure 24. Result of the 3-class model representing the probability patterns in the post-test... 117

Figure 25. Degree of exploration by item type and corectness in the second exploration phase. 118

Figure 26. Degree of exploration by item type and correctness in the first exploration phase. 118

Figure 27. Classification of participants' exploration profiles during the exploration blocks. 120

Figure 28. Setup of the experience... 128

Figure 29. The Augmented Reality effect. 128

Figure 30. The Hollandsche Schouwburg Memorial. 131

Figure 31. Visual summary of the demographical data 132

Figure 32. Experimental protocol. 133

Figure 33. Results on familiarity with the topic and interest 134

Figure 34. Comparison between self-reported feelings before and after interacting with the application. 135

Figure 35. Visitors’ perception of the application presentation. . 136

Figure 36. Visitors' perception of the experience 137

Figure 37. Visitors' perception of the easiness of use 138

Figure 38. Visitors' perception of the provided information. 139

Figure 39. Visited content. 140

Figure 40. Content visited inside each exhibit........................... 141

Figure 41. Technological devices used in CREA. 155

Figure 42. Examples of the six sessions of CREA 158

Figure 43. Examples of the procedure carried out during the interview.. 159

Figure 44. Improvement in programming and robotics knowledge after the intervention 161

Figure 45. Children working on their projects........................... 161

Figure 46. Example of the activity on peripherals observation. .. 164

Figure 47. Example of a trial during the task. 171

Figure 48. Probability to go depending on the relative distance to the target and PSSE.. 173

Figure 49. Representation of the curve of the probability to intercept the target based on the target's location for each of the three proposed agents... 175

Figure 50. Representation of the experimental protocol. 177

Figure 51. Representation of the setup used in the task. 179

Figure 52. PSSE intersections from the TD and ASD groups..... 180

Figure 53. Differences between groups in the difference in shift per agent. 182

Figure 54. Matrix representing the relationship between Phase 2 (actual) and Phase 3 (predicted) behaviour. 184

Figure 55. Differences between agents and groups in perceived agent predictability, target predictability, engagement, and collaboration. 185

Figure 56. Differences in perceived collaboration between the participants that perceived the AI agent as synthetic or as a human player. 186

Figure 57. Frequency of characteristics commented about during the interviews. 187

Figure 58. Hegel's Dialectic model 195

Figure 59. Piaget's Cognitive Disequilibrium model 196

Figure 61. DAC's phases of learning 200

Figure 62. Perceived logical reasoning ability as a function of actual test performance.. 206

Figure 63. DAC’s phases of learning mapped to the evolution of the bias measurement for the whole experiment 207

List of tables

Table 1. Summary of the metacognition models presented.......... 26

Table 2. Comparison between studies.. 54

Table 3. Participants' profile by metacognitive classification....... 60

Table 4. Comparison between the results of the original study and this one. 62

Table 5. Siegler's Rule models and their response patterns by item type.. 94

Table 6. Experimental conditions.. 104

Table 7. Comparison of the five models generated by the LCA for the rules in the pre-test. 115

Table 8. Comparison of the five models generated by the LCA for the rules in the post-test.. 116

Table 9. Profiles created from merging the four simple ones...... 119

Table 10. Children's perception of the different sessions. 162

Table 11. Learners' expected behaviour leading to remaining or moving in each learning phase 205

Table 12. Learners' classification depending on their testing of previous items.. 208

1 INTRODUCTION TO THIS THESIS

1.1 Problem statement

A pivotal part of any students' learning process is knowing when they have already mastered the topic at issue and can move to the following one, or when they still do not know it well and need to continue working on it. For students to do so, they need to be aware of what they know and what they do not know (yet). This ability is crucial for both the academic sector and everyday life, as it involves self-reflection on one's current state, future goals, potential actions and strategies to take, and the evaluation of their results.

The improvement of students' metacognitive skills, that is, the ability to accurately assess one's skill level and to update one's internal model of those skills, has been characterised as an important facet on the way to advance current educational paradigms (Bransford, Brown, & Cocking, 2000). This is not surprising: being aware of one's errors promotes a deeper information processing (Vanlehn, Siler, Murray, Yamauchi, & Baggett, 2003) and active engagement in updating (incorrect) mental models (Chi, 2000). Furthermore, poor metacognition could lead to an under- or over-use of the help provided to the learner. On the one hand, students needing more help are the ones less prone to ask for it (Karabenick & Knapp, 1988; Puustinen, 1998; Ryan, Gheen, & Midgley, 1998). On the other hand, some students may ask for support, even if they have enough knowledge to solve the task by themselves, to facilitate their work without trying to understand the relation between the question and the answer (Aleven & Koedinger, 2000). As a result, we face a vicious circle: students need to improve their metacognition, but they do not possess appropriate metacognitive abilities to self-regulate their learning process.

The reason to focus on Science Technology Engineering and Mathematics (STEM) is that is less affected by language than other subjects (making its results more useful for the international scene) because there are strong perceptions and stereotypes about it (being "bad at maths"). The studies presented in this thesis cover a broad range of ages, as there are three studies with children (chapters 4, 5 and 7), one with teenagers and young adults (chapter 8) and two with adults (chapters 3 and 6).

a) Seven principles of human learning

Indeed, metacognition is one of the seven principles of human learning presented in (Council, 2002), a report created to advance the current educational paradigm. This report, based on previous results of educational psychology, presents seven principles that drive learning in humans. These principles are:

Principle 1: *Principled Conceptual Knowledge*

This principle accounts for the importance of structuring new and existing knowledge around the central concepts of the discipline. This knowledge should be both accessible and usable. Expert's content knowledge is usually structured around the major concepts of the specific domain, as already defended by John Dewey in the early 20th century (Dewey, 1902).

Principle 2: *Prior knowledge*

This principle postulates that learners use their previous knowledge to construct new, by linking the new one to what they already know. This process implies adding, modifying, or reorganising existing knowledge or skills. Moreover, when facing new content, students do not only already possess knowledge but also beliefs and misconceptions about that topic, which can significantly affect how they approach new learning (Wandersee, Mintzes J, & Novak D, 1993).

Principle 3: *Metacognition*

This principle defends the use of metacognitive strategies to identify, monitor and regulate cognitive processes. Learners need to assess what they already know and what else they need to know in any given situation to be effective problem solvers. To do so, they have to consider both factual (about the task, their skills, and the goals) and strategic knowledge (about how and when to use a specific procedure to solve the problem). Considering learners' metacognitive abilities to increase one's understanding of their self-regulation processes was already proposed by researchers like Efklides (Efklides, 2006).

Principle 4: *Differences among learners*

This principle highlights the importance of considering students' different abilities, strategies and approaches when facing a learning task, as previously shown in research (Alhajri, Alhunaiyyan, & AlMousa, 2017). There can be substantial differences in cognitive abilities between learners, even among individuals of the same age. Moreover, these differences can also relate to emotional and motivational characteristics, as we will see in the next principle.

Principle 5: *Motivation*

This principle accounts for the effect of learners' motivation to learn on their learning process. Regardless of this motivation being either extrinsic (performance-oriented) or intrinsic (learning-oriented), it can affect students' willingness to persist in a task, affecting the learners' results. For example, learners' beliefs about their abilities in a subject area strongly relate to their success in learning about that domain. Moreover, those results can also be affected by learners' perception of their abilities as predetermined or as substantially affected by their effort (Dweck & Henderson, 1989).

Principle 6: *Situated learning*

This principle defends that knowledge is contextually situated and shaped by the context and the activity (Lave & Wenger, 1991). Knowledge has thus to be presented in authentic contexts, that is, settings and situations that would usually involve that knowledge. Moreover, the situated learning theory also defends the idea of learning by doing, where learners can observe the implications of their knowledge.

Principle 7: *Learning communities*

In similar lines to the previous one, this principle highlights the importance of socially supported interactions to enhance students' learning process (Smith, MacGregor, Gabelnick, & Matthews, 2004). Providing learners with the opportunity to interact and collaborate with others allows them to discuss their ideas with others and to learn from observing others. Moreover, social interaction is essential for students' metacognitive skills and the formation of their sense of self. Meaningful conversations do not consist only of discussing facts or procedures; but also, of applying ideas and raising questions.

The principles introduced here will be present, to a great or lesser extent, in the research questions and studies of this thesis. Chapter 10 (Conclusions) will explain how each study assessed each of the principles.

1.2 Purpose of this thesis and Research Questions

The primary purpose of this thesis is to understand learning further and more precisely, metacognition, following the principles of learning. As metacognition is linked to self-perception, we wish to understand how learners perceive their performance in a given task. Can learners understand if they are performing correctly? This leads to the first research question:

RQ1. ***How does actual and perceived performance relate to each other when participants evaluate their answers?***

To answer this question, we performed a study where participants reported the perceived correctness of their answers (Chapter 3 – *STUDY 1: Students' difficulties in recognizing their competence level*). This study indicated that there was a mismatch between actual and perceived performance, and participants generally overestimated their competencies. As this assessment seemed a good indicator that reflected an individual's metacognitive abilities, we maintained this metric in the following studies (for example the studies in Chapter 4 – *STUDY 2: How confident are you? A study on feedback and gender*, and Chapter 5 – *STUDY 3: Metacognitive factors behind rule change in the Balance Scale task*).

The outcomes of Chapter 3, along with literature, highlighted the role of gender in this discrepancy between actual and perceived performance. Given that there are differences between genders in error-monitoring (Beyer, 2015), we wished to examine the role of gender and feedback in a learning task, leading to the second research question:

RQ2. ***How do those differences between perceived and actual performance differ depending on students' gender and the feedback received?***

To answer this question, we performed a study where participants partially received feedback in an educational task, and we dedicated part of the analysis in the gender differences. This study highlighted the role of gender and feedback in metacognition, as we found differenced in confidence between genders but not in performance. We consequently included gender and feedback as an assessment in the following studies (for example, the study in Chapter 5 – *STUDY 3: Metacognitive factors behind rule change in the Balance Scale task*).

Thus far, we have identified a relation between metacognition, gender and feedback. We wished to further explore other metacognitive and motivational variables that could potentially affect the learning process. For example, exploration has been shown to be beneficial for learning (L. Schulz, 2012), leading to the third research question:

RQ3. ***Which other distal variables would be related to students' metacognitive abilities and behaviour in an educational task? And how do they relate to children's exploration in a free answer task?***

The study in Chapter 5 (*STUDY 3: Metacognitive factors behind rule change in the Balance Scale task*) was created to answer that question. Here, we studied how children's traits (like intolerance for uncertainty, mathematical anxiety, perceived competence and goal-directedness) related to their behaviour during an educational task and the discrepancy between perceived and actual competence. Results suggested exploration positively affected the learning process. The fourth research question further evaluates the benefits of free exploration:

RQ4. ***How do users navigate through the content of a virtual educational experience based on exploratory learning?***

To answer this question, we created a real-life educational scenario based on a Virtual-Reality (VR) experience providing content about a well-known historical event, the Holocaust (Chapter 6 – *STUDY 4: Virtual Reality for Historical and Cultural Learning*). In this study, users mostly explored content related to personal stories and seemed to change their emotional state towards more positive and empathetic emotions.

The outcomes of these studies have identified variables that affect metacognition and consequently learning. Another aim of this thesis is to implement an exploratory and metacognition-based practice in the educational paradigm, where the learning task capitalises on the concepts we examined this far. This leads to the fifth research question:

RQ5. ***How can we create a methodology to teach programming, in a discovery-oriented way?***

In Chapter 7 (*STUDY 5: A methodology to teach programming and robotics to children*), we created a collaborative methodology, based on Computational Thinking, focused on visual representations, verbal explanations and self-discovery. Children seemed to enjoy more educational activities that involved crafting, creativity, the usage of computers, as well as those that promoted teamwork.

The studies so far evaluated learning and self-perception in neurotypical individuals. Nevertheless, considering the importance of collaborative learning and the increase of inclusive programs in educational centres, we saw the need of assessing collaboration and self-perception in neurodivergent individuals. Here, we focused on autistic spectrum disorder, given its relation to social deficits (substantial in collaborative learning) and the increase in new diagnostics (Avlund, Thomsen, Schendel, Jørgensen, & Clausen, 2020).

RQ6. ***How do behaviour and self-assessment differ in autism during a collaborative task?***

The study in Chapter 8 (*STUDY 6: Collaboration variability in Autism Spectrum Disorder*) was created to answer that question. We found differences in predictive and collaborative behaviour between neurotypical participants and participants in the autistic spectrum, which were not reflected in the self-reports.

1.3 Methodology used

The research presented in this thesis uses behavioural studies to answer the proposed research questions. To evaluate the metacognitive abilities of learners, our studies involve tasks that cover quite broad educational subjects ranging from STEM (more specifically math-related tasks and technology by programming and building robots) to history. The STEM subjects mainly focused on math-related tasks and technology, by programming and building robots. We selected these tasks because they are less affected by language, compared to other subjects, and the outcomes of the studies can be useful internationally. Furthermore, we focused on these subjects because of the strong perceptions and stereotypes around them (like being “bad at maths”), and gender plays a key role in these stereotypes. The studies presented in this thesis cover a broad range of ages including children (chapters 4, 5 and 7), teenagers and young adults (chapter 8) and adults (chapters 3 and 6). The tasks are mainly digitised to allow for automatic retrieval of information. Tasks are adaptive to the user based on the performance of the user

The main methodology used in this thesis follows a quantitative research approach, using several methodologies: self-reported questionnaires, computer logfiles and monitoring judgment tasks. Moreover, two of the studies presented also take a qualitative approach by using interviews. These measures can be categorised as online or off-line. Online ones refer to those measures obtained while the participant is performing the task; they are, thus, always connected to a specific task. Contrarily, offline measures are obtained outside of the process of the task and may not be related to it.

a) Self-reported questionnaires

Self-reported questionnaires are an off-line method, commonly using rating scales. They can be presented in paper or electronic form and can be completed by a larger sample than one-to-one methods. One of the limitations of this type of measure is that participants may be limited to the range of possible answers provided by the experimenter. For that reason, some of our self-report questionnaires also contain open-ended questions.

b) Computer logfiles

Advances in technology have made it possible to use computer logs to obtain online measures automatically and unobtrusively. Like this, participants' behaviour can be automatically coded by the computer (or any other device, as tablets) in a logfile. As a consequence, it is a very efficient method, as there is no need for the researcher to carry out one-on-one testing with the participants. Furthermore, the level of detailed allowed by computer logfiles makes it a very fine-grained method.

c) Monitoring judgment tasks

Monitoring judgement tasks are online measures of metacognition that can be part of a learning task without interrupting it much. Here, the researcher adds self-monitoring questions inside of a task that participants have to answer.

When asking a participant to report their perceived performance, one must consider that the measure contains a sense of magnitude. That is why some authors (Lichtenstein & Fischhoff, 1977) defend the use of an explicit estimate of the probability of being correct rather than verbal confidence reports. Nevertheless, the decision of which kind of report to use would depend on the nature of the task and the purpose of the self-report.

d) Interviews

This work uses open-ended, structure interviews. The reason behind this selection is that open-ended interviews allow participants to use their own words to describe their thoughts and understanding (Pintrich, Wolters, & Baxter, 2000). In contrast, structured interviews allow researchers to probe participants for clarifications or more elaboration in their answers (Groves et al., 2009). Moreover, when the person being interviewed is a child, interviews come in hand by eliminating participants' reading ability as a confounding variable. As a disadvantage, interviews are more time-consuming in both deployment and analysis and can fall into relying on memory processes, as participants have to remember the task to answer about it.

1.4 Main publications

Study 1. Blancas, M., Zucca, R., Maffei, G., & Verschure, P. F. M. J. (2020) *Students' difficulties in recognising their competence level: A revision of the Unskilled and Unaware effect.* In Metacognition and Learning (Submitted)

Study 2. Blancas, M., Vouloutsi, V., Zucca, R., & Verschure, P. F. M. J. (2020). *Learning phases as expressed in the performance of primary-school children in a mathematical task*. In Contemporary Educational Psychology (Submitted)

Study 3. Blancas, M., Jansen, B., & Verschure, P. F. M. J. (2020). *Metacognitive factors behind rule change in the Balance Scale task* (In preparation)

Study 4. Blancas, M., Wierenga, S., Pacheco, D., & Verschure, P. F. M. J. *Virtual Reality for Historical and Cultural Learning: A User Study at the Hollandsche Schouwburg Memorial.* In *EuroVR 2020* (Accepted)

Study 5. Blancas, M., Valero, C., Mura, A., Vouloutsi, V., & Verschure, P. F. M. J. (2019). "CREA": *An inquiry-based methodology to teach robotics to children*. In 10th International Conference on Robotics in Education (Accepted)

Study 6. Blancas, M., Maffei, G., Vouloutsi, V., Sánchez-Fibla, M., & Verschure, P. F. M. J. (2020) *Collaboration variability in Autism Spectrum Disorder*. In Frontiers in Human Neuroscience (Published)

1.5 Other publications

Blancas, M., Wierenga, S., López, H., & Verschure, P. F. M. J. (2020). *Active Learning in Digital Heritage*. In Grace, V. (Eds.), *Digital Holocaust Memory, Research and Education* (In preparation)

Blancas, M., Valero, C., Vouloutsi, V., Mura, A., & Verschure, P. F. M. J. (2020). *Educational Robotics: A journey, not a destination*. In Papadakis, S. & Kalogiannakis, M. (Eds.), *Using Educational Robotics to Facilitate Student Learning* (Accepted)

Blancas, M., Jansen, B., & Verschure, P. F. M. J. (2020). *Metacognitive and motivational factors behind behaviour in a mathematical task*. In 17th International Conference on Motivation and the 9th Conference on Metacognition

Jansen, B., Blancas, M., & Verschure, P. F. M. J. (2020). *How children's math anxiety relates to exploration and learning on a mathematical task*. In 17th International Conference on Motivation and the 9th Conference on Metacognition

Blancas, M., Maffei, G., Vouloutsi, V., Sánchez-Fibla, M., & Verschure, P. F. M. J. (2018). *Measuring collaboration in a predictive game task: a comparison between autistic and neurotypical individuals using the Point of Social Subjective Equality*. In Social cognition in humans and robots

Low, S. C., Blancas, M., Maffei, G., & Verschure, P. F. M. J. (2018). *Focusing Attention on the Mechanisms of Autism.* In Social cognition in humans and robots

Blancas, M., Vouloutsi, V., Zucca, R., Mura, A., & Verschure, P. F. M. J. (2018). *Hints vs Distractions in Intelligent Tutoring Systems: Looking for the proper type of help.* arXiv preprint arXiv:1806.07806

Blancas, M., Vouloutsi, V., Fernando, S., Sánchez-Fibla, M., Zucca, R., Prescott, T. J.,... & Verschure, P. F. M. J. (2017). *Analysing children's expectations from robotic companions in educational settings.* In Humanoid Robotics (Humanoids), 2017 IEEE-RAS 17th International Conference on (pp. 749-755). IEEE

Vouloutsi, V., Blancas, M., Zucca, R., Verschure, P. F. M. J. (2017). *Studying the adaptation of robot's strategies in an educational task.* In Child-Robot Interaction Workshop at HRI 2017

Blancas, M., Vouloutsi, V., Zucca, R., & Verschure, P. F. M. J. (2016). *Considering students' confidence when building a synthetic* *tutoring system.* In conference New Friends 2016.

Reidsma, D., Charisi, V., Davison, D., Wijnen, F., van der Meij, J., Evers, V.,... & Mazzei, D. (2016) *The EASEL project: towards educational human-robot symbiotic interaction*. In Conference on Biomimetic and Biohybrid Systems (pp. 297-306). Springer, Cham.

Blancas, M., Zucca, R., Vouloutsi, V., & Verschure, P. F. M. J. (2016). *Modulating Learning Through Expectation in a Simulated Robotic Setup.* In Conference on Biomimetic and Biohybrid Systems (pp. 400-408). Springer, Cham.

Vouloutsi, V., Blancas, M., Zucca, R., Omedas, P., Reidsma, D., Davison, D.,... & Cameron, D. (2016). *Towards a synthetic tutor assistant: the EASEL project and its architecture.* In Conference on Biomimetic and Biohybrid Systems (pp. 353-364). Springer, Cham.

Blancas, M., Vouloutsi, V., Grechuta, K., & Verschure, P. F. M. J. (2015). *Effects of the robot's role on human-robot interaction in an educational scenario.* In Conference on Biomimetic and Biohybrid Systems (pp. 391-402). Springer, Cham.

Lallée, S., Vouloutsi, V., Blancas, M., Grechuta, K., Llobet, J. Y. P., Sardà, M., & Verschure, P. F. M. J. (2015). *Towards the synthetic self: making others perceive me as another.* Paladyn, Journal of Behavioural Robotics, 6(1).

1.6 Thesis outline

This thesis contains ten chapters. The first chapter is a general introduction. Chapter two presents a theoretical framework of learning theories and metacognition. Six empirical studies were conducted in order to answer the research questions formulated in the introduction. Studies 1, 2 and 3 (chapters 3, 4 and 5, respectively) analyse the relationship between students' performance and their metacognitive abilities (mainly, assessed through confidence reports) in science-related tasks. Finally, the fourth and fifth studies (chapters 6 and 7) present the results of two educational technologies approaches. Chapter 6 presents a user study on a digital experience to teach about the history of the Holocaust, deployed in a museum; while Chapter 7 presents a methodology to teach programming and robotics to elementary-school children; while. Finally, the sixth study (chapter 8) explores prediction, collaboration, and self-evaluation in individuals in the autistic spectrum, and compares it with neurotypicals. Chapter 9 presents the implications for a pedagogical model. Finally, the last chapter presents the general discussion and conclusions of this thesis.

2 GENERAL THEORETICAL FRAMEWORK

2.1. Introduction

This section presents the literature relevant to this thesis that is related to pedagogical theories and metacognition. As this thesis mainly focuses on learning, we first present the learning theories that are relevant not only to this work but also any learning tasks and environments, like Piaget's Constructivism, Vygotsky's Zone of Proximal Development Theory or Dewey's Experiential Learning. This thesis mainly focuses on metacognition, and consequently, we also present models of metacognition, like Flavell's, Schraw's and Ekflides'. Furthermore, we provide information about mathematical learning and the Balance Scale task, as this is an educational task used in this thesis. Finally, we provide a section regarding learning technologies, as the main findings of this thesis can be applied to the creation of adaptive technologies that aim at bootstrapping and improving knowledge acquisition. The information more specific to each study (as, for example, the one about collaboration in autism or about learning technologies for digital heritage) appear in the chapter related to the specific study.

2.2. Learning Theories

As we are interested in learning, we need to understand how humans acquire knowledge. For that reason, in this section, we present several learning theories. All of them focus on an experience-based kind of learning, where learners are active agents rather than passive ones. Many of the theories study how learning occurs through interaction with the world. In the next sections, we will present the learning theories used to scaffold this work: Piaget's Constructivism and its idea how of children learn from the world by building mental models of it through experience; Vygotsky's Zone of Proximal Development, and how supporting children's reasoning can help them move from what they know to what they do not know (yet);

Experiential Learning, and its cyclical process of doing, observing thinking and planning.

a) Piaget's Constructivism: Mental schemes

For constructivists, learners are not empty vessels to be filled with knowledge, but active agents in their learning (Ackermann, 2001). Constructivism sees learners as constructing their knowledge by interacting with the world (Piaget, 1977). Jean Piaget, a well-known developmental psychologist and main precursor of this theory, viewed learning as an active process where learners are constantly recollecting and reorganising information in their brain, creating meaning out of their own experiences (Piaget, 1952). He defined the mental maps and organisations obtained from these experiences as "cognitive schemes". These schemes are not static; they grow and change over time through experience. For Piaget, development is the increase in the number and complexity of a person's schemata.

According to Piaget, growing is a process of adaptation to the world, which happens through three steps: *Assimilation*, *accommodation* and *equilibrium*. When an agent's existing schemas can explain the world that the agent is facing, the learner is in what Piaget called a state of *equilibrium* (that is, a state of cognitive balance). Nevertheless, when a person faces new information, two things can happen: they can match their mental schemes or not. On the one side, if this information can fit their mental schemes, the agent would go through a process of *assimilation* by using their existing schemes to deal with the information. On the other side, if it does not fit their mental schemes, the agent would go through a process of *accommodation* by updating their existing schemes to embrace the new information (Wadsworth, 1996).

Piaget defended that the purpose of these forms of adaptation was to obtain balance, what he defined as *equilibrium,* between the mental schemata of the individual and the world that surrounds them (which, when they crash, cause to the learner a state of *disequilibrium)*. This equilibration is what drives the learning process of students, as they avoid frustration by trying to restore the equilibrium by mastering the new challenge.

In an educational context, there are moments when the learner is not able to advance in the learning process on their own and needs some kind of external guidance. Constructivism emphasizes the guidance of the learner. This theory views the teacher as someone who should not give correct solutions, but instead provide hints and directions to help the learners find solutions themselves. The role of a constructivist teacher is closer to a guide than to the classical teacher (Steffe & Gale, 2012), which enhances the importance of engagement during the class. In Constructivism, learners actively construct knowledge by anchoring new information to their previous knowledge (Ausubel, 1978). It matches problems with multiple correct solutions, with no criteria put forward the solution.

b) Vygotsky's Zone of Proximal Development

Social Constructivism, a part of Constructivism related to communication, stated the importance of language and culture in learning. Lev Vygotsky, the main theorist of Social Constructivism, proposed what is known as the Zone of Proximal Development (ZPD) theory. The ZPD is the figurative space between what learners know and what they could know with the help of a more knowledgeable agent, being a teacher or a peer (Vygotsky & Cole, 1978). The ZPD is the distance between a person's actual developmental level as determined through independent problem solving and their level of potential development as determined by problem-solving under guidance or in collaboration with more capable peers.

To stay in the Zone of Proximal Development while carrying out an educational activity is one of the ways to make an activity engaging to the students (Shernoff, 2003). Such an approach provides learners with enjoyment and challenges without making the activity too difficult, which would result in loss of confidence in their control and abilities (Deshmukh et al., 2013). Providing learners with support specifically tailored to each of them, by incrementally increasing the level of complexity is called “scaffolding” (Bruner, 1978). Sometimes, a human or a computer-based tutor scaffolds a learner to achieve the understanding of a topic and so to cross the Zone of

Proximal Development (Vygotsky & Cole, 1978) through the use of cognitive tools, technologies for knowledge construction used by learners to interact with a task and think (Kim & Reeves, 2007).

Similar to Vygotsky's ZPD, Mercer and colleagues proposed the Intermental Development Zone (IDZ) (Mercer, 2000). This theory follows a similar idea to the one proposed by Vygotsky; however, in this "zone" the student exploring a dynamic space of improvement. Thus, the IDZ zone updates in parallel with the students' learning process. It follows a collaborative approach, were the learner and the teacher are in constant interaction to understand the student's ongoing capabilities and adapt the pace and content accordingly.

c) Dewey, Kolb and Experiential Learning

John Dewey (Dewey, 1938) articulated the value of experience and reflection in education. His work highlighted the importance of developing one's opinions of concepts by interacting with the information. He defended that, by viewing learning as a process only based on the transmission of content, students lose opportunities to develop their own opinions. Consequentially, not all learners will arrive at the very same conclusions, as learning is individualised based on past experiences. Thus, the experiential learning classroom mimics society, where all individuals interact and can have different views on topics and information.

The principle of learning through experience was also defended by the Experiential Learning Theory (ELT) (Kolb, 1984). This theory consists of four phases: *Concrete Experience*; *Reflective Observation*; *Abstract Conceptualisation;* and *Active Experimentation* (***Figure 1***). In the *Concrete Experience* phase, the individual is assigned a task, and learning occurs by feeling and taking part in an experience. In the *Reflective Observation* phase, that is, learning by reflection, learners review what they have done and learned. In the *Abstract Conceptualisation* phase, that is, learning by thinking, learners interpret what has happened and relate it with what they already know. Finally, in *Active Experimentation*, which is learning by doing, they put this learning in practice.

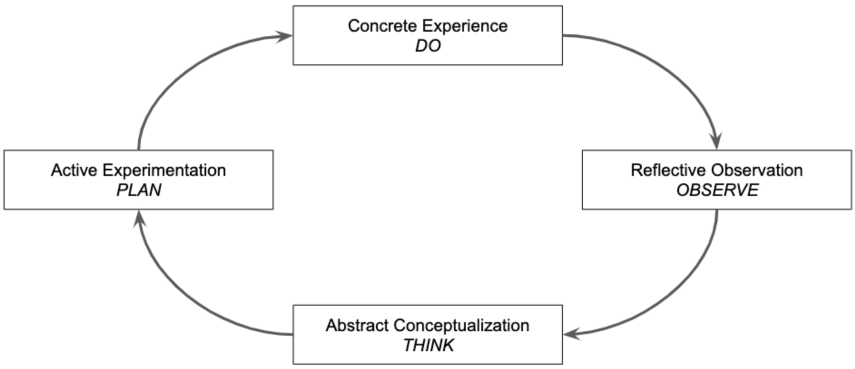


Figure 1. Kolb's experiential learning cycle (adapted from Healey and Jenkins, 2007).

Four types of learning can be defined as a result of combining these previous steps (Kolb & Kolb, 2005). The learner can diverge, combining the use of *Concrete Experience* and *Reflective Observation*; assimilate, combining the use of *Abstract Conceptualization* and *Reflective Observation*; converge, combining the use of *Abstract Conceptualization* and *Active Experimentation*; and accommodate, combining the use of *Active Experimentation* and *Concrete Experience*.

d) Inquiry-Based Learning

Inductive Learning (IL) proposes a self-directed, student-centred approach, where students learn actively by doing, by discussing questions and solving problems (Prince & Felder, 2006). IL contains (among others) Inquiry-Based Learning (IBL), where students observe a phenomenon and generate their procedures and guiding principles.

IBL, developed during the 60s (Bruner, 1961) encourages learners to solve problems and collaboratively practice critical thinking, working with their peers to arrive together to a common goal. In IBL, students build on their previous knowledge, connecting ideas, reflecting about them and sharing them with others. This sharing of ideas is constant during the whole learning process, allowing teachers to work on learners' misconceptions and to build knowledge from them.

Joseph Schwab (Schwab, 1960) was one of the main proponents of the Inquiry-Based Learning (IBL) theory which involves social collaboration and problem solving. IBL proposes that learners gain knowledge by interacting with peers and investigating real-world challenges. Brainstorming is also highly encouraged in students. Moreover, IBL promotes communication, active listening, and reasoning skills.

e) Constructionism

Seymour Papert, the father of Constructionism, criticised already in the 90s how little the educational system had changed in the previous 200 years (mostly, in comparison to other areas like medicine) (Papert, 1993). Nevertheless, the development and easy access to technology undergone in recent years has facilitated this change in pedagogical paradigms (UNESCO, 2015). These newer approaches also show changes in the role of the teacher, who becomes a facilitator (rather than a mere provider of information), and whose function is to help students to pursue their own goals (Martinez & Stager, 2019).

The merging between Papert's interest in Piaget's Constructivism and his knowledge of computation provided the cornerstones for the development of Constructionism. Constructionism builds on Constructivism's notions on how knowledge is constructed in children's mind, adding an understanding of how constructions in the world support that knowledge. Constructionism views students as active agents in their learning process, actively building their knowledge through experience. Constructionism is an educational perspective that defends the idea of learning through making and the use of technology-enhanced environments, instead of learning via direct instruction (Papert, 1980).

Similarly to Piaget, this theory defends how learners obtain information from the outside world and mentally organize it to understand it. Papert added a more tangible flavour to these ideas, defending the importance of students pursuing concrete design and construction of objects for their learning process. Papert believed in the idea of children manipulating technology, extending and applying it. For him, the actions taken during the learning process (in, for example, building a robot) are the means to arrive at the desired outcome (like learning robotics or programming), rather than mere instructions to be followed.

Following this approach, learners use technology to explore their learning goals (Bers, 2008), and despite beginning with the same materials, they can arrive at different outcomes given their interests. Papert used the term "powerful idea" to refer to children's personal and unique understanding of the objects used to interact with. Moreover, reaching a discovery through acting on the world, allows them to cross the ZPD on their own terms and means, making an idea more meaningful to them (Bers, Ponte, Juelich, & Schenker, 2002).

Educational approaches like Papert's Constructionism relate to the importance of self-regulation in education. Giving children the (cognitive and physical) tools to be the very agents of their learning empowers them to pursue more in-depth strategies. The exploration promoted by the approaches mentioned above, allows learners to exercise their problem-solving and reasoning skills. Citing Papert, "The only really competitive skill is the skill of being able to learn". It is with this idea of learning to learn that we move to the following section of this chapter: Metacognition.

2.3. Metacognition

One can divide any task between the cognitive and metacognitive skills used to perform it. The cognitive skills represent the knowledge and procedures required to solve a problem (for example: calculating), while the metacognitive skills represent the systematic use and monitoring of that skill (when to use it and for what) as well as the ability to check the progress, quality and the outcome of the problem-solving process. Accurate and reliable metacognition that is, the knowledge of one's knowledge (Metcalfe, 2009) allows the learner to invest the proper amount of time and effort on a specific topic. This monitoring and control of the progress of study requires that one knows what one knows and what one does not know (yet). To assess metacognitive judgments, we have to analyse the relationship between the learner's prediction of their performance and the actual performance per se. A learner has good metacognition if these two variables, confidence in one's abilities and their real outcome, correlate in a positive way. Nevertheless, the literature shows that students are not good at judging what they know and what they do not know (Maki, 1998).

This section begins by providing a review of the main models of metacognition, many of which share several concepts and features. In this part, three of the main authors on metacognition will be considered: Flavell, Schraw & Denninson and Efklides (for a more detailed analysis of metacognition and self-regulation models, see (Panadero, 2017)). Consequently, we provide information about the metacognitive measure of *confidence*, that is, a judgement about the correctness of an answer, used in several studies presented in this thesis. After that, there will be a subsection on self-regulated learning and another in the relationship between metacognition and gender. Finally, we will examine the possible implications of metacognition in pedagogy.

It is important to highlight that the conceptualisation of "metacognition" can be inconsistent across its own field, lacking coherence among definitions, as it is still a growing field (Veenman, Van Hout-Wolters, & Afflerbach, 2006). The most commonly used, and the one that will be used for this thesis, is proposed by Flavell (Flavell, 1979).

a) Models of metacognition

The term metacognition was coined by Flavell in the 70s. He began referring to this kind of self-reflective processes as metamemory (Flavell & Wellman, 1977), a process which was composed by knowledge, monitoring of memory (which was later also used by Tarricone in 2011) and regulation. It was in his work in 1976 when the term metacognition appeared, to refer to the process related to more than just memory. It must be noted that the author posed two definitions for the term: the first one, in 1979 (Flavell, 1979); and the latter one, at the beginning of the century (Flavell, Miller, & Miller, 2002), after acknowledging other researchers' contributions to the field during the 80s and 90s. Thus, in this subsection, we will see how the metacognition terminology has evolved during the years.

Flavell

John Flavell can be considered a leading figure in the field of metacognition, with his work serving as the foundation of much subsequent research. Flavell distinguishes between three main components of metacognition: *Metacognitive Knowledge* (MK), metacognitive monitoring (also known as *Metacognitive Skills (MS*) (Efklides, 2006a; Veenman et al., 2006)) and *Metacognitive Experiences* (ME) (Flavell et al., 2002). MK comprises students' knowledge and beliefs about their interactions with cognitive tasks and strategies and can be divided into three categories: knowledge about people, tasks, and strategies. Knowledge about people refers to learners' knowledge both about themselves and others. Knowledge about tasks refers to students' understanding of the task's demands and goals. Finally, knowledge of strategies refers to how learners conceptualise thinking and problem-solving strategies to use them to achieve their goals.

Metacognitive Skills allow students to monitor and regulate cognition and learning (Veenman, 2011). Schraw and Moshman have split MS into three subcategories: planning, monitoring and evaluation (Schraw & Moshman, 1995). During planning, learners set goals, select appropriate strategies, make predictions and allocate resources. During monitoring, they observe their behaviour, and during evaluation, they assess the efficiency of their thinking, for example, through reflection. Finally, *Metacognitive Experiences* (ME) represent the "cognitive or affective experiences that pertain to a cognitive enterprise" (Flavell et al., 2002). They reflect the feelings of puzzlement or "eureka moments" undergone by students during a task. Anastasia Efklides, whose work will be presented at the end of this subsection, defines them as judgements and online task-specific knowledge, which are related to cognitive and affective regulatory loops (Efklides, 2006).

A decade later, Brown (Brown, Bransford, Ferrar, & Campione, 1983) proposed a division of metacognition between metacognitive knowledge (which comprised knowledge about one's own and others' cognitive processes) and metacognitive regulation (which referred to the processes of planning, monitoring, and checking). Contrarily to Flavell's theory, neither Brown nor Jacob & Paris did not consider metacognitive experiences as part of metacognition. Nevertheless, contrarily to Brown & Co's theory, they did not consider as metacognition any automatic or implicit cognition that could not be reported. Jacob & Paris divided metacognition into self-appraisal of cognition (which would refer to metacognitive knowledge) and self-management of thinking (which would account for metacognitive regulation). Self-appraisal of cognition represented declarative (what), procedural (how) and conditional (why and when) knowledge; while self-management of thinking represented planning, evaluation, and regulation.

Schraw & Denninson

Similarly to Brown and Jacob & Paris, Schraw & Denninson's (Schraw & Dennison, 1994) model of metacognition was also divided into metacognitive knowledge and metacognitive regulation. The metacognitive knowledge part contained declarative, procedural and conditional knowledge. The metacognitive regulation represented planning, information management strategies, comprehension monitoring, debugging strategies, and evaluation. One of the main differences of their model in comparison to the one from Jacob & Paris is that it is specifically oriented towards learning environments.

Similar to the aforementioned Kolb's phases of experiential learning, appropriate metacognition comprises the mastering of three skills: planning, monitoring, and evaluating (Woolfolk, 2013). Planning refers to the ability to select appropriate strategies for a task, like the amount of time that one needs to spend on the task, the order of steps, or what to focus on. Being aware of one's performance during a task is the main outcome of monitoring. Finally, evaluating is the ability to make proper judgments about the processes undertaken during the task and their results (Kleitman & Stankov, 2007).

Efklides

Until this moment, early theories of metacognition incorporated metacognitive knowledge, regulation or experiences. Anastasia Efklides added one more dimension, metacognitive skills, going back to Flavell's approach. Despite not being the first researcher to propose a three-item model (see, e.g. (Pintrich et al., 2000)), her model represented a great contribution to the fields of metacognition and self-regulated learning. Her theory comprised metacognitive knowledge, experiences and skills. Metacognitive knowledge represented declarative knowledge and contained models of cognitive processes, epistemological beliefs and knowledge of persons (self& others), tasks, strategies (including conditional knowledge) and goals. Metacognitive experiences arise from self-awareness during a task and contain feelings, judgements, estimates and online task-specific knowledge. Finally, metacognitive skills represent those deliberately used strategies supporting the control of cognition: orientation, planning, regulation, monitoring and evaluation.

Efklides also developed the Metacognitive and Affective Model of Self-Regulated Learning (MASRL model, (Efklides, 2011)), which was divided into person level and person x task level. The person-level is relatively stable and influences self-regulated learning in a top-down manner. It includes trait-like knowledge and characteristics of cognition, metacognition, motivation, affect and volition. On the contrary, the Person x task level is dynamic and influences Self-Regulated Learning (SRL) in a bottom-up way. It includes individual task-related processing and any subjective experiences that relate to the task. This top-down/bottom-up reciprocity ensures that each level has the potential to inform and change the other.

Tarricone

In 2011, Tarricone wrote a review on the literature of metacognition, which led to her Taxonomy of metacognition (Tarricone, 2011). Following the main view, she divided metacognition into metacognitive knowledge (MK) and regulation (MR). MK contained declarative, procedural and conditional knowledge, like Jacob & Paris' and Schraw & Denninson's models. MR contained regulation of cognition and executive functioning (both can relate to person, task, and strategy) and metacognitive experiences. The first one is also split into monitoring and control and self-regulation. Metacognitive experiences are considered the results of online monitoring and represent metacognitive feelings and metacognitive judgements. Feelings contain task familiarity and difficulty, confidence, feeling of knowing and satisfaction with performance—judgements estimates of learning, memory accuracy, solution correctness, effort expenditure and strategy effectiveness.

As one can understand from the literature provided, several models to represent students' metacognition have been developed in the last years. **Table 1** presents a summary of the theories presented and the terminology used.

Table 1. Summary of the metacognition models presented.

Author	Year	Terminology
Flavell	70s	*Metacognitive Knowledge* (knowledge about people, tasks, and strategies) *Metacognitive Monitoring/Skills* *Metacognitive Experiences*
Brown & Co	1983	*Metacognitive Knowledge* *Metacognitive Regulation*
Jacob & Paris	1987	*Metacognitive Knowledge* (declarative, procedural and conditional knowledge) *Metacognitive Regulation*
Schraw & Denninson's	1994	*Metacognitive Knowledge* (declarative, procedural and conditional knowledge) *Metacognitive Regulation* (planning, information management strategies, comprehension monitoring, debugging strategies, and evaluation)
Efklides	2008	*Metacognitive Knowledge* (cognitive processes, epistemological beliefs and knowledge of persons -self& others-, tasks, strategies -including conditional knowledge- and goals) *Metacognitive Experiences* (feelings, judgements, estimates and online task-specific knowledge) *Metacognitive Skills* (orientation, planning, regulation, monitoring and evaluation)
Tarricone	2011	*Metacognitive Knowledge* (declarative, procedural and conditional knowledge) *Metacognitive Regulation* (regulation of cognition and executive functioning -both can relate to person, task, and strategy- and metacognitive experiences) *Metacognitive Experiences*:

		Metacognitive Feelings (task familiarity and difficulty, confidence, feeling of knowing and satisfaction with performance) *Metacognitive Judgements* (judgements, estimates of learning, memory accuracy, solution correctness, effort expenditure and strategy effectiveness)

In this thesis, we will use the models coined by Flavell, Schraw, Efklides and Tarricone as a base for our metacognitive exploration. In them, mostly Efklides and Tarricone, consider several measures of metacognitive accuracy. In our case, we will focus on (the feeling of) *confidence*, one of the measures used to analyse students' metacognitive abilities (and individuals' perception of their performance in many decision-making tasks).

b) Confidence

Decision confidence (DC) (from now on, confidence) represents an estimate of the probability that a decision is correct (Boldt, Blundell, & De Martino, 2019). This variable allows the brain to exploit information about the environment (Knill & Pouget, 2004). DC has been studied since the early times of experimental psychology (Henmon, 1911; Peirce & Jastrow, 1884), where a larger part of the research focused on memory and perception. More recent work has extended the study of confidence towards learning and value-based choice (Lebreton, Abitbol, Daunizeau, & Pessiglione, 2015; De Martino, Fleming, & Garrett, 2013; Meyniel, Schlunegger, & Dehaene, 2015). Nevertheless, despite the extensive research in this topic, there is not a unified definition to represent this operation of the mind, and it is often presented as a feeling (Insabato, Pannunzi, Rolls, & Deco, 2010), as "certainty" (Fimbel, Michaud, & Martin, 2009), "subjective probability" (Kahneman & Tversky, 1972) or "introspective accuracy" (Fleming, Weil, Nagy, Dolan, & Rees, 2010).

Confidence has an important role in regulating effort. Nevertheless, it is a subjective value, so it is not always well calibrated with accuracy. The research on confidence and value beliefs has been supported by both behavioural and neuroimaging studies (Lebreton et al., 2015; Vaghi et al., 2017). More recently, it has also been studied by the field of neurocomputation, showing, for example, the contribution of certainty and confidence on perceptual choice (Bang & Fleming, 2018).

The confidence shown by learners on their capabilities has been found to correlate with the way they approach a learning task. On the one hand, confidence is positively correlated with deep learning; that is, the kind of learning focused on understanding and acquiring meaningful knowledge through the elaboration of ideas and the use of critical thinking. On the other hand, it is negatively correlated with surface learning, characterised by being based on mere memorisation and with motives far from the purpose of the task (Geitz, Brinke, & Kirschner, 2016).

The role of confidence in learning becomes enhanced while paired with feedback. Incorrect mental schemes accompanied by high confidence create a surprising feeling in the learner, which in turn leads to better retention of the content as learners pay more attention to feedback. Low-confidence correct responses enhance retention, as they also surprise the learner, by enabling him to strengthen the association between cue and response to inhibit any competing responses (Butler, Karpicke, & Roediger, 2008). When the errors are accompanied by high confidence, they are more likely to be corrected after receiving feedback than the low-confidence ones, or what is called the *hypercorrection effect* (Butterfield & Metcalfe, 2001).

High confidence in one's abilities (but not overconfidence) allows the student to persist in the task. Confidence is composed of several components (John Keller, 1983): *expectancy for success*, *perceived confidence* and *perceived control*. *Expectancy for success* refers to students' estimation of their possibilities to succeed in a task. *Perceived confidence* refers to a more internal belief regarding their prerequisite knowledge and skills (Harter, 1978, 1982). Finally, *perceived control* focuses on the beliefs about the causes of their success (John Keller, 1983). A proper instruction/feedback is the one that gives learners tasks that are aligned with their knowledge, making them neither too easy nor too difficult. Additionally, it should make them feel confident that they can perform the task (provided that the difficulty of the task matches their abilities) and its goal should be to guide students towards the acquisition of the requested knowledge and skills.

Although overconfidence may help to provide the motivation to move forward without giving up (Fast, Sivanathan, Mayer, & Galinsky, 2012), students' learning process can be inhibited by poor metacognition. On the one side, if a learner does not have enough knowledge and skills but is overconfident, they could continue attempting the task without following the provided feedback; consequently, this could lead them to give up or to feel underconfident because of the repeated failures. On the other side, if a learner has enough knowledge but is underconfident of their ability, they could feel as having a low chance to succeed, which in turn could lead them not to put enough effort or even resulting in giving up (Warren, 2012).

The traditional curricula currently taught at schools has been shown to promote over-confidence in the students, as it does not foster confusion of failure (Brickman, Gormally, Armstrong, & Hallar, 2009). A proper system is the one that gives learners tasks that are aligned with their knowledge, making tasks neither too easy nor too difficult, to maintain the Zone of Proximal Development (see previous section). By ensuring that the difficulty of a task matches their abilities, learners can feel confident they can perform the task. Finally, tasks that have a defined goal to guide them towards the acquisition of the requested knowledge and skills are important.

c) Metacognition and Self-Regulated Learning

Self-regulation and self-directed learning skills allow students to help themselves by adapting to changes. Although similar, those terms are not totally interchangeable. In Self-directed learning (SDL), learners define their own learning tasks; while in Self-Regulated Learning (SRL) (Zimmerman, 1989) it may be the facilitator who provides it for them. In both approaches, individuals diagnose their learning needs and strategies to formulate their goals and later evaluate their outcomes (Knowles, 1975). Nevertheless, in SRL, the teacher helps learners by providing initial instructions (Saks & Leijen, 2014) Consequently, due to the autonomy given to the learner in SDL, this approach is better suited for older students, while SRL (coming from educational psychology), is mostly used in school environments. Some authors include SRL inside of SDL (but not the other way around) (Loyens, Magda, & Rikers, 2008; Saks & Leijen, 2014).

d) Gender in Metacognition

The aforementioned differences between actual and perceived performance appear to be larger in female students. (As a note, I would like to clarify that with *gender* I mean the gender the learner identifies with, not their biological sex). Females seem to express higher levels of self-criticism and self-handicapping behaviour (Hirt, McCrea, & Boris, 2003), and to attribute success more frequently than men to higher ability and failure to lower ability (Dickhäuser & Meyer, 2006).

Some studies relate these differences to differences in competitiveness, showing females as less competitive than males (Gneezy, Niederle, & Rustichini, 2003). Females' lack of confidence leads them to low participation in competition while males' overconfidence results in excessive participation in competition (Niederle & Vesterlund, 2007). Nevertheless, we have to be cautious, as it can fall in a circular explanation, considering some studies present these differences in competitiveness as the cause of the differences in perceived ability while other present them as the outcome (Gneezy et al., 2003). Consequently, more research is needed to assess the directionality of the relationship between these two components.

Moreover, the gender ratio in the class can also play a role on females' decision to participate in competitive environments (Balafoutas & Sutter, 2012; Niederle, Segal, & Vesterlund, 2013), regardless of their performance (Gneezy et al., 2003). Nevertheless, as already stated by some authors (Dunning, Heath, & Suls, 2004) the role of gender in meta-reasoning processes needs to be further investigated.

These gender differences in self-perception increase even more in STEM-related tasks. When dealing with mathematical tasks, for example, females' perception of their abilities is significantly lower than those of male regardless of their actual performance (Else-quest, Hyde, & Linn, 2010). It is important to remark the fact that these metacognitive differences are not related to differences in levels of mathematical performance (Else-quest et al., 2010; Hyde, Lindberg, Linn, Ellis, & Williams, 2008; Lindberg, Hyde, Petersen, & Linn, 2010; Meelissen & Luyten, 2008), but to higher levels of mathematical anxiety (that is, a feeling of tension or fear interfering with math performance (Ashcraft, 2002)) and more negative attitudes towards mathematics (Frenzel, Pekrun, & Goetz, 2007; Goetz, Bieg, Lüdtke, Pekrun, & Hall, 2013).

These stereotypes and prejudices about gender might result from the social categorisation children experience in their first years of life (for example, the use of blue and pink to differentiate between boys and girls, respectively) (Biemmi & Satta, 2017). It is important to highlight that gender stereotype is a bipolar construct, which means that it presents mutually-exclusive situations: What is masculine cannot be feminine and vice versa (Deaux & LaFrance, 1998; Renfrow & Howard, 2013). These stereotypes on the relation between gender and career already begin in the early years of elementary school (Sullivan & Bers, 2016). This is in line with the differences in self-perception between genders being stronger in high school.

Already in the 90s, Sadker & Sadker defined what is known as the *confidence gap*, which represents the difference between genders in confidence in STEM-related careers. This self-perception emerges already during adolescence and can influence future academic and career choices (Halpern et al., 2007), as this drop in females' perception of own competence (Pajares, 2003) can lead them to not feeling confident enough to pursue "gender-biased" studies (Perez-felkner, Nix, Thomas, & Kirkham, 2017).

There is not a clear consensus in the metacognitive differences between genders. Some authors report significant differences between genders, while others highlight the dependency of those differences in the knowledge domain (Voyer & Voyer, 2014). For example, high-school male students report higher confidence in STEM-related subjects (which may not correlate with their actual performance (Stankov, Lee, Luo, & Hogan, 2012)), and this difference that increases with age (Pajares, 2002).

As shown in the results of Chapter 3, the gender ratio in the class can also play a role on females' performance. This can be supported by previous results on gender ratio affecting females' decision to participate in competitive environments, (Balafoutas & Sutter, 2012; Niederle et al., 2013), regardless of their performance (Gneezy et al., 2003). These insights should be thus taken into consideration when designing class distributions or even group activities.

The negative gender gap in overconfidence against girls has been shown to be greater for students in the higher quartiles than those in the lower quartiles. This is in line with the differences between genders in the discrepancy between perceived and actual performance shown in Chapter 3, where the results were stronger for the top quartile.

Another social factor that has recently been proposed as a possible cause of gender disparity in school achievement is the potential effect of *stereotype threat* (Voyer & Voyer Susan, 2014). Stereotype threat arises when the performance of a group is influenced by the understanding that its members belong to a social group that is not expected to perform well in a role.

This stereotype arises from an early age. Results from Hartley and Sutton (Hartley & Sutton, 2013) show that even at an early (4-7 years old) age, both genders think that adults expect girls to be better students than boys. Moreover, emphasizing or countering this belief affects (either negatively or positively, respectively) on boy's performance in reading, writing and mathematics. Nevertheless, these manipulations did not seem to affect females' performance.

Other studies (Igbo, Onu, & Obiyo, 2015) observed that while gender had little impact on the students' actual academic achievement, gender expectations had a major influence on how students measured their academic achievement in mathematics. More specifically, despite girls showed higher scores than their male counterparts, the latter evaluated their mathematics' knowledge as higher than the female participants.

It is also important to highlight that this stereotype does not only affect females and does not only relate to STEM. Previous studies (Chatard, Guimond, & Selimbegovic, 2007) explored the impact of gender stereotyping on how high school students perceive their skills in mathematics and the arts, with the purpose of testing self-stereotyping in a stereotypically masculine domain (math) and a stereotypically feminine domain (arts). They showed that students who hold strong beliefs in gender stereotyping, when asked to self-evaluate their grades in these two domains, answer following a consistent pattern with their beliefs.

e) Implications in pedagogy

When students have knowledge and skills but do not believe that they can utilise them to execute the necessary strategy, it is not always clear how their specific mindset can be influenced to increase their so-called *self-efficacy* (Chan & Lam, 2010). Self-efficacy, that is, how a person judges their capabilities to carry on designated types of performances (Bandura, 1986) is considered central to promote students' learning and engagement. Self-efficacy has been associated with several of other learning and motivational-related variables: self-efficacy is positively related to mastery goal orientation (Ross, Shannon, Salisbury-Glennon, & Guarino, 2002) and promotes more self-regulatory strategies (Pintrich & de Groot, 1990; Zimmerman, 2000).

The importance of considering students' beliefs about themselves is not limited only to pedagogical theories meant for children; it is but also visible in adult learning theories used in work environments. Victor Vroom, the father of the *Expectancy theory*, highlighted the importance of considering individuals' expectations and beliefs about themselves, as they can affect their behaviour during a task. Vroom highlighted the importance of differentiating between extrinsic and intrinsic rewards and strengthening learners' beliefs in themselves. In a similar line, Berne's *Values model*, also considered learners' confidence in themselves (and others) as a pivotal influence on their ways of behaving and interacting.

Errors in metacognition are defined by a discrepancy between the subjective and objective correctness of learners' answers. This bias measure (the difference between confidence and performance) can have a positive or a negative value. When the learner makes a correct response but is not confident in it, they are underconfident; when, on the contrary, they have high confidence in an error, they are overconfident of their abilities. Overconfident learners may make errors of commission (that is, they act when they should not) while underconfident ones may make errors of omission (not acting when they are correct).

Feedback can help in both cases by correcting this metacognitive discrepancy. On the one side, feedback enhances retention of correct responses accompanied by low-confidence by strengthening the association between question and answer and inhibiting competing responses (Butler, Karpicke, & Roediger III, 2008); on the other side, it helps to correct high-confidence errors, promoting what is called the *hyper-correction effect* (Butler, Karpicke, & Roediger III, 2008). Thus, feedback promotes learners' metacognition by providing them with reinsurance of their theories or a contradiction to them. As a result, feedback can improve resolution and calibration of confidence judgments, as claimed by (Butler, Karpicke, & Roediger III, 2008).

Children can still be overconfident after experience and feedback, not because of a lack of metacognition, but as a protection against loss of motivation (Bjorklund, 1997). Contrarily, adults may shift to underconfidence (which is known as *underconfidence with practice*) (Koriat, Sheffer, & Ma'ayan, 2002). There can be several reasons a student remains overconfident in their performance There can be self-enhancement motives, showing motivational biases as wishful thinking (Schneider, 1998); chronic self-views and informational differences between self and others. There is also the possibility that the learner suffers from a lack of competence to understand one's level of incompetence.

Maintaining a state of overconfidence after feedback can lead to making poor study choices, impeding learning (Dunlosky & Rawson, 2012). Training students' metacognitive abilities has been previously shown to have beneficial effects, both in reading and mathematics-related activities (Roll, Aleven, McLaren, &

Koedinger, 2007). What is more, Hattie, in his well-known review of instructional approaches (Hattie, n.d.), highlighted metacognitive and self-regulated approaches among the most effective ones (mean effect size of 0.67).

Accordingly, failures of metacognition may lead to maladaptive decision-making: people who are overconfident of their knowledge about information security (a positive metacognitive bias) are more likely to take risks when using the internet (Heyes, Bang, Shea, Frith, & Fleming, 2020) and people with weaker metacognitive sensitivity are more likely to hold radical beliefs at both ends of the political spectrum (Rollwage, Dolan, & Fleming, 2018).

One of the subjects that seems to be more affected with self-beliefs is Mathematics, the topic of the next subsection.

2.4. Mathematical Learning

As stated in the previous chapter, many of the tasks used in this thesis are STEM-related and mostly mathematics. Mathematical learning seems to be one of the subjects more affected by self-perception of one's abilities. Moreover, it is a subject less related to language and presents a good framework for international projects on learning.

This section will present the relationship between students' metacognitive abilities and mathematics learning and how learning is affected by mathematical anxiety. Moreover, it will present the Balance Scale task, a mathematical task used in this thesis.

a) Metacognition in Mathematics

The discrepancy between perceived and actual ability is also evident in mathematical and science-related learning. Many preconceptions and biases surround mathematical understanding, generating doubts about one's skills even in confident learners (Betz, 1978; Cvencek, Meltzoff, & Greenwald, n.d.; Rubinsten, Bialik, & Solar, 2012). This is evident in statements as “I am just not a math person” (Dweck, 2008; Rattan, Good, & Dweck, 2012). This identification (or not) comes from students’ confidence in maths skills and in their ability to learn math, as well as from mathematical anxiety (Ashcraft, 2002; Jansen et al., 2013).

The term “mathematics confidence” was defined by Pierce and (Pierce & Stacey, 2004) and represented “a student’s perception of their ability to attain good results and their assurance that they can handle difficulties in mathematics”. Part of this mathematics confidence comprises mathematical self-efficacy, that is, a learner’s belief in their likelihood of carrying out a mathematical task in a correct manner. Lastly, “mathematical resilience” (Johnston-Wilder, Brindley, & Dent, 2014) denotes how students overcome barriers to learn mathematics. Another motivational measure related to mathematical learning is Mathematical Anxiety, which will be further presented in the study of Chapter 5.

b) Balance-Beam Task

Two of the studies presented in this task are based on the Balance Beam task (Piaget & Inhelder, 1958; Siegler, Strauss, & Levin, 1981). This mathematics task was originally created by Inhelder and Piaget as a way of assessing if a child was in the *Formal Operational Stage* previously mentioned and it has been extensively used to research about children’s cognitive development. It consists of understanding the relationship between weight and distance by solving where will a balance fall (left, right or if it will remain in equilibrium) when the weights are placed on the scale. In order to succeed in the task, the learner has to be able to understand the physical concepts of weight and distance and the (multiplicative) relation between them.

As previously mentioned, Piaget proposed that learning is a constructive activity comprising successive stages of adaptation to the world, during which learners actively construct knowledge by creating and testing their own theories. In similar lines, developmental stages can be understood as constructions of learners' re-organisation (Fosnot & Perry, 2005). Piaget defined four Stages of Cognitive Development (Piaget, 1972), related to the way the child constructs the previously mentioned schemes of the world. Although the stages include age in their description, age is only an indication, and strict boundaries should not be set between them. Piaget claims that the speed at which children go from one stage to the other may vary due to individual differences.

- In the first stage, the *Sensorimotor Stage*, the (birth – 2 years old) infants begin to form mental representations (schemes) of the objects.

- In the second stage, the *Preoperational Stage* (from 2 to 7 years old), children are able to think about things in a symbolic way, with language being the most obvious form of symbolism; but their type of thinking is still egocentric, they are not able to think from the point of view of another person (Huitt. 2003).

- It is in the *Concrete Operational Stage* (from 7 to 9 years old) when the first logical thoughts appear. This is an important step in the cognitive development of the child, as now they can work things out in their head (instead of only trying them physically). In this stage, children are able to apply conservation (the ability of understanding that something conserves its quantity although its appearance changes) to numbers, mass and weights.

- In the last stage, the *Formal Operational Stage* (from 11 years to adulthood), children can think about abstract concepts and test hypothesis. An experiment concerning this stage will be explained in the following sections.

The usage of the Balance Beam task in the present work is twofold: on the one hand, it constitutes a simple inquiry-based learning task where children's performance can be fully described through the application of the hierarchy of rules of increasing complexity that can be operationally controlled. On the other hand, the Balance Beam problem involves reasoning about physical properties (such as weight and distance) that can be exploited through different means (e.g., manipulation of physical objects or graphical material). ***Figure 2*** depicts the task's four rules. More specifically, to correctly solve the Balance Beam task, learners need to solve five kinds of items (Siegler et al., 1981), which can be divided between non-conflict and conflict items:

- **Non-conflict items:** In this kind of setup, one dimension is equal on both sides, so the balance tilts to the side where the other dimension is larger:
    - **Dominant / Weight item:** The distance between the sides is the same; the only difference between them is the number of weights per side. Thus, the balance tilts to the side with the largest weight.
    - **Subordinate / Distance item:** There is no difference in weight between sides; the only difference is the distance between the weights. Thus, the balance tilts to the side with the weights positioned further from the fulcrum (the centre).
- **Conflict items:** In this kind of setup, the dimension that is larger in one side is smaller in the other, creating a "conflict" between both.
    - **Conflict-dominant / Conflict-weight item:** In this case, the balance will tilt to the side with the larger weight
    - **Conflict-subordinate / Conflict-distance item:** In this case, the balance will tilt to the side with the larger distance
    - **Conflict-equal / Conflict-balance item:** In this case, the balance will remain in balance

Considering this, four patterns of answer arise, depending on the rule (strategy) used to answer the items:

- **Rule I:** Following this rule, judgements are based solely on weight, regardless of the distance between the items. If the sides have equal weight, the beam will remain in balance; if not, it will fall to the one with the largest weight. In this case, only dominant / weight items and conflict-dominant / conflict-weight items would be answered correctly

- **Rule II:** Children in this rule would follow a similar pattern but also answering subordinate / distance items correctly
- **Rule III:** In this rule, children would answer non-conflict items correctly and perform at chance level (33%) in the conflict items
- **Rule IV:** This last rule represents perfect performance in all items

According to Siegler (Siegler & Chen, 1998), one factor that influences the learning of the rules is the discrepancy between the existing and targeted knowledge. Children who already make use of Rule I are more likely to acquire a more complex rule if presented with a setup that can be solved by Rule II than if presented with a kind of problem that can be only solved by applying Rule III. This hierarchical exploitation of rules, therefore allows a higher degree of control over the different stages of learning that will be presented in the next subsection. On the other hand, the balance beam problem involves reasoning about physical properties (such as weight and distance) that can be exploited through different means (e.g., manipulation of physical objects or graphical material).

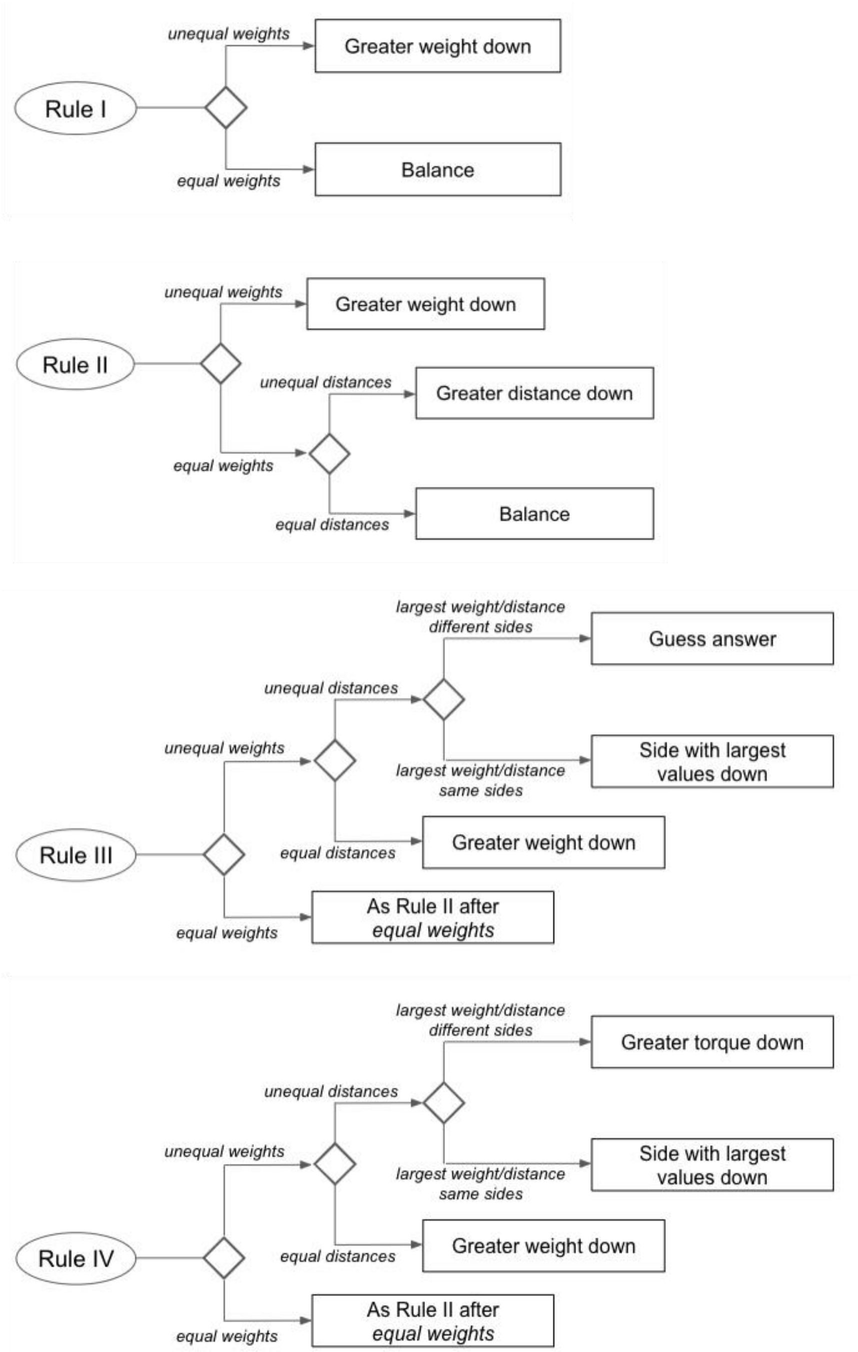


Figure 2. Rules explaining the behaviour during the task, by (Siegler et al., 1981). Adapted from (van Rijn, van Someren, & van der Maas, 2003).

2.5. Learning Technologies

a) Individualized Learning and Intelligent Tutoring Systems

Despite the development of the theories of learning and the effort of teachers, the current educational system still focuses on specific learning goals based on predefined methodologies and has a structure where the individual differences of the learners are poorly managed (Hattie and Yates, 2013). This educational paradigm can be caused by the lack of a consistent scientific framework to ground these goals on the mechanisms of learning, memory, and cognition. Tiberius and Billson (Tiberius & Billson, 1991) defined two levels of teaching: teaching as a mere knowledge transfer and teaching as a social dialogue. Here, teaching is seen as a social dialogue, and teachers help students to achieve higher learning performances by giving them feedback and influencing their motivation. Skinner has highlighted the importance of feedback in education and argued that when students receive contingent responses and help from their teachers, they become more engaged in school and have better grades; especially if the interaction has been individualised (Skinner & Belmont, 1993).

Unfortunately, traditional educational methods like paper-based tests miss the possibility of getting that much into detail when assessing the characteristics of the student. It is for that reason that approaches like the use of Intelligent Tutoring Systems allow for a deeper retrieval of the data provided by the student. Apart from the influence that technology can have on the performance and motivation of the student, Intelligent Tutoring Systems allow for individual assistance to the learner, conveying the paradigm shift that is leaving behind the "one-size-fits-all” approach (Grant & Basye, 2014).

Technology allows educators to monitor the learning process of the student. By monitoring their problem-solving skills, educators can identify which are their specific areas of difficulty, and generate exercises targeting the individual’s needs (Corbett, Anderson, Carver, & Brancolini, 1994; Koedinger & Anderson, 1997). Intelligent Tutoring Systems can become a tool to improve learners’ metacognition by training them on recognising their own mistakes and achievements.

Monitoring the learning process of the student through technology gives access to more detailed information hidden at first sight that is not available by traditional methods. This information allows creating a model of the student able to predict the possible outcome of their actions and adapt in the best way to the learner. Current models usually use the performance or the affective state of the student (via the analysis of facial expressions, physiological data, et.c) to improve their learning, but miss the assessment of the confidence the student has on their performance.

The remarkable role of the analysis and regulation of affective states of the student to foster effective and long-term learning was already defended by Yerkes and Dodson at the beginning of the 20th Century (Yerkes & Dodson, 1908), and this idea has been extensively researched after them (Barth & Funke, 2010; Isen & Labroo, 2012). Nevertheless, the study of affective states during the learning process can become a challenging task, as a result of the subjective nature of emotions (Porayska-Pomsta, Mavrikis, D'Mello, Conati, & Baker, 2013). It is for that reason that the possibility to automatically detect and log data offered by technology, as well as facilitating the real-time adaptation to the learner, appears as a remarkable tool to improve the study of affective states related to learning processes (Conati & Zhou, 2002; Mello, Craig, Witherspoon, & Mcdaniel, 2008).

A number of synthetic tutors have been built that use pedagogical approaches other than individualised tutored problem-solving: educational games, collaborative learning, learning with an external problem-solving environment, activities that target sense-making and fluency building, and guided invention activities. Most of the instructional decisions taken by automated tutors adapting to students' paces and needs, selecting proper tasks and exercises, or organising instructional messages- are based on the information coming from student models (Koedinger, McLaughlin, & Stamper, 2012).

A good student model fosters learning, which, in turn, is critical to promote accurate self-assessment of current skills or states. It is a circular process: the more accurate the skill diagnosis is, the better the model can predict what the learner knows, which in turn allows for a better assessment, leading to more efficient learning (Koedinger et al., 2012). As Stamper and Koedinger (Stamper & Koedinger, 2011)suggested, a proper student model can promote more effective learning by using resequencing, knowledge tracing, creating new tasks and changing instructional messages, feedback or hints.

One way of promoting efficient metacognition is developing a system that gives feedback after each answer. It has been shown that a correctly timed feedback, provided attempt by attempt, allows for faster improvement in metacognition (Rawson and Dunlosky, 2007). The feeling of confidence is contained inside metacognition, which includes judgments, feelings and thoughts during the task (Kleitman and Gibson, 2011). When students have high confidence in their abilities, they see difficult tasks as challenges to be mastered, which helps to foster an intrinsic interest and deep fixation in activities (Shannon, 2008).

The learner's confidence in their ability has been shown to play an important role in improving their performance (among other requirements, like facing tasks aligned with their knowledge and being goal-oriented to improve) (Geitz et al., 2016). These beliefs can influence the learner's behaviour towards achievements, like effort or persistence. Individuals with higher confidence are more prone to get involved in tasks and, when presented with a difficult task, they consider it a challenge to master rather than a threat to be avoided, which increases the possibilities to complete it.

2.6. The Distributed Adaptive Control (DAC) Architecture

The development of a Synthetic Tutor Assistant to scaffold the learning process of students' by adapting to their individual differences is the focus of the European project Expressive Agents for Symbiotic Education and Learning (EASEL). The work summarised in this section is based on the development of a Synthetic Tutor Assistant (STA) (Vouloutsi et al., 2016). The STA's main goal is to guide the learner through an educational task, by offering diverse tutoring strategies based on the performance and capabilities of each individual student. The STA's implementation and pedagogical model is grounded in the Distributed Adaptive Control (DAC) theory of mind, brain and body nexus (Verschure, 2012; Verschure, Voegtlin, & Douglas, 2003) which serves as a real-time neuronal model for perception, cognition, and action (P. F. M.. Verschure, Pennartz, & Pezzulo, 2014). According to DAC, the brain is viewed as a layered control architecture, dominated by parallel and distributed control loops, that is subdivided into functional segments that reflect how we process the states of the world, the self, and the interaction through action (Verschure et al., 2003). As a control system, DAC comprises four layers: Somatic, Reactive, Adaptive, and Contextual (***Figure 3***). Across the layers, we distinguish three functional columns of organisation: the world (exosensing), the self (endosensing) and the interface between the self and the world through action.

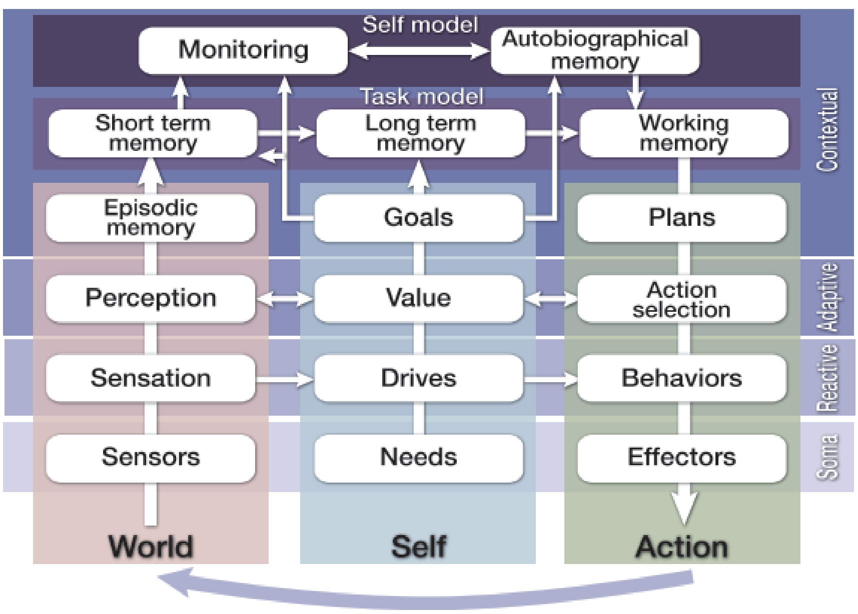


Figure 3. DAC architecture

a) DAC as a pedagogical model

As a pedagogical model, DAC views the learner as a complex individual whose behaviour derives from the interplay of sensation, reaction, adaptation, and contextualisation. DAC predicts that in order to learn and consolidate new material, the learner undergoes a sequence of learning phases: *resistance*, *confusion*, and *abduction*. Resistance refers to a mechanism that results from defending one's own (in)competence level. Students tend to hold overly optimistic and confused views about their level of knowledge: those with a good understanding of a topic tend to underestimate their capabilities and those who do not, tend to overestimate them (Kidd & Monk, 2009; Kruger & Dunning, 1999; Grant, Malloy, & Murphy, 2009). This feeling is what we refer to as resistance, and what consequently leads to a state of confusion.

In similar lines, Peirce established the difference between the ego, the non-ego and their perception. Considering Peirce's theory (Peirce, 1997), every agent has their own idea of the world (what he defines as ego), in contrast with the real way the world is (define as non-ego). When individuals are faced with a situation that does not fit their inner world model (ego), they would be surprised. DAC proposes an equivalent paradigm, where the difference between the inner world of the learner derived from their previous knowledge (similar to Peirce's ego) and the world itself, which presents the new knowledge (Peirce's non-ego) causes confusion, what the learner has to overcome in order to abduct and acquire knowledge. Peirce's surprise sets off a struggle.

Confusion is what creates the necessity to resolve the problem and learn through re-adapting. Human learners show significant variability in their performance and aptitude (Felder & Brent, 2005). There are moments when the learner is not able to advance in the learning process on their own and needs external guidance. Thus, for learning to be effective and suitable for a wider scope of individuals, learning technologies need to adapt to the abilities and the progress of every student. Adapting to the skills and progress of individual students helps to maintain the process of learning acquisition; it is thus essential to sustain a challenging enough task. Monitoring, controlling and adjusting the confusion is what we define as *shaping the landscape of success*. Such approach is consistent with the concept of *scaffolding*, Vygotsky's *Zone of Proximal Development (ZPD).*

Hence, confusion needs to be controlled so that it does not lead to a complete loss of motivation or development of *learned helplessness* (Abramson, Seligman, & Teasdale, 1978); the student needs to believe that he can be effective in controlling the relevant events within the learning process (Seligman, 2007).To do so, the system needs to adjust to the skills and progress of the learner, allowing for the task to remain challenging enough. Confusion is necessary to discover and generate theories and assess them later, that is, to be able to perform *abduction*. *Abduction* is the very process of acquiring and stabilising new knowledge. This DAC-derived learning reflects the core notions of Piaget's theory of cognitive development assimilation and accommodation through a process of equilibration (Piaget, 1952; Wadsworth, 1996).

Given the literature presented above, we will first begin with a study on students' discrepancy between actual and perceived performance.

3. STUDY 1: Students' difficulties in recognizing their competence level

The first study presented in this thesis was designed with the aim of assessing the relation between perceived and actual performance in learning. To do so, we carried out a task, based on the Dunning-Kruger effect, where participants answered a 15-items test and report their confidence in each of their answers, apart from providing their general confidence at the end of the test.

3.1 Introduction

a) The original study

Back in 1999, Kruger and Dunning presented a study on how the difficulties in recognizing one's own incompetence lead to inflated self-assessment (Kruger & Dunning, 1999). They determined three regularities: the "unskilled" (those with low performance) overestimated their performance, while the "skilled" (those with high performance) underestimated it; and these discrepancies were asymmetric: a higher number of unskilled overestimated their performance. The study discusses how the necessary skills to correctly judge performance are the ones already necessary to perform correctly in the first side.

The original study was divided between three parts, each with a different test topic: Humor, logical reasoning, and English grammar. The participants were 45 undergraduates from an introductory psychology course.

- Poor performers (the ''unskilled'') overestimated their absolute and relative performance;

- Top performers (the ''skilled'') underestimated their absolute and relative performance;

- These miscalibrations were typically highly asymmetric in that many more unskilled overestimated their performance than the skilled underestimated theirs; often the unskilled did so quite dramatically

Several studies, in line with the original study, have confirmed the Dunning & Kruger effect where performance can be seen as a predictor of judgement accuracy. In (Bol & Hacker, 2001; Hacker, Bol, & Keener, 2008; Nietfeld, Cao, & Osborne, 2005), high performers showed more accurate metacognitive judgements than low-performers. Similar patterns of miscalibration have been reported in other domains (see (Ehrlinger, Johnson, Banner, Dunning, & Kruger, 2008) for a review). These include real-world settings outside education, such as drivers failing their driving tests (Mynttinen et al., 2009), laboratory technicians quizzed about their work knowledge (Haun, Zeringue, Leach, & Foley, 2000), or chess players during tournaments (Park & Santos-Pinto, 2010).

The purpose of the current study is to both assess the relationship between performance and confidence in a science-related environment and look more specifically at the specific effects of performance level, sex and group composition.

Metacognition

Metacognition, defined as the *cognition of cognition* (Flavell, 1979), can be divided into monitoring and control processes. Monitoring comprises both *metacognitive experiences* and *metacognitive knowledge*. The first category represents metacognitive feelings, such as the feeling of confidence, and judgements. The latter represents declarative knowledge about tasks, skills and strategies. Judgments about performance can anticipate future or assess past performance, i.e. predictive versus postdictive. Performance *predictions,* can be more dependent on general beliefs about one’s competence, not only reflecting the future expected performance in a specific task (Hacker, Bol, Horgan, & Rakow, 2000). *Postdictions* are based on actual and often recent performance and as a result appear to be better indicators of metacognitive skills (Schraw, 2009)(Hacker et al., 2000). Metacognitive judgments further be split into *micro*- and *macrojudgements,* depending on whether they refer to local judgements, that is, to single items; or global ones, referring to an entire test (Nietfeld et al., 2005). Several studies (Gigerenzer, Hoffrage, & Kleinbolting, 1991; Mazzoni & Nelson, 1995; Nietfeld et al., 2005; Schraw & Dennison, 1994; Stankov & Crawford, 1996) showed global judgments as more accurate than local ones, usually reflecting overconfidence (Liberman, 2004), or the *confidence-frequency* effect (Gigerenzer et al., 1991). Moreover, learners' performance itself can be considered a predictor of metacognitive accuracy, as high-performers report more accurately their performance than their low-performer counterparts (Bol & Hacker, 2001; Hacker et al., 2008; Nietfeld et al., 2005) even as the polarity of the judgement tends to be opposite.

a) Confidence

The subjective judgement of confidence is commonly used to study postdictive monitoring processes and is related to specific decisions. Apart from their timing, metacognitive judgements can also be differentiated between item- level or local/micro and test-level global/macro judgments (Nietfeld et al., 2005). Considering confidence as a measure of perceived performance, one can differentiate between overconfident and underconfident individuals. The miscalibration suffered by both of these groups could lead those students to an under- or over-use of provided learning opportunities and to give up a task. Conversely, a learner showing overconfidence, could ignore the provided feedback and continue to attempt to complete a task without changing their strategy, leading, eventually, to giving up. An underconfident learner could in turn expect a low chance of success, which could make them put less effort in a task and give up early. As a result, time on task will be reduced and fewer opportunities to learn explored, lowering their performance and eventually, leading to disengagement. Consequentially, previous literature has highlighted the potential contribution that metacognitive processes could provide to the field of instruction (Georghiades, 2004), which is reflected in the increase of the numbers of studies on metacognition in this field during the last years

b) The importance of assessing confidence in education

The discrepancy between perceived and actual performance is crucial in educational environments, as learners' metacognitive abilities to both monitor their performance and control their behaviour accordingly (Flavell, 1979)) have been shown to be an essential vehicle for self-regulated learning (Chi, 2000; Metcalfe, 2017; Vanlehn et al., 2003). For the learner to be able to assess how well they did or did not do on a task is a vehicle to identify errors and take corrective action and to change strategy. Consequentially, learners with more accurate metacognition have larger learning gains (Rinne & Mazzocco, 2014). On the contrary, errors in metacognition of performance can be detrimental for learning processes (Warren, 2012), as it can lead to a misuse and misallocation of provided resources (Aleven & Koedinger, 2000; Ryan et al., 1998). Hence, measurements of metacognitive abilities to increase our understanding of the learners' self-regulation processes should be included in assessment of education (Efklides, 2006) and an explicit ingredient of a curriculum itself.

A large part of the research on performance judgements has been undertaken with university students (Furnham, Chamorro-Premuzic, & McDougall, 2002; Kisac & Budak, 2014; Mengelkamp & Bannert, 2010). This is not surprising: beyond their easy availability to academic researchers, proper metacognition is especially useful for university students, as they experience more freedom in the regulation of their study than elementary- or high-school ones (Händel & Fritzsche, 2016). Thus, if students judge their performance accurately, they will have more chances of properly regulating future learning processes (Zimmerman, 2000).

c) Metacognitive differences between genders

The aforementioned metacognitive miscalibration has been shown to also reflect a gender effect, i.e. the gender students identify with. Females' perception of their abilities in math-related tasks has been shown to be significantly lower than those of males regardless of their actual performance (Else-Quest, Hyde, & Linn, 2010). Likewise, females tend to attribute successes and failures to their abilities more than males (Dickhäuser & Meyer, 2006). Moreover, previous research has also shown a relationship between gender and self-handicapping behaviour, with women expressing higher levels of self-criticism (Hirt et al., 2003).

The so-called *confidence gap* (Sadker & Sadker, 1995) expresses that female students' confidence in their abilities are consistently and significantly lower than those of their male peers. This gender specific drop in the assessment of own competence, which usually emerges during adolescence, can influence career choices, consequentially leading female students to avoid pursuing more demanding and high-risk STEM subjects (Halpern et al., 2007; Pajares, 2003; Perez-felkner et al., 2017). However, some authors interpret the existence of gender differences in terms of bias, while others highlight the dependence of that bias on the task domain (Voyer & Voyer Susan D., 2014). For instance, male students seem to report higher confidence in mathematics- and science-related tasks (Pajares, 2002) despite their performance (Stankov et al., 2012). On the contrary, they show lower confidence in their language- and arts-related abilities already in middle school (Im, 2013; Wigfield & Eccles, 1989).

d) Hypotheses

In this study we want to assess the relationship between gender, actual performance and metacognitive skills. In addition, we want to assess the operational impact of this relationship. First, we will seek to replicate the original Kruger and Dunning hypothesise that low-performers will overestimate their performance while high-performers will underestimate it (Kruger & Dunning, 1999) to show that the effect is stable over the last two decades and also present in our sample of participants. Second, that test-level global confidence is a more accurate predictor of performance than item-level local confidence. Third, we hypothesise that females will report less confidence as compared to males with the same performance and that they will be more sensitive to error, showing lower confidence in incorrect responses than males.

3.2 Materials and Methods

a) Overview

The test was provided at the beginning of the introductory class of a Cognitive Systems master, during eight consecutive years (from 2010 to 2018). They were not given any compensation (neither economical nor extra credit) for their participation. They were given 30 minutes to solve the test.

b) Participants

A test was provided at the beginning of the introductory class of a postgraduate master course[1], during nine consecutive years (from 2010 to 2018). Participants received no compensation, e.g. economical nor extra credit, for their participation. A total of 224 students participated in the test. From these, 12 were removed from the final poll of data due to incomplete answers (those with more than 10% of non-answered questions). After that, if any participant had a non-answered item, that missing value was filled with that item's mean value (for the whole group of that year). Thus, this led to a total of 212 students being included in the study (age: 26.1 ± 4.16, 83 females).

[1] Cognitive Systems and Interactive Media master of University Pompeu Fabra, http://csim.upf.edu

c) The test

The test contained 15 multiple-choice questions. It must be noted, though, that our test is not a one-to-one mapping of the original one. As we could not obtain the questions of the original study, we used standard questions used in decision-making (see Supplementary Materials). The items were extracted from Massaro's Survey on Psychological Literacy (Massaro, n.d.), with a range of three to five answers provided, and only one correct. The types of questions could be divided among the following categories:

- Logic, i.e. syllogisms (5 items);
- Bias, i.e. examples of the classical cognitive bias studies (5 items);
- Statistics (3 items);
- Memory (2 items);

d) Experimental Protocol

After providing the answer by marking it, they had to report their confidence on their answer providing a number from 0 to 100. They had been previously instructed that 0 represented "not confident at all" and 100, "totally confident". Moreover, they had to report a "general confidence" measure at the end of the test, also from 0 to 100. They were given an hour to solve the test and they were told that it was not a graded activity (to avoid provoking test anxiety). Students were debriefed on the purpose of the study in the following class and given the correct answers, which were discussed together.

Table *2* shows a comparison between the original study and this one. As it is shown, our study focuses on the logical reasoning part of the original study, increases almost by five times the number of participants and has similar number of questions (although ours is shorter). We chose the logical reasoning part because not all our participants have English as their first language, so the English grammar and humour sections could lead to individual differences related to the participant's nationality. In comparison to the original experiment, which had 45 participants, our study presents 212 subjects. As for their studies, our research presents similar context to the original one, as our participants are students of a cognitive science master's degree, while the original study had Psychology students. Nevertheless, the main difference between both studies is that in the original, participants reported their perceived performance compare to their classmates, while in ours, they report they perceived performance in each question.

Table 2. Comparison between studies

	Original study	**Our study**
Tests	Humour, logical reasoning, and English grammar	Logical reasoning
Participants (N)	45	212
Participants (age)	Undergraduate	26.1 ± 4.16
Participants (background)	Psychology Bachelor	Cognitive Sciences Master
Number of questions	20	15

e) Measures

The measures used in this test are:

- **Performance:** as the number of correct answers in the test;
- **Local confidence:** reported after providing each answer;
- **Global confidence:** after completion of the test;

3.3 Results

A D'Agostino-Pearson normality test of the variables showed that the performance data follows a normal distribution (0.03, p = 0.99); while the confidence data follows a left-skewed (negative skewness) distribution (8.29, p = 0.02) requiring a non-parametric statistical test. Consequentially, performance was analysed through independent samples t-test for differences between groups and

Pearson R for correlations; while confidence was analysed through Mann-Whitney U test for differences between groups and Spearman for correlations. For parametric data, reports between parenthesis represent the mean and the standard deviation; for non-parametric data, the median and the median absolute deviation (MAD), a measure of spread used for non-parametric data. We did not consider age as a variable due to the skewness of the data. Although our age range comprises between 21 and 44 years old, the majority of the sample falls around the median of 23.56 years old (MAD: 5.79). The analysis was carried out in Python; specifically, using the libraries *scipy*[2] and *astropy*[3].

a) General results

Low-performers overestimate their performance, high-performers underestimate their performance, and global confidence is more accurate than local confidence

Following the analysis of the original study, we have categorised participants' accuracy into quartiles; with performers in the bottom 25th percentile and high performers in the top 25th percentile. When comparing perceived against actual accuracy, mean confidence remains similar (between 40% and 60%) among performance quartiles (***Figure 4***). On average, participants rated their performance at around the 50th confidence percentile (median local confidence: 74.56%±12.49%), while their actual performance falls around 38.71%±12.50%. The low-performing participants with performance in the bottom quartile (n = 53), largely overestimated their performance in comparison to the rest; as they on average rate their performance at 47.33% while actually showing an average correct score of 12.98% (***Figure 4***). This result is consistent with the original result, but also higher than the mean confidence itself (Wilcoxon, Z=86.0, $p < 0.001$) believing they were above average. Those in the top quartile (n = 53) underestimated their ability: they

[2] https://www.scipy.org/
[3] https://www.astropy.org/

perceived their performance in the 47th percentile despite actually being in the 86th (Wilcoxon, Z=35.0, $p < 0.001$).

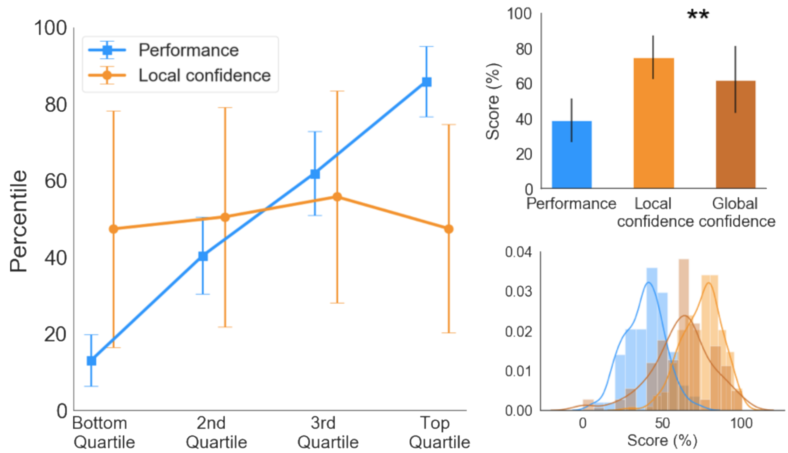


Figure 4. Comparison between performance and confidence. Left: Perceived local performance (orange) as a function of actual test performance (blue). Right, top: Mean values for performance, local and global confidence. Right, bottom: Distribution of Performance, Local confidence and Global confidence scores. ** = $p < 0.001$

b) Results on gender

Females will report less confidence than males with the same performance and females are more sensitive to error, showing lower confidence in incorrect responses than males

In the original study, participants' gender was not considered as a variable, as it "failed to qualify any results" (Kruger & Dunning, 1999). In our case, we do see differences in participants' assessment of their own performance depending on their gender (***Figure 5***). We observed differences between genders in performance (females: 36.3 ± 12.51, males: 40.26 ± 12.24; independent samples t-test: -2.26, p = 0.02) and differences in local confidence (females: 71.42 ± 12.74, males: 76.57 ± 11.90; Mann-Whitney U: -2.98, p = 0.003). When divided among quartiles, we can see that although having similar patterns in the bottom quartile and second quartile, there are large differences between genders in local confidence reports in the second and top quartiles. Females always place themselves under the 50th percentile regardless of the quartile they are in, while males systematically place themselves above it. Moreover, female students' performance correlates with the ratio females-males per group, i.e. female performance is higher when more female students are in the class (spearman R = 0.933, p < 0.001) (***Figure 6***).

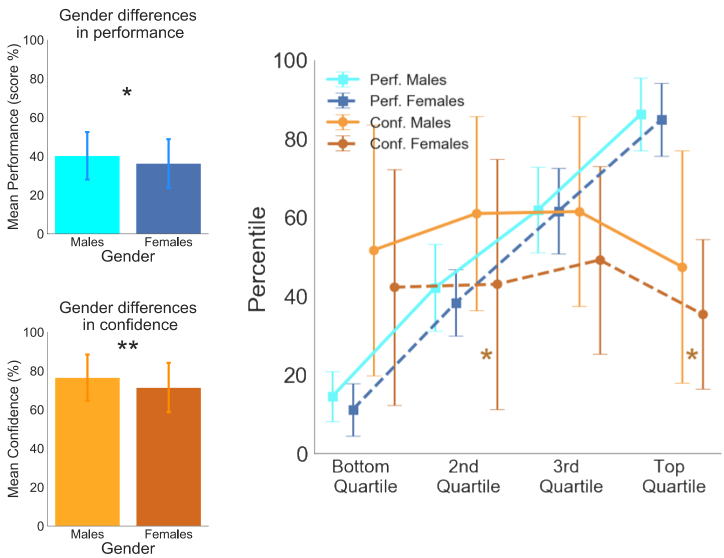


Figure 5. Differences between genders in performance and confidence, overall and divided by performance quartiles. Left, top: Difference between genders in performance. Males (40.26±12.24) had significantly higher performance (t = 2.26, p = 0.02) than females (36.31±12.51). Left, bottom: difference between genders in confidence. Males (76.57±11.9) showed significantly higher confidence (t = 2.98, p = 0.003) than females (71.42±12.74). Right: Differences between genders in perceived logical reasoning ability as a function of actual test performance. The second and top quartiles show significant differences between genders. Second quartile (Males, 64.97±12.26; females, 34.91±28.53; Mann-Whintney U: 241.0, p = 0.03), Top quartile (Males, 53.77±25.24; females, 88., std; test, p) * = p < 0.05

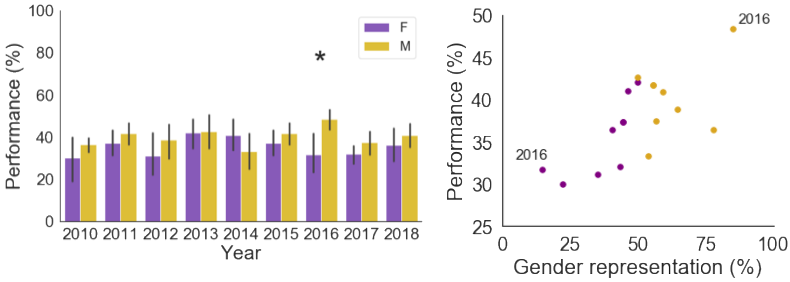


Figure 6. Differences between genders in performance by year and its relation to the class' gender ratio. Left: Performance by year and gender. Significant differences between males' and females' performance in 2016. Right: Correlation between the percentage of each gender in class per year against performance. * = p < 0.05

To measure whether participants assess their performance differently when giving a correct or wrong answer, we computed their confidence discrimination, i.e. to what degree do confidence ratings for correct answers differ from confidence ratings for incorrect ones (Boekaerts & Rozendaal, 2010). ***Figure 7*** (left) depicts the general difference in confidence reports between correct and incorrect answers (correct: 80.0, ±10.0; incorrect: 73.3, ±10.83; Mann-Whitney U: 17982.0, p = 0.0002). Considering the expected differences in performance perception between genders, we also calculated differences in confidence reports per gender, depending on the correctness of the answer (***Figure 7***, males: 76.7, MAD: 9.52; females: 69.2, MAD: 9.67; Mann-Whitney U: 4068.5, p = 0.002). The difference between the confidence in correct and wrong answers is larger for females (correct: 76.67, MAD: 8.67; wrong: 69.22, MAD: 9.67; Mann-Whitney U: 2495.5, p = 0.001) than for males (correct: 80.0, MAD: 10.0; wrong: 76.67, MAD: 9.52; Mann-Whitney U: 6976, p = 0.012).

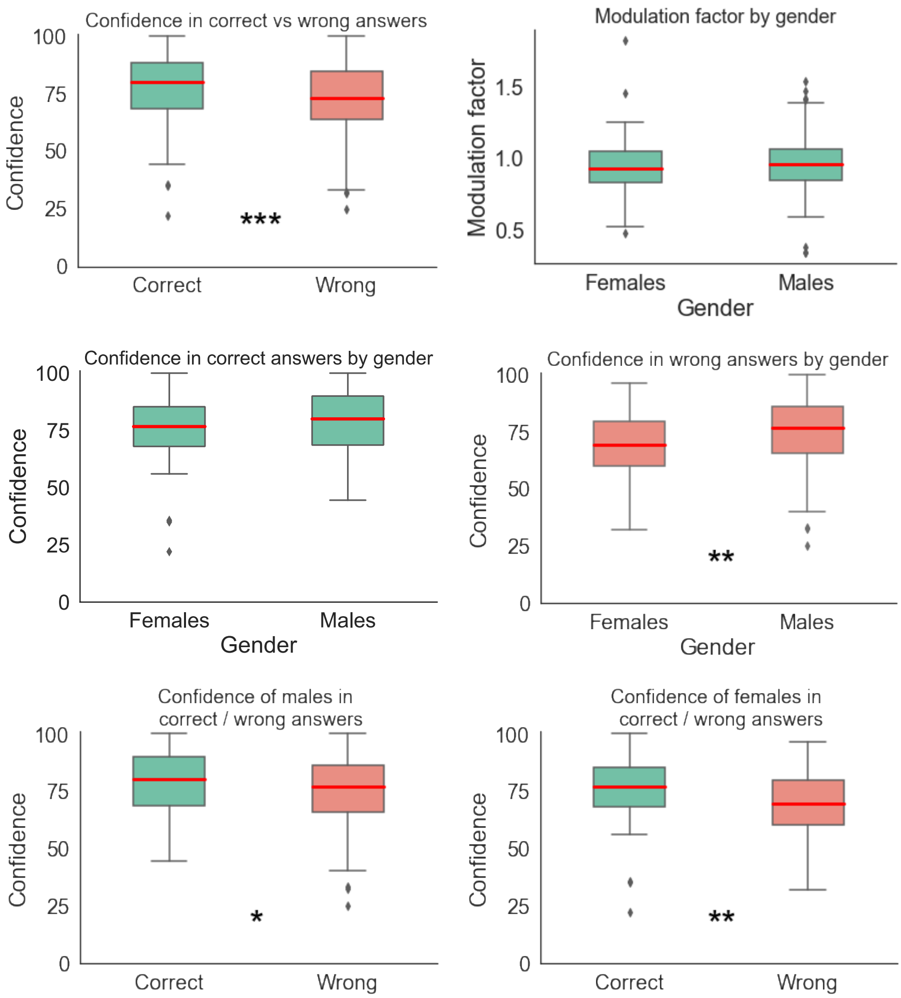


Figure 7. Differences in confidence reports between genders and correct and wrong answers. Left, top: Differences between confidence reports in correct and incorrect items. Correct items are reported with significantly higher confidence than incorrect ones. Left, middle: There are no differences between genders in confidence reports for correct answers. Right, middle: Females report significantly lower confidence than males in wrong items. Left, bottom: There are significant differences in confidence reports between correct and incorrect answers for males. Right, bottom: There are significant differences in confidence reports between correct and incorrect answers for females.

Following the criteria defined in (Aghababyan, Lewkow, & Baker, 2017), we classified participants among four profiles, considering the relation between their self-ascribed and actual performance. To split our data between these groups, we used the criteria of classifying the answers as correct and incorrect and below or above the median local confidence. Like this, answers can be divided in *Overconfident*, i.e. a correct answer with high confidence (above the median local confidence), *Underconfident*, i.e. a correct answer with low confidence (below the median local confidence). When confidence and performance are aligned, we can distinguish between *Knowledgeable*, a correct answer with high confidence (above the median local confidence), and *Realistic*, a wrong answer with low confidence (below the median local confidence). After that, we sum the number of each answer type by participant and classify him/her given the answer type with a higher frequency. ***Table 3*** represents this distribution for our sample. In general, participants show a tendency for overconfidence. When divided by genders, males fall in the overconfident profile while females, under the realistic one.

Table 3. Participants' profile by metacognitive classification (based on (Aghababyan et al., 2017)).

Confidence Profile	Count	Percentage	Males (perc. from profile)	Females (perc. from profile)
Realistic	70	33.02	40.0	60.0
Under-confident	5	2.36	20.0	80.0
Over-confident	87	41.04	73.56	26.44
Knowledgeable	34	16.04	91.18	8.82
Other	16	7.55	31.25	68.75

The questions of the test can be equally divided by their type, i.e. a third of the questions are syllogisms; another third, logic problems; and the other third, questions based on memory or computing probabilities. When comparing between question types (***Figure 8***), we can see large differences in performance between these three types of questions (syllogisms: 20.0, MAD: 20.0; logic problems: 40.0, MAD: 20.0; other: 40.0, MAD: 20.0; Kruskal-Wallis χ^2= 177.75, p < 0.001). More specifically, there are differences between the syllogisms and the other two types (Mann-Whitney U of logic problems: U = 7171.5, p < 0.001; Mann-Whitney U of memory and computation: U = 9699.0, p < 0.001). Nevertheless, there are no differences in local confidence reports among the three types of questions (syllogisms: 76.0, MAD: 14.0; logic problems: 78.0, MAD: 10.0; other: 78.0, MAD: 10.0; Kruskal-Wallis χ^2= 3.06, p = 0.22).

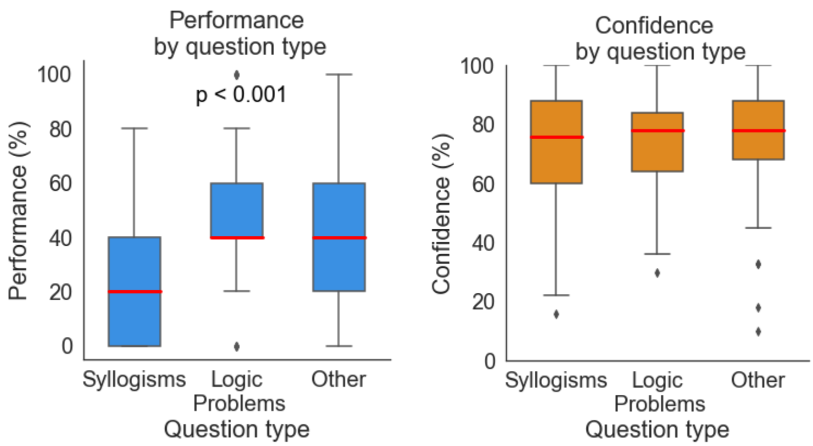


Figure 8. Participants' performance and confidence by question type. Red bars represent the median. Left: Significant differences in performance by question type. Right: No differences in confidence by question type.

3.4 Discussion

This study, based on the original Kruger and Dunning task, assessed the relation between performance and confidence in a logic test. Our participants, a total of 212 master students, answered a 15-items test and were requested to provide the confidence in their answer. ***Table 4*** shows a comparison between the original results and the ones of this study. We first hypothesised that low-performers would show larger perceived performance than the actual one and that high-performers would perceive their performance as lower than the actual one. We also hypothesised that that global confidence (reported for the whole test) would be more accurate than local confidence (reported after each item). Finally, we wanted to address the analysis of gender differences missing in the original study. We hypothesise that females will report less confidence than males with the same performance and that they will be more sensitive to error, showing lower confidence in incorrect responses than males.

Table 4. Comparison between the results of the original study (Kruger & Dunning, 1999) and this one.

	Original study	**Our study**
Bottom quartile	Overestimated their abilities the most, by approximately 50%	Similar results
Top quartile	Underestimated their abilities by about 10-15%	Underestimated their abilities by about 40%
Gender	No gender differences reported (not even gender distribution)	Differences in confidence and performance

a) General discussion

As in Dunning and Kruger, we report a similar pattern, with participants showing a tendency for overconfidence. Moreover, we used the Item-Level Metacognitive Discrimination method to analyse differences in confidence reports between correct and incorrect answers. Our results show that participants significantly reported higher confidence in correct answers than in wrong answers. Nevertheless, we should highlight the smaller difference between reports for both types of responses, what is in line with the tendency for overconfidence.

As occurred in the original study, the discrepancy between actual and perceived ability is more prominent in the lower quartile (the "unskilled") than the upper one (the "skilled"). Those in the bottom quartile failed to recognise their inability, overestimating themselves, almost above the mean. Similar results were found in (Ehrlinger et al., 2008), where it was more difficult for participants in the bottom quartile to estimate their actual accuracy correctly. This effect was also seen in other fields, as (Pavel, Robertson, & Harrison, 2012), with similar patterns found in aviation studies, where students who low-scored in a FAA (Federal Aviation Administration) test showed higher pre- and post-estimates of their performance than their more knowledgeable peers. Akin regularities resulted in (Grant et al., 2009), where low-performance students self-reported higher computing skills than their actual ones. Other examples comprise driving performance (Kidd & Monk, 2009) and medical knowledge in doctors (Mehdizadeh, Sturrock, Myers, Khatib, & Dacre, 2014).

Following the same line, participants in the top quartile also shown a mismatch between perceived and actual performance, placing themselves around the mean. Similar to the original experiment, students in the top quartile underestimated their ability. Nevertheless, in the original study, this underestimation differed from the actual data by 10-15%, whereas in our results, this discrepancy is larger, around 35%. Moreover, the lack of differences in local confidence between question type supports the idea that confidence does not depend on the type of question but on the group comparison.

b) Discussion on gender differences

Contrarily to what generally reported in literature, we found differences in performance between genders. One explanation could be, as (Keller, 2007) mentions, that females' performance could have been affected by the design of the test, where they had to report their gender before answering the questions. Another possibility could be the large discrepancy in the ratio between males and females in the class. This is further validated by analysing the correlation between female-to-male ratio in class and performance, seeing that, the smaller the difference in number, the higher the performance in females. Similar results have shown that for women both the performance in (Gneezy et al., 2003) as well as the selection into competitive environments (Balafoutas & Sutter, 2012; Niederle et al., 2013) is sensitive to the gender composition of the group.

Results show a larger difference between performance and confidence in females, both when assessed by percentiles and generally. This difference is also shown between confidence reports for wrong answers between genders. In line with the literature about larger error-sensitivity in females (Beyer, 2015), females felt less confident than males when answering a question wrongly. When considering this difference by gender, we can see that there are differences between genders in confidence reports for wrong answers. Females report significantly lower confidence than males after providing an incorrect answer. This is in line with females showing larger error sensitivity than males.

3.5 Conclusion

The original Unskilled-and-Unaware study ignited further research on this effect in different settings and domains (Ehrlinger et al., 2008). Nevertheless, not all previous studies found a similar effect (Hacker et al., 2000; Hartwig & Dunlosky, 2014). These different results may come from differences in group performance levels, as students' classified as low- or high-performing depends on the rest of the class. Despite the Dunning-Kruger effect being known as unskilled being unable to perceive their own performance, another explanation has also arisen in the literature. Poor performers may possess the metacognitive abilities to recognise their incompetence; nevertheless, their self-enhancement needs may lead them to ignore it when the domain of that ability is not important for them (Kim, Chiu, & Bregant, 2015).

In the original study, participants rated their "general logical ability" as higher than the one of their peers (placing themselves, on average, in the 66^{th}, higher than the actual mean of 50) and also their percentile rank on the test. Given the comparative nature of percentile ranking, the original study poses two possible sources for that miscalibration: either participants overestimated their abilities or underestimated the ones of others. To solve that issue, they also asked participants to estimate their perceived number of correct answers in the test. Participants did not overestimate the number of questions they answered correctly. This reflects a more erroneous peer assessment rather than erroneous self-assessment.

This more accurate calibration of the number of correct questions goes in line with our results, as participants showed more accurate global confidence (the general one for all the test) than their mean local one (the confidence reported after each question). The pattern of our results for the bottom and second quartile is similar to the one of the original. Nevertheless, our results are different in the third and top quartile. This can be related to the original study asking participants to rate their abilities in comparison to their peers and ours, to rate them on their own. One possible explanation would be that participants in the bottom and second quartile perceive their abilities in similar ways, regardless if being asked to compare to others or not. Contrarily, students in the third and top quartiles may be more influenced by their "position" in the class to evaluate their abilities.

a) Strengths and limitations of the study

The main strength of this study is the sample size. In comparison to the original research, our study has about four times more participants than the original study. This allowed us to run, not only the general analysis but more specific ones, considering, for example, participants' gender or differentiating among question types. Another strength is the study of gender differences, which were not present in the original research.

One of the main limitations of this study, as happened in the original one, is the sample profile. The use of student sampling presents a biased sample despite its benefits (as easy access and low cost -if any- of data collation), which may not be representative of the general population. Another limitation could be the design of the test, which, as previously mentioned, requested students to report their demographics (age and gender) at the beginning of the test. This could have an effect on females' performance and confidence reports, as it has been shown to have an impact on how females answer questions (mostly, if these contain science-related topics). This is known as the "stereotype threat", which refers to individuals' fear of being prejudged due to a negative stereotype about one's social group (Steele & Aronson, 1995). For that reason, new versions of the test ask students for their demographical data at the end.

Something to consider is that in the original study, participants did not specifically report their confidence per answer, as they did here. Firstly, they completed the test before answering the questions about their perceived performance. They first compared their "general logical reasoning ability" to the ones of their peers by providing their (perceived) percentile ranking; they then estimated how their test score would compare with that of their peers and assessed the number of questions they thought they had answered correctly. It must be noted that some studies (Gignac & Zajenkowski, 2020; Krueger & Mueller, 2002) have studied if the magnitude of the Dunning-Kruger effect could be much smaller than reported previously as a result of a statistical artefact. These studies explain how the effect is influenced by the *better-than-average* effect, that is, people considering themselves above average across several skills and abilities (Mabe & West, 1982). In our case, we think our study should not be that highly influenced by the *better-than-average* effect, as our participants reported the confidence in their answers, not how well they did it in comparison to their peers. Nevertheless, it is something to consider for further research on the topic.

We would also like to comment on the terminology used for the classification of the profiles. More specifically, the participants who were classified as *realistic* and *knowledgeable*. We discuss a possible negative predisposition that may not have to do with their performance directly. How can we distinguish from a person that has good metacognition or is just negatively (in the case of the *realistic* profile) or positively (in the case of the *knowledgeable* profile) biased towards reporting lower or higher confidence, respectively? One way of controlling for this would be using personality scales (see, for example, the Big Five questionnaire). On another note, the fact that a large number of the confidence reports are in around the 50^{th} percentile (for confidence, we have to remember that), could that be an indirect comparison to the rest of the class? Our results show a constant value for confidence reports and a variable performance. Moreover, performance (mostly in females) seems to be affected by group ratio but not self-assessment. Could the reported confidence be a prior?

4. STUDY 2: How confident are you? A study on feedback and gender

The previous study showed results on the discrepancy between actual and perceived performance and their differences depending on students' gender. Nevertheless, that study did not include any feedback, as participants were not informed about the correctness of their answers. With that in mind, we wanted to carry a study exploring those three variables (performance, confidence and gender) in relation to feedback.

4.1 Introduction

The improvement of students' metacognitive abilities, that is, the ability to accurately assess one's skill level and to update one's internal model of those skills, has been considered as a critical aspect of how to advance educational paradigms (Bransford et al., 2000; Desoete & De Craene, 2019; Schneider, 2008). This view is not surprising: helping students to become aware of their errors promotes deeper information processing and active engagement in updating their mental models (Chi, 2000; Metcalfe, 2017; Vanlehn et al., 2003). Also, poor metacognition may lead to an under- or over-use of the provided support. On the one hand, students needing more help are the ones less prone to ask for it (Karabenick & Knapp, 1988; Puustinen, 1998; Ryan et al., 1998). On the other hand, students having enough knowledge to solve the task by themselves may anyway ask for support to ease their work without trying to understand the relation between the question and the answer (Aleven & Koedinger, 2000). As a result, we face a vicious circle; learners need to improve their metacognition, but they may not show the appropriate metacognitive skills to self-regulate their learning process.

The main goal of this study is to understand what affects metacognition, in this case, represented by students' confidence reports. Given the role of metacognition in learning, in this study, we first wish to elucidate the relationship between confidence and performance in an educational task. To do so, we designed an educational scenario based on the Piagetian Balance Scale, where we asked learners to report confidence about their performance. This study continues previous knowledge to develop a learning tool able to adapt to each individual and use the appropriate strategies to present the various knowledge elements, according to the needs, preferences, and performance of each learner (for more information about this research, check: Reidsma et al., 2016; Vouloutsi et al., 2016). Here, we are mainly interested in assessing the required inputs that would potentially allow for this adaptation to the user's metacognitive state.

a) The role of confidence, feedback and gender in metacognition

Metacognition commonly refers to the process of monitoring one's thoughts and controlling the allocation of mental resources. Flavell defines metacognition as "cognition about cognition" (Flavell, 1979), as it encompasses skills that allow learners to understand and monitor their cognitive processes (Schraw, Crippen, & Hartley, 2006). It enables the student to self-regulate their learning (Fetsco, Thomas, 2005), and it has been previously demonstrated as a cornerstone to promote a greater understanding of the taught content (Aleven & Koedinger, 2002). Although there exist different dimensions characterizing metacognition, one of the most prominent is represented by confidence. Confidence, a feeling coming from the subjective assessments of how well a cognitive task is being carried out, represents an individual's internal measure of certainty that a just-made decision was correct (Jonsson & Allwood, 2003). Two main aspects characterize confidence: it has to be decision-specific, and it is reported after a decision has been taken.

The confidence manifested by learners on their capabilities has been found to correlate with the way they approach a learning task. On the one hand, it positively correlates with understanding and acquiring meaningful knowledge and learning through the elaboration of ideas and the use of critical thinking. On the other hand, confidence negatively correlates with learning based on mere memorization and with motives far from the task's purpose (Geitz et al., 2016). Errors in metacognition are defined by the discrepancy between subjective and objective performance. Underconfident learners respond correctly with low confidence, whereas overconfident learners answer wrongly with high confidence.

The role of confidence in learning improves when paired with feedback. Low-confidence correct responses enhance retention, as they create a feeling of surprise in the learners by enabling them to strengthen the association between cue and response to inhibit any competing responses (Butler, Karpicke, & Roediger III, 2008). Incorrect answers accompanied by high confidence also surprise the learners, which in turn leads to better content retention as they pay more attention to feedback. This discrepancy between actual and perceived performance can help to correct high-confidence errors by promoting what is called the *hypercorrection effect* (Butterfield & Metcalfe, 2001). That is, errors accompanied by high confidence are more likely to be adjusted when receiving feedback than those reported with low confidence. Thus, feedback promotes metacognition by reinsuring or by contradicting the theories of the learner. As a result, feedback can improve the resolution and calibration of confidence judgments (Butler, Karpicke, & Roediger III, 2008). If feedback is correctly timed and provided after each attempt, it allows for faster improvement in metacognition (Rawson & Dunlosky, 2007).

Considering how confidence and learning improve when paired with timely feedback (as it allows for the correction of potential errors or the strengthening of already correct decisions), we therefore also want to understand how previously received feedback affects performance and confidence when no more feedback is provided. For that reason, we follow a pre-test, task, post-test design, where participants receive feedback only during the task but not in the pre- and post-tests.

The difference between actual and perceived performance appears to be biased in female students. Before we proceed, we would like to clarify that with *gender*, we mean the gender the learner identifies with. A relationship has been observed between gender and self-handicapping behaviour, with women expressing higher levels of self-criticism (Hirt et al., 2003). Women's perception of their abilities in math-related tasks is significantly lower than those of men regardless of their actual performance (Else-quest et al., 2010). This difference is apparent in females reporting higher levels of math anxiety (i.e., a feeling of tension or fear interfering with performance (Ashcraft, 2002)) and more negative attitudes towards math than males (Frenzel et al., 2007; Goetz et al., 2013). Despite these differences in abilities estimation, no gender differences have been found in levels of math achievement (Else-quest et al., 2010; Hyde et al., 2008; Lindberg et al., 2010; Meelissen & Luyten, 2008).

In educational contexts, this can lead to female students showing a drop in their perception of their competence (Pajares, 2003) for not feeling confident enough to pursue “gender-biased” career paths such as engineering (Perez-felkner et al., 2017). This so-called confidence gap (Sadker & Sadker, 1995), usually emerging during adolescence, influences future academic and career choices, making female students less prone to initiate science-related studies (Halpern et al., 2007). However, no clear consensus emerged, with some scholars reporting significant differences in bias between genders and others claiming that those differences are unstable and dependent on the knowledge domain (Voyer & Voyer, 2014). When differentiating by academic subjects in elementary and high-school years, boys report higher confidence in their mathematical and scientific skills, with this difference being more substantial with age (Pajares, 2002), not frequently following their performance (Stankov et al., 2012). Though, as already stated by some scholars (Dunning et al., 2004), the role of gender in meta-reasoning processes must be further investigated.

Considering the effect of gender in the assessment of one’s abilities, we, therefore, wish to evaluate the role of gender and feedback in confidence and performance improvement in this task. Thus, in accordance with existing literature, we expect to find gender differences in confidence. More specifically, we hypothesize that females will show a tendency to report lower confidence compared to males, regardless of performance.

The information about the learner can be either proximal (variables related to the task, like an individual’s performance, confidence or pace) or distal (representing students’ characteristics, like their gender or age). Both proximal and distal variables can serve as inputs to a cognitive architecture. A cognitive architecture represents a theory of the human brain, including learning, attention mechanisms, and problem-solving skills. Usually, the theoretical framework of a cognitive architecture is consistent across a variety of tasks and has two main functions. A cognitive architecture can help understand the process of learning (for example, how memory organization and problem-solving can work in individuals with determined characteristics). Furthermore, it can serve as the base to create intelligent artificial agents, that either employ an architecture to learn or, in the case of a tutoring system, use that architecture to support students' learning process. It serves as a way to summarize the results of cognitive psychology into a comprehensive computer model.

A cognitive architecture can make use of its internal states to generate online reactive and adaptive behaviours to scaffold students' learning processes. It could allow to not only generate the system's actions but to predict the ones of the learners, understand their internal states based on previous or current interactions or infer possible reasons for their behaviour. A cognitive architecture can thus learn from the student it is interacting with, providing individualized content to the learner, adapted to their needs and capabilities. Considering this, we will later relate the results of this study to DAC's architecture in Chapter 9.

Our goal is to understand how metacognition, and more specifically, as indicated by the *resistance* learning phase, confidence, as well as performance, play a role in learning contexts. For this reason, we aim to assess the relationship and differences between performance and confidence reports in a mathematical task. Although the proposed educational task is not adaptive to the learners' needs, based on the DAC architecture, we attempt to control *confusion* by gradually increasing the difficulty of the task. To analyse the relation of performance to confidence, we asked learners to report their confidence (as a "metacognition variable") in each answer in a trial-by-trial base.

Our research questions are:

- Which is the relationship between participants' metacognitive skills and their performance?

- How does metacognition and performance during the task (while receiving feedback) relate to initial metacognition and performance (in a pre-test, when no feedback is provided)?
- What are the changes in performance and metacognition after the task?
- Are there gender differences in these variables? If so, how do the differences, as suggested by the literature, between genders in confidence reports (but not in performance) relate to differences in performance improvement?

To answer the first question, we asked the participants to report their confidence in their answers both at the intervention (on a trial-by-trial basis) and during the pre- and post-intervention sessions. To answer the second question, all participants filled a pre-test questionnaire before the task to know participants' previous metacognition and performance. At the same time, their performance and confidence were also tracked during the task. To answer the third question, a post-test was provided after the task. To answer the fourth question, the participant's gender was considered in the analysis.

In the following sections, we present the educational scenario devised to analyse the relationship between confidence and performance along with the metacognitive-related variables employed. Following the pedagogical model grounded in the Distributed Adaptive Control (DAC) architecture (Verschure, 2012; Verschure, Voegtlin, & Douglas, 2003; Vouloutsi et al., 2016) where the *resistance* phase reflects high-confidence errors, and the Dunning-Kruger effect (Kruger & Dunning, 1999), we expect that low-performance individuals will show higher confidence and, therefore, overestimate their abilities.

4.2 Methods

In this section, we present the educational scenario employed to teach children about weight and distance, based on the classical Piagetian Balance Scale task (Piaget & Inhelder, 1958; Siegler et al., 1981). The purpose of the educational scenario is to introduce the student progressively to concepts that are key to understanding how to solve the Balance Scale problem, such as the role of weights, distance, and their interaction. The task consists of discrete parts including exercise presentation, the correct placement of the weights, assessment of the outcome of the scale, report of the user's confidence, and receive feedback. As a task, it also served for the validation of a content delivery platform, part of complementary research on the development of an adaptive learning system, what is outside the scope of this current work. Here, we focus on assessing the relationship between confidence and performance, and it is affected by gender and feedback.

a) The Balance Scale task

The usage of the Balance Scale task in the present work is twofold: on the one hand, it constitutes a simple inquiry-based learning task, where children's performance can be fully described through the application of the hierarchy of rules of increasing complexity that can be operationally controlled. On the other hand, the Balance Scale problem involves reasoning about physical properties (such as weight and distance) that can be exploited through different means (e.g., manipulation of physical objects and/or graphical material).

For this study, we focused on the first three rules of the original task, as they have been claimed to be the ones fitting our age-range (8-9 years old) (Jansen & van der Maas, 2002). As seen in (Hardiman, Pollatsek, & Well, 1986), we included an additional rule where there can be two weights per side instead of one (with the idea of, in future versions of this research, generate "half-baked" exercises, that is, incomplete exercises where the learner has to finish the setup). Thus, our version of the Balance Scale task comprises these rules:

- **Rule I:** The outcome of this task is defined by the difference between two weights (one per side), placed at the same distance from the fulcrum. The student has to understand that the scale will fall to the side with the greatest weight or remain in equilibrium if each side has the same weight.

- **Rule II:** Here, the weights are the same on both sides of the scale; what determines the scale's movement is the distance from the fulcrum: the scale will fall to the side with the largest distance to the fulcrum or remain in equilibrium if both weights have the same distance.

- **Rule III:** Here, different weights can be placed at different distances to the fulcrum. To solve this setup, children need to merge both previous rules, taking into account both the weight and distance. In this rule, there is only one weight per side.

- **Rule IV:** In this rule, there are two weights per side. The outcome of this setup comes from calculating and comparing the products of weights and distances per each side. This rule allows us to assess differences in performance between having one weight per side or two weights per side.

One factor influencing rule-learning is the discrepancy between existing and targeted knowledge (Siegler & Chen, 1998). Children who already make use of Rule I are more likely to acquire a more complex rule if presented with a setup that can be solved by Rule II than if presented with a kind of problem that can only be solved by applying Rule III. This hierarchical exploitation of rules, therefore, allows for more control over the different stages of learning postulated within the DAC framework.

b) Experimental Procedure

The sample consisted of 63 participants (31 female), between nine and ten years old (M = 9.49, SD = 0.53). All children were in the fourth year of two elementary schools in Barcelona (Spain). The local ethical committee (Comitè Ètic d'Investigació Clínica, CEIC - Parc de Salut Mar-) approved this study. The participants' parents received information sheets together with consent forms that all the children had to bring signed before taking part in the experiment. Despite not being the main aim of this paper, one part of the study was to assess possible differences in the learning outcomes between the three different types of content presentation tools where we compared three conditions: a physical balance, a virtual reality balance (content was provided using a handheld device) and an augmented reality balance that utilized the device coupled with a motorized balance. The timeline of the experimental protocol is illustrated in ***Figure 9*** (bottom).

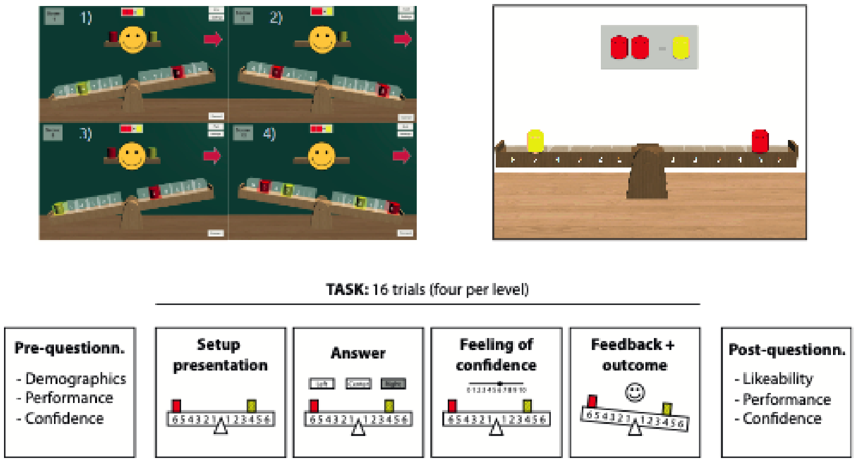


Figure 9. Representation of the task rules, example and experimental protocol. Top left: Representation of the four rules of the task. The images come from screenshots of the handheld device. The number represents each of the rules of the task: 1) Use of weight, 2) Use of distance, 3) Use of weight and distance together, 4) Use of weight and distance together with two weights per side. Top right: Example of the images appearing in the questionnaire. Below them, participants would be requested to report where the balance would fall and their confidence in their answer. Bottom: Illustration of the experimental setup procedure consisting of three main parts: the pre-task questionnaire that included demographics, a few exercises of the Balance Scale task that assessed the existing knowledge of the children, and for each exercise children had to report their confidence. The main task consisted of 16 trials of increased difficulty (4 trials per rule) in which children also had to report the confidence in their responses. Similarly, the post-questionnaire consisted of six exercises of the Balance Scale task, and children had to report their confidence. Additionally, participants had to answer, using a 5-point Likert scale, if they would do the task again, if they would suggest it to their friends and if they liked it. The pre- and post- questionnaires were not followed by feedback.

At the beginning of the experiment, the participants had to fill in a pre-task questionnaire consisting of demographics (like age, gender, etc.), their preferred activities, and access to technology (like smartphones or tablets). Additionally, to assess their previous knowledge on the topic, children filled in a questionnaire related to the Balance Scale task. This questionnaire consisted of five questions where a setup was presented (see ***Figure 9***, Top Right), and children had to report the outcome of the balance: whether it would fall to the left, right, or remain in equilibrium. Additionally, they reported their confidence in their answer (11-point Likert scale where 0 was “not sure at all” and 10 was “totally sure”). A similar test (same test but with different exercises) was provided at the end of the experiment, to measure how much their performance and confidence had varied after the task.

Both the pre- and post- questionnaires were made through Google Forms and provided in a tablet. Participants did not receive any feedback about their answers. Before the pre-test questionnaire, we presented children the two different weights that they would see during the whole study: a red weight and a yellow weight (the yellow was twice as heavy as the red). We ensured that children understood the relation between the two weights by asking “if two red weights are as heavy as a yellow one, which one is heavier?”.

The main task consisted of 16 exercises of increasing difficulty (four for each rule). Each exercise provided a setup that the learner had to recreate (place the appropriate weights at the corresponding position). When the weights were set correctly, learners had to predict the scale’s outcome (left, balance, right) and report their confidence (11-point Likert scale where 0 was “not sure at all” and 10 was “totally sure”). Despite using a 0-10 scale to ask for confidence self-reports may be perceived as a very fine discernment of levels, the format of the scale was decided (together with the teachers) to facilitate students’ understanding, similar to the range used in school grades, as done in (Foster, 2016).

Upon the exercise completion, learners received feedback about their answer by viewing the scale’s outcome. We chose to provide feedback to every answer to ensure trial-by-trial monitoring. Finally, in the post-task questionnaire, participants were asked to solve five exercises, similar to the ones in the pre-questionnaire, again without receiving feedback about them. The exercises in the pre-task questionnaire allow us to evaluate the existing knowledge of each student regarding the task. The post-task questionnaire enables us to assess both the task and the learner’s knowledge acquisition.

To analyse participants’ metacognitive skills, we assessed how accurately they were reporting their perceived performance. To do so, we computed their bias from their answers. The bias represents the “realism of confidence”, the relationship between performance and confidence, and considers learners as being able to properly calibrate their metacognition when the value is zero. The bias is computed by subtracting the normalized performance values from the normalized confidence values. Thus, an overconfident learner will have a bias higher than zero, whereas an underconfident learner will present a bias lesser than zero.

c) Measures

Here we provide a list of the measures extracted in this experiment that are part of the analysis:

- **Performance:** Is the number of correct answers. Obtained from the task and the pre- and post-questionnaires. To control for ceiling effects, the analysis of improvement in performance is not the difference between the performance of the pre- and post-assessment questionnaires. Instead, we consider improvement as a “normalized improvement”, that is, how much could the learner improve from its initial score in the pre-test questionnaire.

- **Normalized improvement:** This measurement represents how much participants could improve given their initial performance (equation (1), represents the total possible score one could obtain in the pre-questionnaire, and depicts the participant’s score in pre-questionnaire). That is, we subtracted the performance (%) to the maximum possible score (100%). Later, we calculated how much of this improvement was achieved (equation (2), representing the participant’s performance in the post-questionnaire). To do so, we calculated the difference between the pre and post questionnaire and multiplied it by 100 and divided by the WoI to obtain the percentage of improvement.

$$WoI = t_{pre} - p_{pre} \quad (1)$$

$$improvement = \frac{(p_{post} - p_{pre}) * 100}{WoI} \quad (2)$$

- **Confidence:** Participants' provided confidence in their answers. We asked participants to report their confidence after each exercise during the task and the pre- and post-questionnaires. It is represented on a discrete scale from 0 to 10, with 0 meaning *Not sure at all* and 10, *Totally sure.*

- **Bias:** represents the difference between the perceived and actual performance of an individual. We computed the bias by subtracting the performance in each trial (1 if it is correct and 0 if the answer is incorrect) from the normalized confidence values (the reported confidence divided by the maximum score of confidence). We calculated the bias for both the task and the pre- and post-questionnaires. This measure extracted from (Fischhoff, 2013). Equation (3) represents this variable (where *n* is the total number of trials; *c*, the confidence reports; and *p*, the performance). Thus, the bias represents the signed difference between mean confidence and mean performance.
$$bias = \frac{1}{n} \sum_{i=1}^{n} (c_i - p_i) \quad (3)$$
The bias represents the "realism of confidence", the relationship between performance and confidence, and considers the learner as being able to properly calibrate their metacognition when the bias is zero. Consequently, an overconfident learner will have a bias higher than zero, whereas an underconfident learner will present a bias lesser than zero.

- **Item-Level Metacognitive Discrimination:** This index represents the difference in confidence for correct items compared to confidence for wrong items by calculating a discrimination score (mean confidence correct items - mean confidence wrong items). It is based on (Destan & Roebers, 2015).

4.3 Results

To explore the effects of confidence on performance during the experimental procedure, we divided the results in pre and post questionnaires, the task's impact on their metacognitive skills, and the metacognition's evolution during the task. We also investigated the role of gender in confidence.

We first computed the participants' normalized improvement (see Measures section) and assessed the difference between platforms. Since we did not find any significant differences in performance improvement among the different platforms (PB: Median = 50, MAD = 50; VR: Median = 0, MAD = 50; AR: Median = 0, MAD = 100; Kruskal-Wallis, H = 1.78, p = 0.41), we merged the datasets from the three conditions into a single pool of data for the following analysis.

a) Results on performance

First, we could observe an improvement in performance after the task (***Figure 10***, Top Left). A Wilcoxon signed-rank test shows a significant difference between the results of the pre- (median: 80.0, MAD: 20.0) and post- (median: 80.0, MAD: 20.0) questionnaires (Z = 326.0, p = 0.028). Nevertheless, as the baseline of all participants was already high, we divided the sample between participants that scored low (less than 50% of the maximum possible score) in the pre-questionnaire and high (more than 50% of the maximum possible score). This division allowed us to evaluate the differences in improvement, depending on the participants' baseline. A Mann-Whitney U test showed a highly significant difference in improvement between the Low group (the participants that scored less than the 50% of the maximum score in the pre-questionnaire) (median: 60.0, MAD: 0.0) and the High group (the ones who scored 50% or more) (median: 0.0, MAD: 20.0) (Z = 77.5, p < 0.001) (***Figure 10***, Top Right). These results suggested that the so-called "low-performers" improved significantly more than "high-performers", as can be expected due to a ceiling effect.

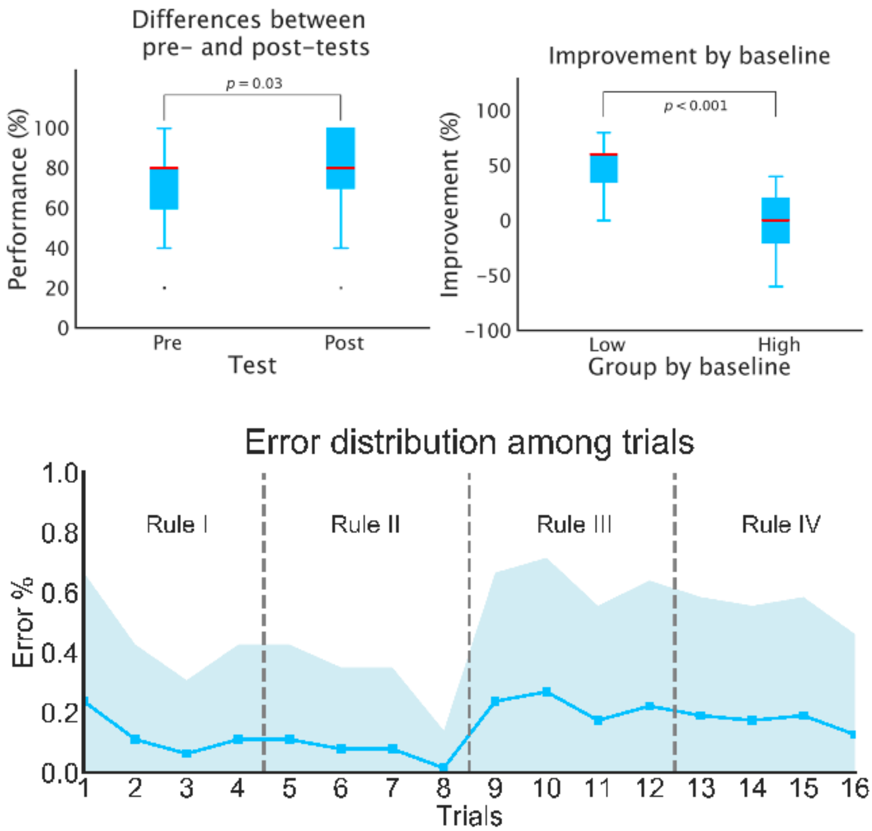


Figure 10. General results on performance: Improvement between pre- and pos-tests and error distribution among trials.***Top left:*** Differences in performance before (pre-questionnaire) and after (post-questionnaire) the task. ***Top right:*** Comparison of the improvement in performance between participants that scored low (less than 50% of the maximum score) in the pre-questionnaire and high (more than 50% of the maximum score). Significantly higher performance in the post compared to the pre-questionnaire. ***Bottom:*** Error distribution among trials. The shaded area represents the SD. The jump from trial 8 to 9 represents the difficulty increase from rule 2 to rule 3, where children merged both of the previously used strategies.

We then looked at the distribution of errors across the experimental session (***Figure 10***, bottom). Despite the general high performance (depicted in the low error distribution), we could see an error increase from trials 8 to 9. This jump represents the transition to rule 3 (in trial 9), where participants begin to face conflicting items.

b) Which is the relationship between participants' metacognitive skills and their performance?

We explored students' metacognitive skills by assessing the accuracy of the reported perceived performance (***Figure 11***, Top Left). A normality test showed that this data was normally distributed. A dependent-samples T-test was run between every pair of results (Pre-Task, Task-Post, and Pre-Post). We found a positive bias (M = 0.022, SD = 0.265) in the pre-questionnaire, which means that participants were slightly more confident about their responses compared to their performance. During the task, we observed a significant decrease in the bias between the pre-questionnaire and the task (M = -0.086, SD = 0.166; $p = 0.007$). Finally, we found a close to significance increase bias from the task (close to significance, $p = 0.063$) in the post-questionnaire phase.

c) How does metacognition and performance during the task (while receiving feedback) relate to initial metacognition and performance (when no feedback is provided)?

To see how individuals' metacognition evolved from the experiment's beginning to the end, we divided participants into two subgroups: overconfident students, those showing a positive bias during the pre-test questionnaire, and underconfident, who had a negative bias in the same test. Splitting participants based on their initial bias revealed a significant difference in reported confidence between groups in the pre-questionnaire results (independent samples T-test, overconfident: M = 0.21, SD = 0.20; underconfident: M = -0.19, SD = 0.15; $p < 0.001$, ***Figure 11***, Top Right). Overconfident participants were significantly more confident than the underconfident ones. These substantial differences were masked as a bias closer to zero (see ***Figure 11***, top left) when all participants were taken into consideration, when, in fact, this was not the case.

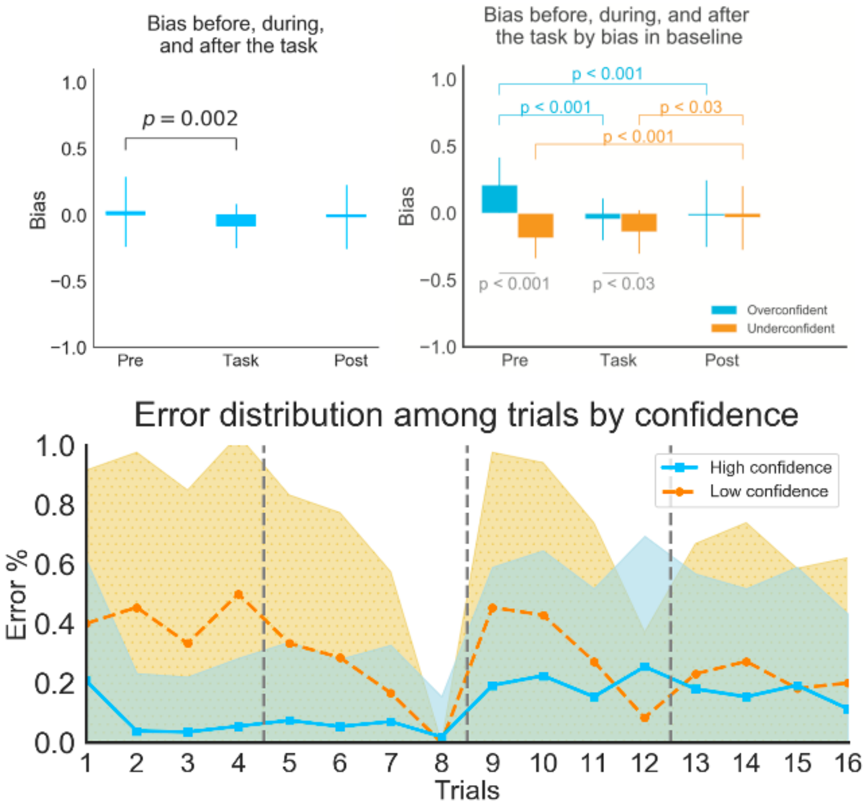


Figure 11. Bias evolution during the experiment and its relation to confidence. ***Top left:*** Participants' bias during the three parts of the experiment. ***Top right:*** Evolution of the bias measurement for the whole experiment when participants were classified as overconfident and underconfident. The results illustrate how the differences in bias in the pre-questionnaire disappear in the post-questionnaire. These results match the DAC's three phases of learning predictions regarding the confidence of the learner. ***Bottom:*** Error distribution among trials by confidence group. Both groups show the aforementioned performance jump from trial 8 to 9. Participants in the high confidence group have better performance and more stable confidence, while participants in the low confidence group show worse performance and more variable confidence.

Overconfident individuals showed a highly significant decrease in bias from the pre-questionnaire to the task setup, where they began to receive feedback about their performance (pre-questionnaire: M = 0.21, SD = 0.20; task: M = -0.05, SD = 0.17; $p < 0.001$, ***Figure 11***, Top Right). For underconfident learners, although their bias did increase (got closer to zero, hence became more accurate), it did not reveal a significant difference from the pre-task questionnaire to the main task. Hence, during the task, overconfident students were significantly more accurate (that is, they can estimate their performance closer to the real one) than the underconfident ones (independent samples t-test, overconfident: M = -0.05, SD = 0.17; underconfident: M = -0.14, SD = 0.16; p = 0.03, ***Figure 11***, Top Right).

The initial significant differences in bias between overconfident and underconfident participants disappeared after the task, as after the intervention, both groups became more accurate in the estimation of their performance. Thus, we observe a significant difference between initial bias and bias after the task, in both overconfident (pre: M = 0.21, SD = 0.20; post: M = -0.04, SD = 0.25; $p < 0.001$) and underconfident individuals (pre: M = -0.19, SD = 0.15; post: M = -0.04, SD = 0.25; $p < 0.001$) (***Figure 11***, Top Right). During the task, we can see participants in the high confidence group showing better performance and more stable confidence. In comparison, participants in the low confidence group show worse performance and more variable confidence (***Figure 11***, Bottom). Moreover, there is a performance jump from trial 8 to 9, representing participants moving from applying rule II to applying rule III.

d) How does the relationship between confidence and performance differ between genders?

To assess the differences between genders in their metacognitive abilities, we divided participants depending on their gender and compared their performance and confidence reports in the pre-questionnaire. Males reported significantly higher scores in their confidence compared to females in the pre-questionnaire (males: Median: 80.0, MAD: 11.6; females: Median: 72.0 MAD: 8.8; p = 0.02, Mann-Whitney U test). However, as shown in ***Figure 12*** (Top Left), we did not observe any differences in performance among genders (males: Median: 80.0 MAD: 20.0; females: Median: 80.0, MAD: 20.0, p = 0.4, Mann-Whitney U test). These differences in confidence remained during the task (p = 0.047), when feedback was provided, with no significant differences in performance (p = 0.14, Mann-Whitney U test).

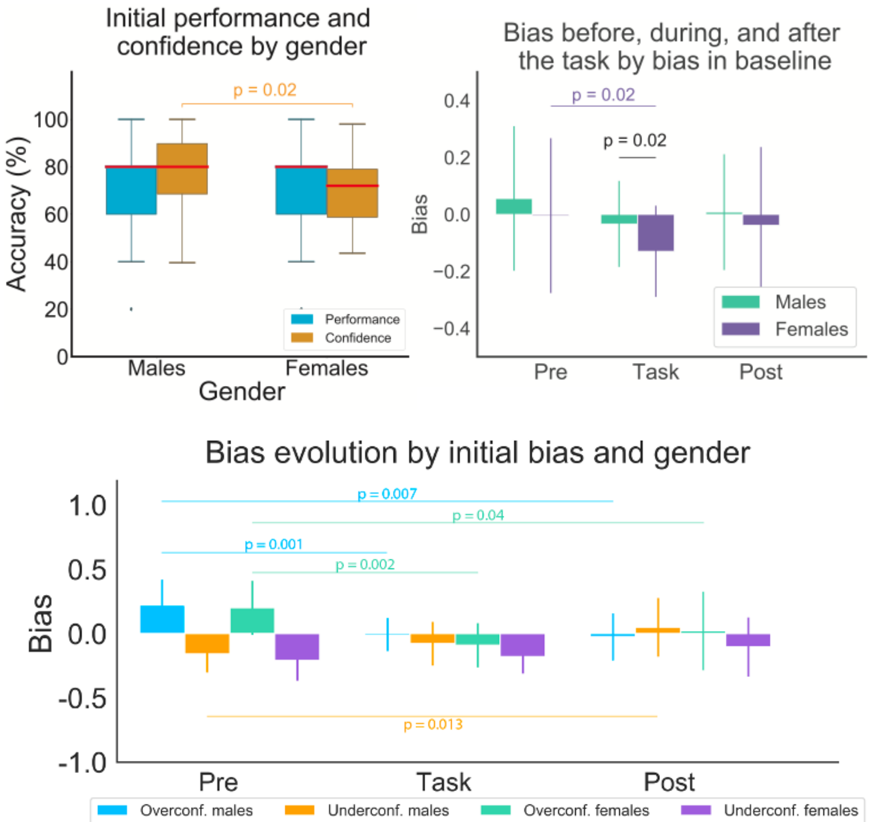


Figure 12. Differences between genders in performance, confidence and bias. **Top Left:** Performance and confidence in the pre-test split by gender. There are significant differences between genders in confidence but not in performance. **Top Right:** Evolution of the bias measurement during the whole experiment, divided by gender. There are significant differences between bias in the pre-test and the task for females, and between genders during the task. **Bottom:** Evolution of the bias measurement among the whole experiment by gender and bias in the pre-questionnaire. In the end, all the groups significantly improve their bias, but the one of the underconfident females, who remain underconfident.

A dependent samples t-test showed significant differences between the pre-questionnaire (M = -0.01, SD = 0.27) and task (M = -0.13, SD = 0.16) biases in females (p = 0.018). There was also a significant difference between genders' bias during the task *per se*, when females had lower bias (M = -0.13, SD = 0.16) than males (M = -0.03, SD = 0.15) (p = 0.02) (***Figure 12***, Top Right).

Finally, we wanted to determine if there exist any interaction among bias and gender. To do so, we analysed how participants' metacognitive skills evolved during the experiment by classifying the learners both by gender and previous bias. Participants were divided into overconfident males (n = 17), underconfident males (n = 13), overconfident females (n = 15) and underconfident females (n = 15). ***Figure 12*** (bottom) represents this distinction. Differences in the pre-questionnaire between over- and underconfident participants were found for both genders (overconfident males: M = 0.22, SD = 0.2; underconfident males: M = -0.16, SD = 0.14; $p < 0.001$) (overconfident females: M = 0.2, SD = 0.21; underconfident females: - 0.21, SD = 0.15; $p < 0.001$).

We found statistically significant differences in bias between the pre-questionnaire and the post-questionnaire both for over (pre-questionnaire M = 0.22, SD = 0.2; post-questionnaire M= -0.03, SD = 0.18; $p = 0.007$) and underconfident males (pre-questionnaire M = -0.15, SD = 0.14; post-questionnaire M= -0.05, SD = 0.23; $p = 0.013$). Additionally, statistically significant differences were found between pre- and post-questionnaire also for overconfident females (pre-questionnaire M = 0.20, SD = 0.21; post-questionnaire M= 0.02, SD = 0.31; $p = 0.04$), but not for underconfident ones. Finally, we found differences between bias before and during the task for overconfident individuals (both males and females) (males: bias before the task M = 0.22, SD = 0.2; bias in task M = -0.01, SD = 0.13; $p = 0.001$) (females: bias before the task M = 0.20, SD = 0.21; bias in task M = -0.09, SD = 0.17; $p = 0.002$) but not for underconfident ones.

e) Confidence differences between correct and incorrect answers. Gender differences in confidence for wrong answers but not for correct ones

To assess participants' perception of correct and wrong answers, we calculated the difference between confidence reports for these two kinds of answers. ***Figure 13*** (left) shows that participants report significantly higher confidence (p = 0.002, Mann-Whitney U test) for correct (median: 8.42, MAD: 1.24) than incorrect answers (median: 7.16, MAD: 1.75). Nevertheless, the range of answers for correct answers covered from 5 to 10 points in the Likert scale (that is, the upper half of the whole possible range), while the one for wrong answers covered almost the entire range of the Likert scale (that is, from 1 to 10, in this case). When split by genders (***Figure 13***, right), both genders report larger confidence for correct than wrong trials. Nevertheless, we can see that males report significantly higher confidence (median: 8.97, MAD: 0.88) than females (median: 8.13, MAD: 1.13) for correct trials (p = 0.03, Mann-Whitney U test), whereas their reports for wrong trials do not differ. Seeing it from another perspective, males differ in their self-reports for correct and wrong answers, while females do not.

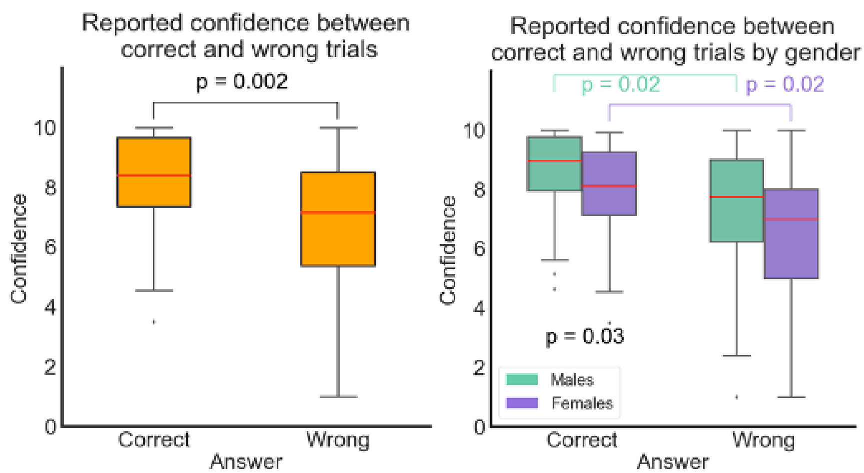


Figure 13. Differences in confidence reports between genders and between correct and wrong answers. **Left:** Differences in reported confidence between correct and wrong trials. **Right:** Differences in reported confidence between correct and wrong trials split by gender.

To further study students' metacognitive abilities, we computed the *Item-Level Metacognitive Discrimination (ILMD)* (Destan & Roebers, 2015), explained in the methods. A Pearson R test showed a significant correlation between ILMD during the task and the improvement in the bias measure from the pre to the post-test (r = 0.28, p = 0.04) (***Figure 14***, left). Moreover, following the method used in (Aghababyan, Lewkow, & Baker, 2018), we assessed the relationship between participants' overconfidence ratio and their average score during the task. Contrarily to the results of the original paper, this resulted in a negative correlation (r = -0.68, p < 0.001) (***Figure 14***, right).

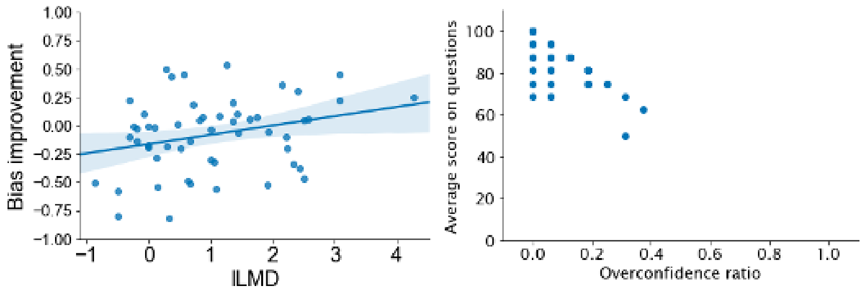


Figure 14. Other measures of metacognition. **Left:** Correlation between participants' ILMD during the task and the bias improvement after the task (comparing between pre and post-questionnaires). **Right:** Correlation between overconfidence ratio and the average score on questions during the task.

4.4 Discussion

The purpose of this study is to understand the relationship between performance and self-reported confidence in a mathematical task. To do so, we designed a scenario based on the Balance Scale task (Piaget & Inhelder, 1958; Siegler et al., 1981) with clearly defined levels of complexity. The scenario consisted of three main parts: a pre-task questionnaire, the intervention, and a post-task questionnaire. The main task's purpose was to introduce the student progressively to concepts that are key to understanding how to solve the Balance Scale problem, such as the role of weights, distance, and the interaction between them. The way the task was formulated allowed users to monitor their performance trial by trial and hence facilitated the extraction of information regarding their metacognitive state. The post-task questionnaire allowed us to evaluate the student's current knowledge on the task and compare it with the pre-task to measure individual improvements. At the same time, it served as a way to evaluate the interaction.

Part of the study (that is not the scope of this work) was to evaluate the three different platforms for content delivery (AR, VR, PB). As we found no differences between students' performance among the three original conditions, we only focused on the student's bias and gender and not the educational content's presentation tool. We can observe a ceiling effect, as most students already performed well from the beginning. Children did not find any major problems in solving the puzzles provided, including those of the third and fourth rules. However, according to Siegler (Siegler et al., 1981), children of 9-10 years old already find problems in assimilating rule III, when in our experiment, this was not observed.

The observed ceiling effect highlights the importance of adapting the content to each student, as the majority of the children that participated in our experiments were able to solve the task despite their young age. At the same time, the pre-task phase allowed us to categorize the learners by their inherent bias and enabled us to understand better how their metacognitive skills evolve during the task.

Finally, we observed a significant change of bias from trial 8 to 9, where there is a change from rule II to rule III. In rule III, the exercises provided include the application of rules I and II; hence they must employ strategies that they have already used in the previous trials. These findings fit well with previous literature on prior knowledge: a possible explanation is that, as stated by (van Kesteren, Krabbendam, & Meeter, 2018) the reactivation of previously learned information during new learning leads to successful knowledge construction.

When observing the difference between perceived accuracy in correct and incorrect trials, we can see that, despite being different, the variability of confidence reports for wrong answers encompasses double the range of the one for correct answers. Males report significantly larger confidence for correct answers than females, compared to confidence in wrong answers. Considering the literature, one would expect to find significantly lower confidence in incorrect answers in females, given their larger error-sensitivity (Beyer, 2015). This lack of difference could come from the larger representation of corrects answers compared to the wrong ones, which could account for lower statistical power.

a) Other variables to explore: Gender

Due to the relationship between gender and bias and the task's science-related nature (which, as mentioned before, can lead to differences in metacognition between genders), we wanted to explore the possible dissimilarities in metacognition between male and female students before, during, and after the task. We observed pre-existing disparities between genders in their metacognitive abilities (as there were differences in confidence reports among genders but not in performance). Similar results were found in (Else-quest et al., 2010; Hyde et al., 2008; Lindberg et al., 2010; Meelissen & Luyten, 2008). These differences evolved during the experiment. The results illustrate that in males, improvement after the task positively correlated to initial metacognition during the pre-questionnaire (when they have not yet received feedback about their performance). Contrarily, female participants' improvement after the task did not correlate to their initial bias but their bias during the task (while receiving feedback about their performance).

The difference between genders in the timing of the correlation between bias and improvement could be related to the stronger effects that negative feedback has on females, as reported in different psychophysiological studies (Ding et al., 2017; Robinson, Standing, DeVito, Cools, & Sahakian, 2010). This effect exists regardless of feedback being representative (that is, being congruent to the answer's correctness) and it has been related to an increased punishment sensitivity (Cross, Copping, & Campbell, 2014; Li, Yuan, & Lin, 2008; Moeller & Robinson, 2010; Schirmer, Zysset, Kotz, & Von Cramon, 2004; Weller, Levin, & Bechara, 2010; Yuan et al., 2009). Moreover, it can also be affected by gender differences in stress responses (Wang et al., 2007). To our knowledge, there are no previous studies reporting differences between genders in the relation between performance improvement and the difference between performance and confidence.

Although not significant, the results showed a difference in bias between genders during the pre-questionnaire. When analysing the classification by both bias and gender, only the underconfident females did not significantly improve their metacognition at the end of the experiment. Moreover, the results regarding the evolution of bias during the task demonstrate significant differences in bias between genders during the first two rules that disappear in rule three, where the exercises merge to the two dimensions previously explored in the two earlier rules.

4.5 Conclusions

Traditional educational paradigms promote learning as a mere knowledge transfer; however, equally important is scaffolding the student's learning process. In other words, the process of acquiring new information is as essential as knowledge acquisition itself. The assessment of confidence in a learning task permits a better understanding of the learning process undergone by the student. At the same time, it allows us to examine the implications of self-perception of one's abilities on knowledge acquisition.

Externally reporting the learners' current confidence is a way of making them aware of their learning process. Our primary goal is not to directly increase the confidence learners have on their ability, as it would be counterproductive if the skills are low. In contrast, our goal is to promote the improvement of students' metacognitive skills to help them to be aware of their needs and achievements. This would help them cope better with failure, as they would not see their knowledge as something static, as happens with learners with a fixed mindset (Mueller & Dweck, 1998), but as something that they can improve. Enabling learners to be aware of their current abilities and properly self-monitor their improvement may serve as a way to set the ground for a proper learning process. Providing learners with content that suits their current needs and skills can fail in increasing their knowledge if they are not aware of those needs. A learner that does not have a proper model of their abilities may not be able to obtain appropriate profit from the provided feedback. Thus, it is essential to provide students with tools that allow them to detect their capabilities. By doing so, they may foster a mastery-oriented approach in which they will look for new challenges whenever they have mastered the previous content.

5. STUDY 3: Metacognitive factors behind rule change in the Balance Scale task

The previous study showed results on the discrepancy between actual and perceived performance, their differences depending on students' gender and the effect of feedback. Nevertheless, that study did only considered participants' confidence as an expression of their metacognitive abilities. With that in mind, we designed a study assessing the effect of other metacognitive and motivational variables in a similar task. Moreover, we assessed their relation to students' exploration patterns.

5.1 Introduction

Rules (or strategies) play a crucial role in developmental psychology and cognitive science in general (Reber, 1993). Nevertheless, there is yet no clear consensus on the optimal method to assess children's rules. One way to measure them is the rule assessment methodology, which uses pattern matching. This method was used by Siegler in the Balance Scale task (Siegler et al., 1981). In this task, children are classified as using one rule or another depending on their response pattern to different problem types.

The aim of the task (presented in the 2.4.c of the Chapter 2, "Pedagogical Framework") is teaching children about the concepts of weight and distance. Given those dimensions, several patterns appear. ***Table 5*** represents the possible rules used by the children depending on the combination of accuracy per item type. Like this, a child using Rule I would answer correctly weight (W) and conflict-weight (CW) items; and a child using Rule II, W, distance (D) and CW items. If they used rule III, they would answer correctly W and D items and guess in CW, conflict-distance (CD) and conflict-balance (CB) items; and a child using Rule IV would answer all items types correctly.

Table 5. Siegler's Rule models and their response patterns by item type. Adapted from (Jansen et al., 2007)

Item types	**RI**	**RII**	**RIII**	**RIV**
Weight	1.00	1.00	1.00	1.00
Distance	.00	1.00	1.00	1.00
Conflict-weight	1.00	1.00	.33	1.00
Conflict-distance	.00	.00	.33	1.00
Conflict-balance	.00	.00	.33	1.00

Previous research has shown that children's improvement in this task can be due to three factors: After seeing movement in the balance scale (Siegler & Chen, 1998) (that is, after feedback), after manipulation (Philips & Tolmie, 2007), after seeing challenging configurations (e.g., "distance items" for children using Rule I) (Jansen, Raijmakers, & Visser, 2007) or after a combination of the above three. Nevertheless, not all children improved in this task, even with the same manipulation and previous knowledge. Consequentially, it raises a question: Why do some children improve on the Balance Scale task and others do not? We propose that children's prior knowledge is not the only factor affecting rule change in this task but that they also may be affected by metacognitive and motivational factors.

Metacognition, the process of monitoring our thoughts and controlling the allocation of mental resources (Flavell, 1979), is considered one of the main factors influencing learning (Wang, Haertel, & Walberg, 1997). It encompasses skills that allow learners to understand and monitor their cognitive processes (Schraw et al., 2006), supporting self-regulated learning (Fetsco, Thomas, 2005). Confidence is one of the most common measures in metacognition. It represents an individual’s measure of how sure they are that a just-made decision was correct (Jonsson & Allwood, 2003). Errors in confidence calibration denote the discrepancy between perceived and actual performance, dividing learners among underconfident (high performance and low confidence) and overconfident (low performance and high confidence). Previous research has shown that learners (especially, males) tend to be in the last group (van Loon, de Bruin, Leppink, & Roebers, 2017). Substantial discrepancies between expected and actual outcomes have been shown to induce exploratory behaviour in infants (L. Schulz, 2015). However, less is known about its effects on mathematical learning and the relation between exploration and other metacognitive and motivational measurements.

a) Metacognition in Mathematics

The discrepancy between perceived and actual ability is also evident in mathematical and science-related learning. Many preconceptions and biases surround mathematical understanding, generating doubts about one's skills even in confident learners (Betz, 1978; Cvencek et al., n.d.; Rubinsten et al., 2012). This is evident in statements as "I am just not a math person" (Dweck, 2008; Rattan et al., 2012). This identification (or not) comes from students' confidence in maths skills and in their ability to learn math, as well as from mathematical anxiety (Ashcraft, 2002; Jansen et al., 2013).

The term "mathematics confidence" was defined by Pierce and (Pierce & Stacey, 2004) and represented "a student's perception of their ability to attain good results and their assurance that they can handle difficulties in mathematics". Part of this mathematics confidence comprises mathematical self-efficacy, that is, a learner's belief in their likelihood of carrying out a mathematical task in a correct manner. Lastly, "mathematical resilience" (Johnston-Wilder et al., 2014) denotes how students overcome barriers to learn mathematics.

b) Exploration

Promoting learners' exploration has been previously shown to be beneficial for their learning process. The main reasons behind it are that it encourages broad hypothesis testing and improves the dept of understanding. The wrestling with the similarities and differences between experiences and prior knowledge (leading to noticing inconsistencies) can help to dispel learners' illusions of competence by helping them to acknowledge their lack of understanding.

Previous research has shown how inconsistency between current evidence and prior knowledge engages learners in building explanations (Cristine H. Legare, 2014). One of the vital parts of this process is exploration, which help learners to generate evidence to disambiguate between possible causal variables (L. E. Schulz & Bonawitz, 2007). This exploration has been shown to be more effective when it is used to explain inconsistent outcomes rather than confirming previous observations (Legare, 2012). This is because when learners' inductive generalisations crash with the current evident, the following exploration is rational, as they look for an explanation to the phenomenon.

It should be noted, also, that despite some guidance helps to support learners' exploration, teaching can also constrain their exploration and discovery. For example, in (Bonawitz et al., 2011), children who were instructed a possible function of a toy later performed less exploration with that toy than the ones who did not receive a demonstration. Considering this, a condition where learners are directly instructed should lead to less exploration than one where they are not instructed, as an instructional condition would lead to explore more the previously seen examples.

c) Active/Passive learning

Another comparison in this study is between active and passive learning. In active learning, students are responsible for their learning process. This kind of learning encompasses various tasks like carrying out quizzes (and receiving feedback) and pausing for self-reflection and to consolidate the content (Bonwell & Eison, 1991; Ebert-May, Brewer, & Allred, 1997; Sarason & Banbury, 2004). It provides several benefits compared to passive learning: learners are more involved and engaged in activities and more motivated (Bonwell and Eison (1991)).

Contrarily, in passive learning, students have little opportunity to provide their input through discussion or exercises (Stewart-Wingfield & Black, 2005), as they passively receive the information from the teacher and internalize it by memorization. In this kind of learning, students seem to decrease their attention, as observed by many educators in their classes (Dorestani, 2005), which, at the end, can lead to learners retaining less material than in an active learning class (Van Eynde & Spencer, 1988).

Seeing the effect of self-beliefs in learning, we consider several possible metacognitive and motivational factors that could account for a student not moving to the following rule, presented in the next section.

5.2 Metacognitive and motivational factors

Here we propose four metacognitive and motivational factors that could be affecting rule change (or staying in the same rule). These are mathematical anxiety, perceived competence, goal orientation and intolerance of uncertainty.

a) Mathematical anxiety

Mathematical anxiety (MA) is represented by a state of discomfort around the performance of mathematics tasks (Ma & Xu, 2004). It represents the feelings of tension when people are faced with mathematical problems, regardless if it is in a school setting or a daily life activity (Ramirez & Beilock, 2011). Some authors propose that MA affects performance by occupying working memory with intrusive worries about the task, not allowing students to carry it out (Ashcraft, 2002). It is thought to be one of the main reasons for students that do not like math and want to avoid it (Ashcraft, 2002).

MA can be divided into two dimensions: an affective and a cognitive one (Choi & Clark, 2006; Ho et al., 2000; Wigfield & Eccles, 1989). The affective dimension refers to a feeling of nervousness, tension and fear towards mathematics. The cognitive one refers to the negative expectancy of performing correctly in mathematics. The affective dimension (negatively) correlates to mathematical performance and self-perception of ability, while the cognitive one correlates to the value students give to maths and their actual effort (Cates & Rhymer, 2003; Ho et al., 2000). MA has thus detrimental effects in learners' academic outcomes: They purposely choose fewer math courses, obtain lower grades in the mathematics-related subjects and choose college majors that are less related to mathematics than their low math anxiety peers (Ashcraft, Kirk, & Hopko, 2000).

Nevertheless, it is not clear yet the directionality of the relationship between performance in mathematics and MA. Some authors have discussed this topic, expressed as the *Deficit Theory* vs *Debilitating Anxiety Model. Deficit Theory* one presents mathematical incompetence as the cause of MA; while in *Debilitating Anxiety Model*, it is MA what causes lower mathematical performance. Carey and colleagues (Carey, Hill, Devine, & Szücs, 2016) present a third possibility, known as the *Reciprocal Theory,* where this relationship is bidirectional and can influence each other.

b) Perceived competence

Competence beliefs are ability self-concepts referring to one's cognitive representations of how good one is at a given activity (Wigfield & Eccles, 2000). Several researchers (Bandura, 1997; Pajares, 1996a) claimed that perceptions of self-competence are a core determinant of a person's ability to spend more time when the role becomes challenging, even when an error arises. These judgements of one's competence have been shown to be even more robust predictors of behaviour than prior knowledge (Multon, Brown, & Lent, 1991; Pajares, 1996b; Schunk, 1991). They have been extensively studied in young children, as it is in the early school years when they begin to form domain-specific competence beliefs for newly introduced subjects as mathematics (Guay, Ratelle, Roy, & Litalien, 2010; Wigfield et al., 1997). Moreover, it has been shown that those competence beliefs are not general perceptions, as they differ among subjects (Harter, 1982; Marsh & Martin, 2011; Valentine, DuBois, & Cooper, 2004).

One of the subjects most affected by one's competence beliefs is mathematics. In contrast to mathematical anxiety, which can be considered an emotional measure, mathematical self-concept is a motivational one. Nevertheless, both measures are closely linked (Pietsch, Walker, & Chapman, 2003). Despite not being as related to test anxiety as mathematical anxiety, perceived competence in mathematics tends to be expressed different between genders, with females reporting lower self-concept than males (Kenney-Benson, Pomerantz, Ryan, & Patrick, 2006; Marsh & Yeung, 1998).

c) Goal orientation

Any problem can be divided into three parts: the givens, the goals, and the obstacles (Anderson, Boyle, & Reiser, 1985). The givens are the different elements that form a problem, and the goals are the desired outcome. The obstacles are the characteristics of the problem (or the student) that make it difficult for the learner to transform the givens to the desired goal, or to realise that this transformation has happened.

There are different ways in which learners monitor and reflect on their work and choose their goals: either being performance- or mastery-oriented. Performance-oriented students focus more on watching their peers, leading to missing cues about their own learning. On the contrary, mastery-oriented learners consider their own effort and process and act consequentially, either changing strategies or applying more effort.

Further development of this theory led researchers to add one dimension: approach vs avoidance. Like this, a 2x2 framework arises (Harackiewicz, Barron, Pintrich, Elliot, & Thrash, 2002; Pintrich et al., 2000). This framework divides learners among four profiles: Performance-Approach (PAp), Performance-Avoidant (PAv), Mastery-Approach (MAp), and Mastery-Avoidant (MAv). The focus of the first group, PAp, is to do better than their classmates, leading them to pursue tasks that will ensure success over their peers. Contrarily, a child with a PAv approach will focus on not doing it worse than their peers, avoiding tasks with a higher chance of failure (or that they think that they are not capable of doing). This kind of goals is associated with task withdrawal and self-handicapping behaviour, with the idea of avoiding failure (Harackiewicz et al., 2002) MAp have the goal to develop competence and skills. Finally, MAv represents students' fear of failure due to lack of understanding, which can lead to disorganised studying (Elliot & McGregor, 2001).

In general, all four goals but PAv can lead to positive outcomes (Harackiewicz et al., 2002).

Mastery-oriented approaches promote greater engagement and understanding, and appropriate help request (Ames, 1992; Pintrich, 2000, 2003). Contrarily, a performance-oriented approach can lead the student to depend more on external feedback and focus more on comparison to other than understanding and improvement in the task (Church, Elliot, & Gable, 2001). This theory follows a similar pattern than Dweck's mindset theory on views of intelligence (Dweck & Henderson, 1989) According to Dweck, learners having an incremental view/growth mindset view intelligence as something malleable that can be developed over time. Consequentially, students are more open to feedback (both positive and negative), and their learning goals focus on improving in the task, leading to greater learning. On the contrary, for students that have an entity view / fixed mindset, intelligence is fixed and cannot be improved, leading them to focus on validating this intelligence (and paying less attention to errors), what can lead to overconfidence.

d) Intolerance of uncertainty

Any decision-making process has an implicit uncertainty over the best rule to be followed to obtain the desired outcome. Resolving this uncertainty is thus one of the most challenging aspects of the decision-making process. The presence of uncertainty indicates that an individual has only partial knowledge about a given piece of information. Some authors defended the view of uncertainty as a motor for cognitive development, engendered by the awareness of internal dissonances, as it can help in maintaining and stimulating curiosity (Pallasmaa, 2010). It has been shown to guide social learning in infants (Harris & Lane, 2014) and young children (Sobel & Kushnir, 2013).

Intolerance of uncertainty (IU) represents "the tendency to react negatively on an emotional, cognitive, and behavioural level to uncertain situations and events" (Dugas & Koerner, 2005). Individuals with high intolerance of uncertainty find ambiguity distressing and have difficulties in dealing with uncertain situations, showing a tendency to react negatively to them (Dugas, Buhr, & Ladouceur, 2004). While experiencing uncertainty, individuals with high IU may experience distorted contingency beliefs, perceiving the expectancy of threat disproportionately to the expectancy of safety. Consequentially, they may perceive neutral or even positive cues as dangerous (Dugas et al., 2004).

e) Hypotheses

The present study analyses the relationship between learners' perception of their skills and performance and their actual accuracy while solving a mathematical task. It also addresses the question of whether learners, when provided with different difficulty levels and types of task, engage in different patterns of exploration, and if whether these patterns depend on their characteristics.

- About confidence:
    - Students will be inaccurate in their perceived accuracy (confidence), showing a tendency for overconfidence
    - Learners' confidence will relate to their goal-orientation (positively correlated to mastery-oriented and negatively to performance-oriented)
    - No differences between genders in performance but in confidence

- About gender:
    - Female students will show higher mathematical anxiety and lower perceived competence
    - Female students will tend to avoid negative feedback to a higher extent, leading to less exploration

- About exploration:
    - Learners presented with more items that challenge their rule use will show more informative exploratory

behaviour in the Free exploration part (more testing of previous errors)

- This group will also show more performance improvement in the post-test (compared to the pre-)

5.3 Materials and Methods

a) Participants

161 children (mean age: 8.7±0.5 years old, 75 females) participated in this study. The sample comes from three schools in Barcelona, Spain. Neither participants, their parents or school got any reward for participating in the study. The participating schools were contacted by email or phone for recruitment. Parents/caregivers were previously provided with an information sheet and signed a consent form to allow their children to participate in the study. Participants also received an information sheet and a verbal explanation before the experiment.

b) Conditions

This study contains a total of six conditions. The conditions come from merging two dimensions tested: active/passive learning and three levels of difficulty.

Active/passive dimension:

- **Active:** In this condition, participants answer exercises like the ones in the pre- and post-test (blocks 1 and 5, respectively) and get feedback about them.
- **Passive:** In this condition, participants see the setup of the exercise and then, the outcome of the balance.

The exercises provided are the same for both conditions. These items are chosen given the three levels of the difficulty dimension:

- **Same:** In this condition, participants receive items from the same rule as the one they classified for in the pre-test.
- **Middle:** In this condition, half of the items participants receive come from the rule they classified for in the pre-test and the other half is one rule higher.
- **Higher:** In this condition, participants receive all the items from one rule higher than the rule they classified for in the pre-test.

The difficulty dimension has two exceptions: Not being able to classify the child or them classifying as rule IV (as it is the highest one). If a participant could not be classified, they would receive four items of each type, as in the pre- and post-tests. If they were classified as rule IV, they would receive all the items from rule IV and be later considered as "Same" condition for the analysis. Nevertheless, this last option was never carried out, as no child classified as rule IV in the pre-test.

Considering these two dimensions, six conditions appear. In the first one, Passive Same (PS), participants observed items of the same rule to the one they classified for in the pre-test. In the second, Passive Middle (PM), half of the items participants observed came from the same rule they scored in the pre-test and the other half were one rule higher. In the third condition, Passive High (PH), participants observed items of one rue higher than the one they classified for in the pre-test. In the fourth, Active Same (AS), participants answered exercises of the same rule to the one they classified for in the pre-test. In the fifth condition, Active Middle (AM), half of the exercises came from the same rule they scored in the pre-test and the other half were one rule higher. Finally, in the sixth condition, Active High (AH), participants answered exercises of one rue higher than the one they classified for in the pre-test. ***Table 6*** presents a summary of the conditions.

Table 6. Experimental conditions. Merging the difficulty (same as the rule extracted in the pre-test, 50% same and 50% more difficult than the rule extracted in the pre-test, all more difficult than the rule extracted in the pre-test) and the instructional (passive/active) dimensions, six conditions appear.

	Rule used in pre-test	**½ Rule used in pre-test, ½ +1**	**Rule used in pre-test +1**
Observation of setup + outcome	Passive Same (PS)	Passive Middle (PM)	Passive High (PH)
Provide answer (L/B/R)	Active Same (AS)	Active Middle (AM)	Active High (AH)

c) Experimental Protocol

The experiment comprised two sessions of an hour each of them. There were two weeks between sessions. ***Figure 15*** represents the timeline of each session. Both sessions began with an introduction lasting between 15 and 20 minutes, in which the participants could also ask questions to make sure the procedure was clear before beginning. The purpose of Session 1 was to obtain, on the one hand, their baseline on the mathematical knowledge required in the task (basic arithmetic). Participants are evenly distributed in the second session’s conditions using the results of the first one.

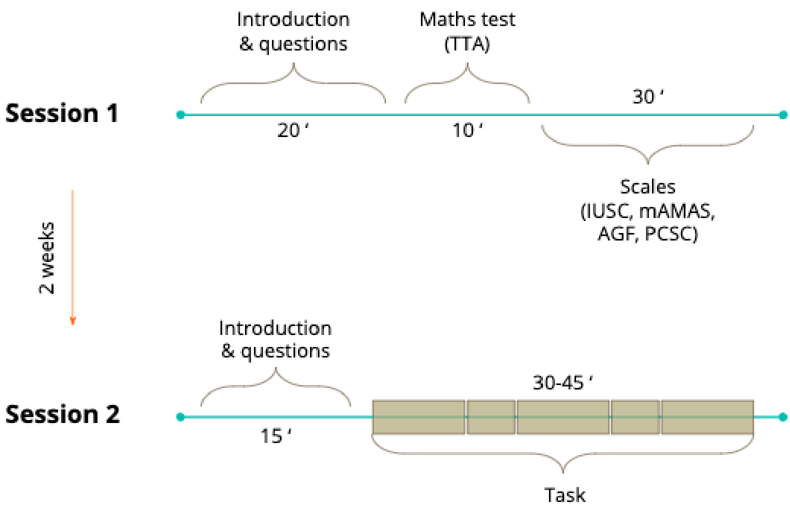


Figure 15. Experimental protocol. The upper part depicts the first session, where the mathematical test and the psychological scales were provided. The bottom part depicts the second session, where the Balance Scale task was carried out. There were two weeks between sessions.

In each school, children were divided into groups of around 20 participants to ensure space between participants and a more controlled environment. Each participant filled in the questionnaires and performed the task individually in a computer/laptop. The questionnaires were completed online using the SoSci[4] survey service, an online survey service following the GDPR legislation on data protection. Children carried out the task in a Unity-based application programmed for the study, installed locally in each computer.

[4] https://www.soscisurvey.de/

	Pretest: Determine rule	Free exploration	Training (Manipulation)	Free exploration	Posttest: Determine rule
Answer	L/B/R		L/B/R | Observation		L/B/R
Outcome	No	Yes	Yes	Yes	No
Tailored	No	No	Yes	No	No
Items	20	10	20	10	20
	Jansen et al., 2002				

Figure 16. Blocks inside of the computerized Balance Scale task

The task of session 2 uses a computerised version of the Balance Scale task, which automatically detects participants' strategies to provide tailored exercises. This task contains five parts, depicted in ***Figure 16***. First, participants answer the expected outcome (left, right or equilibrium) of the balance and report their confidence in their answer (without receiving feedback later). Then, they can freely explore the balance to test their hypotheses. In the third part, students see the outcome of the balance's setup, either by directly seeing the answer or by being provided feedback on theirs. Following that, they can explore the balance again. Finally, they are asked again to report the expected outcome, without receiving feedback about it.

d) Measures and Tools

Measures can be divided into distal and proximal variables. Distal variables are those variables related to participants' characteristics, similar to their traits. For that reason, these measures remain more stable over sessions. On the contrary, proximal variables relate to participants' behaviour in the task and are obtained through their interaction with the task.

The distal measures of this task serve to measure the metacognitive and motivational variables presented before. To do so, we used already established psychological scales (always choosing the version for children, if possible). The scales, representing participants' motivational and affective factors, are the Intolerance of Uncertainty Scale for Children (Comer et al., 2009), the Modified Abbreviated Math Anxiety Scale (Carey, Hill, Devine, & Szucs, 2017), the Achievement Goals Questionnaire (Elliot & McGregor, 2001), and the Perceived Competence Scale for Children (Harter, 1982). The mathematical test is the Tempotoets Automatiseren (De Vos, 2010), a timed arithmetic test. All the scales and the test can be found in the Appendix.

The proximal measures obtained during the task were performance, obtained as the score during the task; confidence, obtained through asking participants for self-reports after each answer; and response time. From them, we obtain a measure of exploration, explained in the Results subsection. To obtain them, we developed a computerized version of the Balance Scale task in Unity3D[5], previously tested in a pilot experiment with 24 children.

5.4 Results

a) Results on confidence

Students are inaccurate in their perceived accuracy (confidence), showing a tendency for overconfidence

To assess participants' initial accuracy (or lack of) in self-evaluating their skills, we analysed the relationship between performance and confidence reports in the pre-test (Block 1, when they do not receive feedback). As we can see in ***Figure 17*** (Left), these variables do not seem to follow a linear relation (Pearson R=0.11, p = 0.18). Confidence reports are similar across all the performance range. Less

[5] https://unity.com

knowledgeable students seem not to be aware of their incompetence, as shown in (Comer et al., 2009). Contrarily, more knowledgeable participants seem to show better-calibrated metacognition, despite some of them reporting less confidence than their actual performance.

We then assessed their final accuracy in self-evaluating their skills by computing the same correlation in the post-test (Block 5, where they did not receive feedback either). In the right part of ***Figure 17***, we can see how the general metacognition improved, as there is a significant correlation between performance and confidence (Pearson R=0.15, p = 0.06).

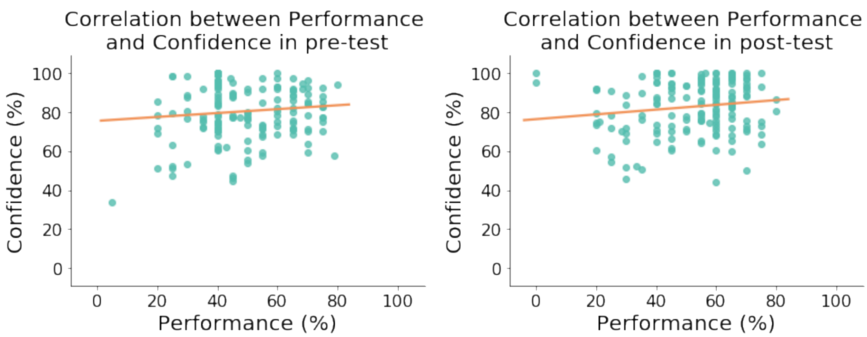


Figure 17. Correlation between performance and confidence in pre- and post-tests. Left: No correlation between Performance and Confidence in the pre-test (Pearson R=0.11, p = 0.18). Right: Almost significant correlation between Performance and Confidence in the post-test (Pearson R=0.15, p = 0.06).

Differences between genders in performance and confidence

When we plot the relationship between performance and confidence considering the participant's gender, we can see that females follow a positive trend (that is, their performance and confidence are more aligned) than males, whose relationship between performance and confidence shows a negative direction. ***Figure 18*** depicts these differences. Left: Negative trend in the correlation between Performance and Confidence in the pre-test in males (Pearson's r=-0.08, p = 0.44) and significant positive trend in females (Pearson's r=0.23, p = 0.04). Right: Positive trend in the correlation between performance and confidence in the post-test in males (Pearson's r=0.12, p = 0.28) and females (Pearson's r=0.17, p = 0.15).

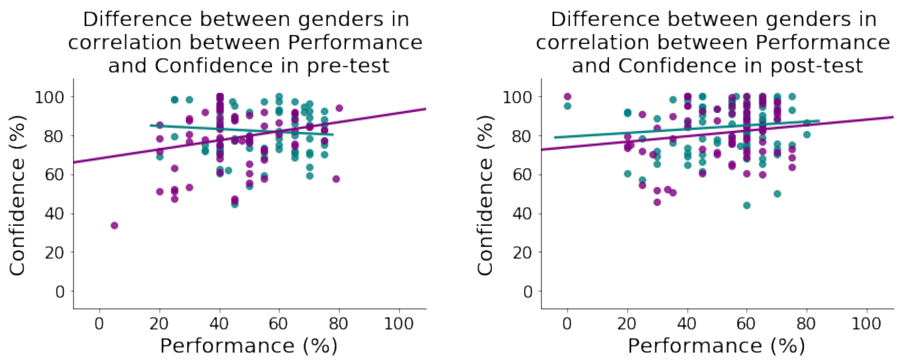


Figure 18. Differences between genders in the correlation between performance and confidence in pre- and post-tests. ***Left:*** Negative trend in the correlation between Performance and Confidence in the pre-test in males (Pearson's r=-0.08, p = 0.44) and significant positive trend in females (Pearson's r=0.23, p = 0.04). ***Right:*** Positive trend in the correlation between performance and confidence in the post-test in males (Pearson's r=0.12, p = 0.28) and females (Pearson's r=0.17, p = 0.15).

Learners' confidence does not directly relate to their goal-orientation (positively correlated to mastery-oriented and negatively to performance-oriented)

To analyse the relationship between participant's confidence in their answers and their goal-orientation, we analysed the correlation between confidence reports in Block 1 (when they have not received feedback yet) and their scores in each dimension of the AGF scale. ***Figure 19*** represents these correlations. Top left: Negative trend in the correlation between bias in the pre-test and Performance-Approach goal (Pearson's R = -0.08, p = 0.30). Top right: Negative trend in the correlation between bias in the pre-test and Performance-Avoidance goal ((Pearson's R = -0.11, p = 0.15). Bottom left: Negative trend in the correlation between bias in the pre-test and Mastery-Approach goal (Pearson's R = -0.14, p = 0.08). Bottom right: Negative trend in the correlation between bias in the pre-test and Mastery-Avoidance goal (Pearson's R = -0.08, p = 0.33).

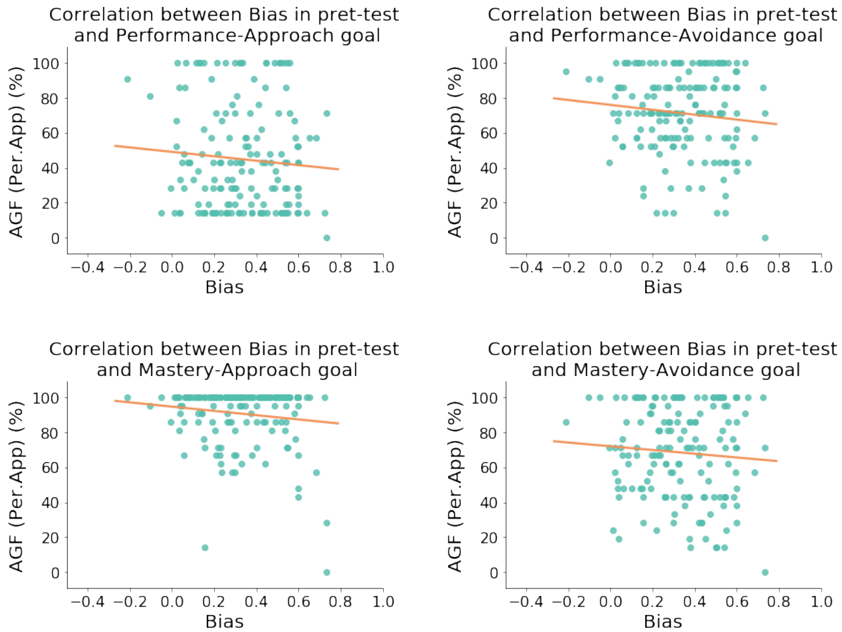


Figure 19. Correlations between bias in pre-test and the four dimensions of the AGF scale. Top left: Negative trend in the correlation between bias in the pre-test and Performance-Approach goal (Pearson's R = -0.08, p = 0.30). Top right: Negative trend in the correlation between bias in the pre-test and Performance-Avoidance goal ((Pearson's R = -0.11, p = 0.15). Bottom left: Negative trend in the correlation between bias in the pre-test and Mastery-Approach goal (Pearson's R = -0.14, p = 0.08). Bottom right: Negative trend in the correlation between bias in the pre-test and Mastery-Avoidance goal (Pearson's R = -0.08, p = 0.33).

b) Results on gender differences

Female students will show higher mathematical anxiety and lower perceived competence

As hypothesised, female students (58.7% ± 16.66) showed higher mathematical anxiety than male students (59029% ± 17.84), as shown by their scores in the mAMAS (independent samples t-test: -3.04, p = 0.003). They also showed lower perceived competence in mathematics (Males: median= 70.8, MAD=16.7; Females:

median=62.5, MAD=12.5; Mann-Whitney U: U = 1465.5, p < 0.001) and in the Cognitive (Males: median= 70.8, MAD=12.5; Females: median=62.5, MAD=8.3; Mann-Whitney U: U = 1682.0, p = 0.001) and Self (Males: median= 79.2, MAD=12.5; Females: median=62.5, MAD=16.7; Mann-Whitney U: U = 2326.0, p = 0.001) dimensions of the PCSC, as shown in ***Figure 20***.

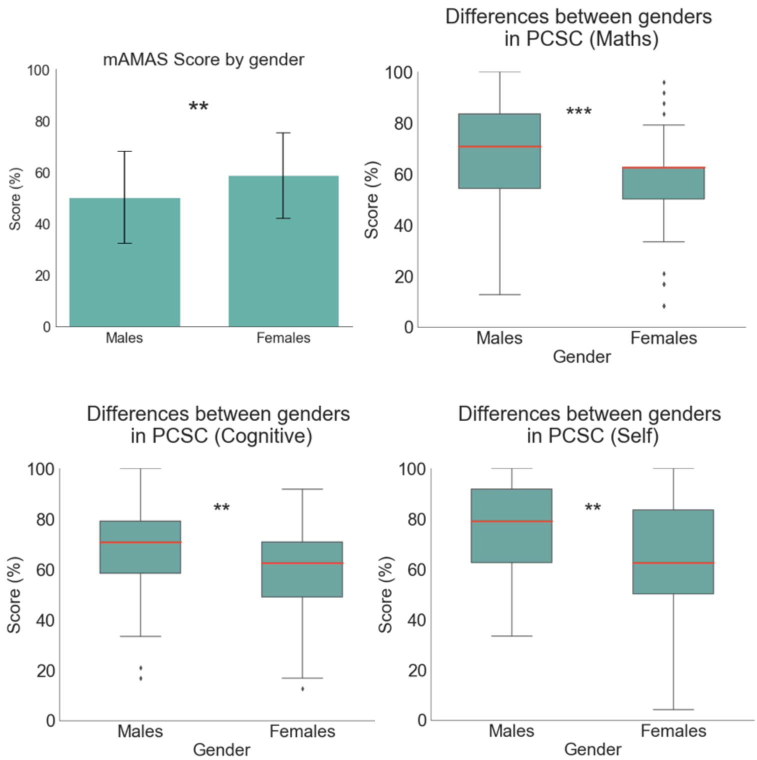


Figure 20. Differences between genders in mathematical anxiety and perceived competence in cognitive, self, and mathematics. ***Top left:*** Significant differences (independent samples t-test: -3.04, p = 0.003) in mAMAS score between males (50.29% ± 17.84) and females (58.7% ± 16.66). ***Top right:*** Significant differences (Mann-Whitney U: U = 1465.5, p < 0.001) between males (median= 70.8, MAD=16.7) and females (median=62.5, MAD=12.5) in the mathematical dimension of the PCSC. ***Bottom left:*** Significant differences (Mann-Whitney U: U = 1682.0, p = 0.001) between males (median= 70.8, MAD=12.5) and females (median=62.5, MAD=8.3) in the cognitive dimension of the PCSC. ***Bottom right:*** Significant differences (Mann-Whitney U: U = 2326.0, p = 0.001) between males (median= 79.2, MAD=12.5) and females (median=62.5, MAD=16.7) in the self dimension of the PCSC.

Female students will tend to avoid negative feedback to a higher extent, leading to less exploration

To assess if there were differences between genders in sensitivity to negative feedback, we first measured the differences between male and female participants in the avoidance-related profiles of the AGF. Like this, we can see in (***Figure 21***, Left) that there are no differences between males (median= 71.4, MAD=14.3) and females (median=71.4, MAD=14.3) in their score for the Performance-Avoidance goal orientation (Mann-Whitney U: U=2930.0, p = 0.26). Nevertheless, we can see in (***Figure 21***, Right) that there are differences between males (median=66.7, MAD=23.8) and females (median=81.0, MAD=19.0) in the Mastery-Avoidance goal orientation (Mann-Whitney U: U=2471.0, p=0.01).

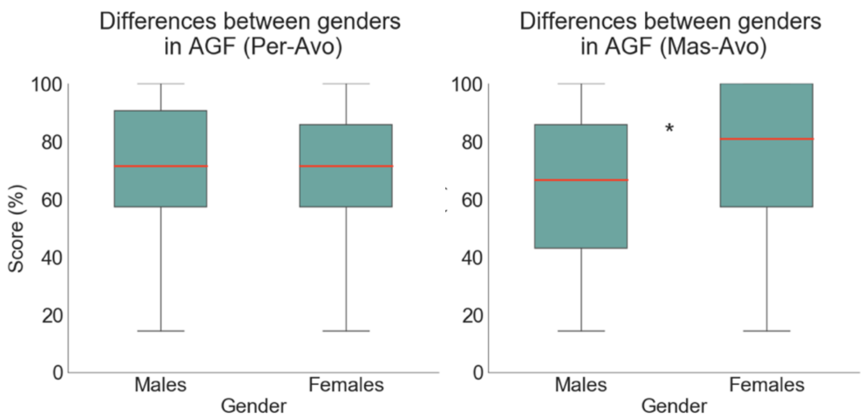


Figure 21. Differences between genders in the avoidant profiles of the AGF. ***Left:*** No difference between males (median= 71.4, MAD=14.3) and females (median=71.4, MAD=14.3) in their score for the Performance-Avoidance goal orientation (Mann-Whitney U: U=2930.0, p = 0.26). ***Right:*** Significant differences between males (median=66.7, MAD=23.8) and females (median=81.0, MAD=19.0) in the Mastery-Avoidance goal orientation (Mann-Whitney U: U=2471.0, p=0.01)

c) Results on exploration

Learners presented with more items that challenge their rule use will show more informative exploratory behaviour in the Free exploration part (more testing of previous errors)

For this part, we only considered the results of the participants who were classified into one of the four rules during the pre-test (Block 1), which leaves us with half of the sample (N = 79). ***Figure 22*** shows the proportion between the three groups in choosing a lower (blue), the same (orange) or one rule more (green) than the one they used in the pre-test.

Learners presented with more items that challenge their rule use will also show more performance improvement in the post-test (compared to the pre-)

We wanted to assess the effect of providing challenging items in participants' improvement after the task. To do so, we split the data into the three levels of difficulty: the ones that received all the items from their same rule ("All same rule"), the ones that received half of the items from their same rule and the other half from a higher rule ("½ same ½ higher") and the ones that received all the items from a higher rule ("All higher rule"). ***Figure 22*** shows these differences.

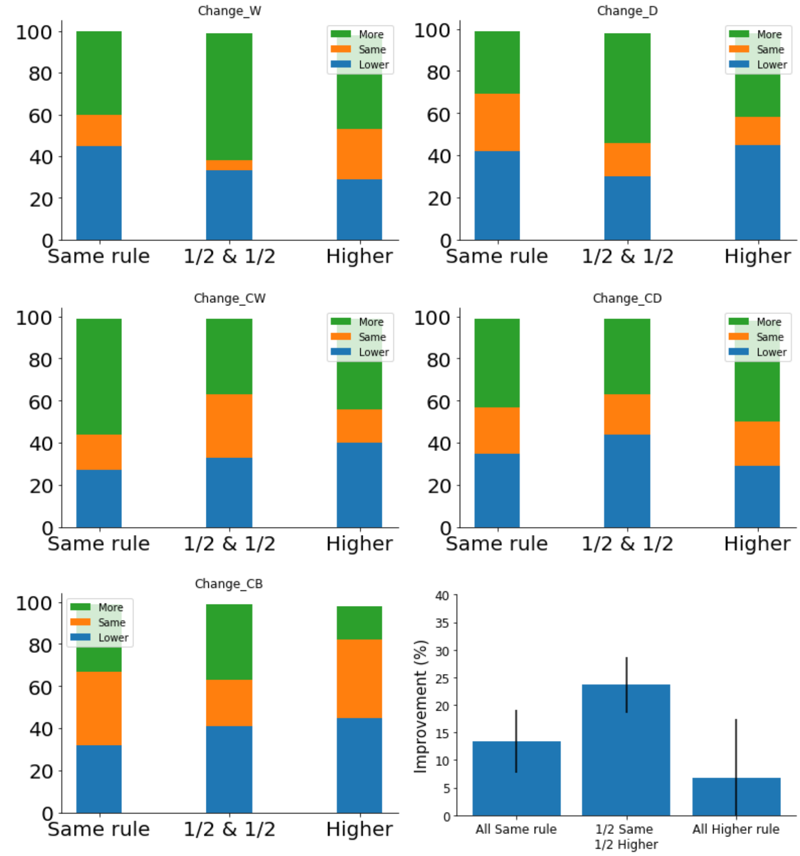


Figure 22. Proportion between the three groups in choosing a lower (blue), the same (orange) or one rule more (green) than the one they used in the pre-test. From top to bottom, the plots show weight, conflict-weight and conflict-balance items, respectively. ***Top and middle right:*** distance and conflict distance items. ***Bottom right:*** Each group's improvement after the task depending on the items they received.

We used a Latent Class Analysis (LCA) to analyse the patterns of children's responses. LCA is a type of finite mixture distribution model (Kruger & Dunning, 1999) It divides the sample into a limited number of classes, each of them characterised by different patterns of probability to answer each item type correctly. From this, one can extract the use of particular cognitive rules, which are not limited to the original ones, as LCA detects clusters of unexpected response patterns that can be interpreted as alternative strategies. We used Rstudio for this analysis and the poLCA (Geary, McLachlan, & Basford, 1989) library, among others.

An LCA distinguishes between manifest variables (the observed behavioural measures) and latent ones (the unobserved variables), both of them, categorical. In this study, the manifest variables are the balance scale items: Weight (W), distance (D), conflict-weight (CW), conflict-distance (CD), and conflict-balance (CB); while the latent ones are the rules extracted from the patterns.

We first analysed the initial rules used. To do so, we analysed the patterns of item responses in Block 1, when no feedback is given yet. As we have five item types, we tested five models, generating from one up to five classes. ***Table 7***shows the results of each model. Considering the fit measures, we chose the model with two classes (M2) as the best option to represent the data.

Table 7. Comparison of the five models generated by the LCA for the rules in the pre-test.

Model	Log-likelihood	df	BIC	aBIC	cAIC	Likelihood-ratio
M1	-274.31	26	572.26	556.46	577.26	89.55
M2	-248.74	20	549.47	514.71	560.47	38.40
M3	-235.73	14	498.09	498.09	568.82	12.38
M4	-232.12	8	500.28	500.28	595.97	5.17
M5	-230.35	2	506.13	506.13	626.79	1.62

In ***Figure 23***, we can see the two classes created by the model. The x-axis represents each class; the y-axis, the probability per item and the z-axis, each item type. There seem to be two different patterns: At the left, the mix of item types depicts the pattern of Rule I (high performance in W and CW); at the right, it seems to represent a similar pattern to the one of Rule III (high W, D, and D, decrease in CW).

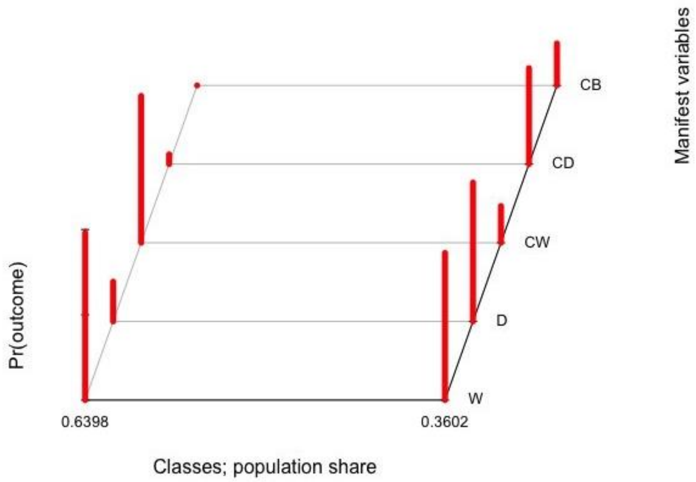


Figure 23. Result of the 2-class model representing the probability patterns in the pre-test.

Next, we assessed the final rule use (Block 5, also without feedback). ***Table 8*** shows the results of the LCA for that block. A three-class model is the model that better fits the data. It has the lowest, aBIC, BIC and cAIC without having the lowest likelihood-ratio.

Table 8. Comparison of the five models generated by the LCA for the rules in the post-test.

Model	Log-likelihood	df	BIC	aBIC	cAIC	Likelihood-ratio
M1	-312.09	26	647.81	632.01	652.81	106.74
M2	-286.43	20	624.86	590.09	635.86	55.42
M3	-268.42	14	617.20	563.47	634.20	19.40
M4	-262.68	8	634.09	561.39	657.09	7.92
M5	-260.22	2	657.53	565.87	686.53	3.00

Figure 24 depicts the probabilities of the three classes created by the M3 model. For the first latent class, the conditional probabilities of weight and conflict-weight items are high, depicting a pattern expected in Rule I. For the third class, the probabilities were high for weight, distance and conflict distance, a pattern similar to Rule III.

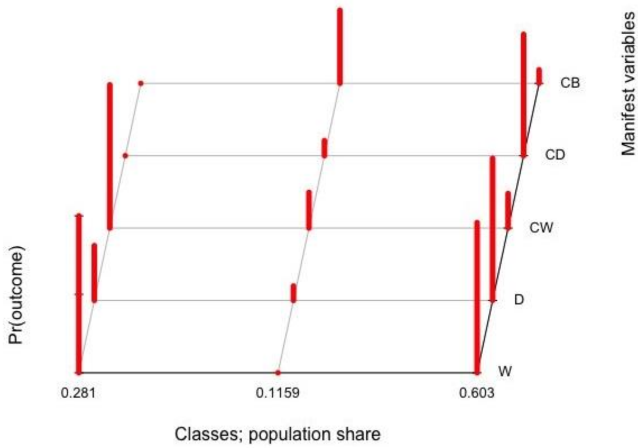


Figure 24. Result of the 3-class model representing the probability patterns in the post-test

When comparing the models from the pre- and the post-test, we can see two main differences: The change from Rule I (left) with a probability of 0.6 in the pre-test to Rule I with a probability of 0.3 in post-test; and the shift from Rule III (right) with a probability of 0.4 in the pre-test to probability of 0.6 in post-test. This change in pattern, going from a simpler to a more demanding rule, could represent participants' learning process during the experiment.

We then wanted to assess the exploration patterns during the two exploratory parts (blocks 2 and 4) the task. To do so, we first calculated the item types with minimum and maximum score during the pre-test (block 1) and the item types with the minimum and maximum examples created in the exploration phases (blocks 2 and 4). With this, we could see the variability of exploration per item type in both blocks. ***Figure 26*** depicts the exploration by correctness and item type in the first exploration phase (block 2), while ***Figure 25*** depicts the exploration by correctness and item type in the second exploration phase (block 4).

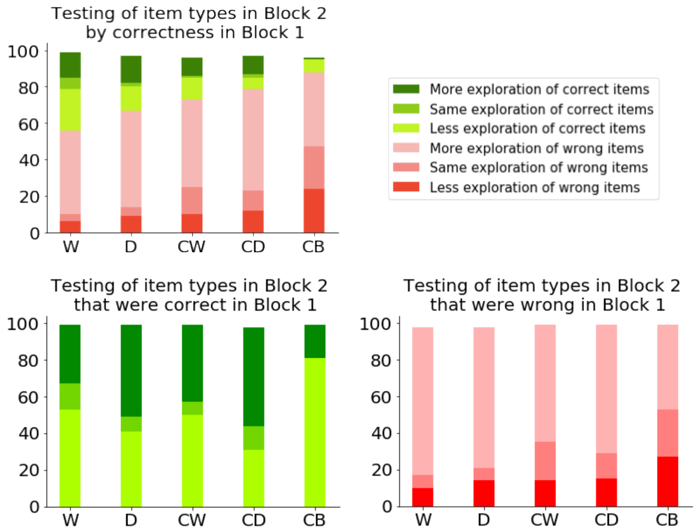


Figure 26. Degree of exploration by item type and correctness in the first exploration phase.

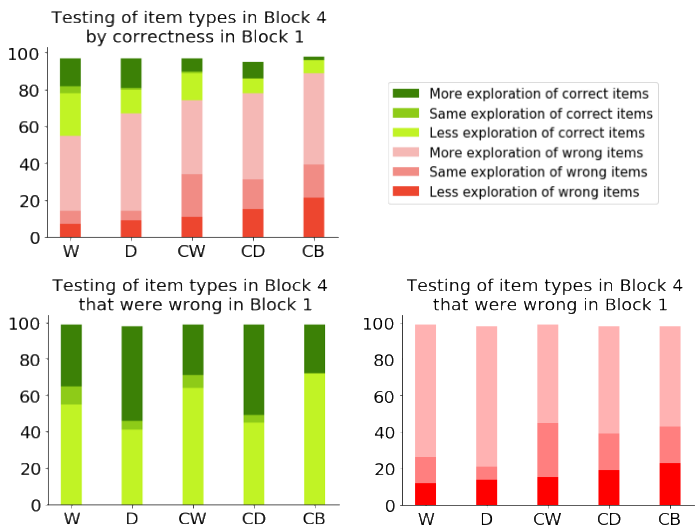


Figure 25. Degree of exploration by item type and corectness in the second exploration phase.

Nevertheless, this provides us with general information about the distribution. If we consider the minimum and maximum percentage per item type (either as correct answer in the pre-test block or as number of setups tested in the exploratory blocks), we can compare between blocks. We then can classify students considering the testing (or not) of previous item types. Like this, there could be:

- If a learner is not testing previously wrong items, they can be considered as being in an "***avoidant***" profile, as they are avoiding negative feedback
- If a learner is testing previously wrong items, they can be considered as being in an "***informative***" profile, as they are collecting information about previously wrong hypotheses
- If a learner is not testing previously correct items, they can be considered as being in an "***exploratory***" profile, as they are testing different item types than the ones they already master
- If a learner is testing previously correct items, they can be considered as being in a "***self-enhancement***" profile, as they are reinforcing their previous hypotheses

Combining them, we obtain the profiles presented in ***Table 9***.

Table 9. Profiles created from merging the four simple ones.

	Not testing previous wrongs	**Testing previous wrongs**	**Other**
Not testing previous corrects	Avoidant Exploratory	Informative Exploratory	Just exploratory, no specific strategy for wrong items
Testing previous corrects	Avoidant Self-enhancement	Informative Self-enhancement	Just Self-enhancement, no specific strategy for wrong items
Other	Just Avoidant, no specific strategy for correct items	Just Informative, no specific strategy for correct items	No specific strategy, neither for correct nor wrong items

We can see that in both exploratory phases (blocks 2 and 4), the main profile is “self only”, that is, the learners who tested previously correct items but did not took a specific strategy (either avoiding them or focusing on testing them) for the wrong items (***Figure 27***). Nevertheless, we can see an increase of “exp only” profile in the second exploration phase compared to the first one. In this profile, learners stopped testing previously correct items and tested other ones (despite not being the ones that they did the worst, as would happen with any “informative” profile). Despite this, they did not take a specific strategy with wrong items, neither testing them to inform their schemes nor avoiding them. Moreover, there is a decrease in “avo self” that is, the profile in which learners avoid testing previously wrong items and maximize testing previously correct items. We also have to comment, though, that the “inf exp” profile (when learners test previously wrong items and do not test previously correct ones) also increases.

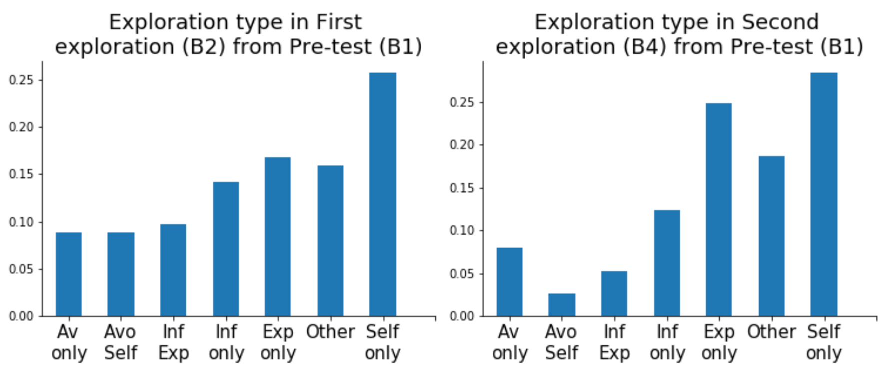


Figure 27. Classification of participants' exploration profiles during the exploration blocks.

6. STUDY 4: Virtual Reality for Historical and Cultural Learning

The previous study showed results on students' exploration patterns during their learning process and how they relate to their metacognitive abilities. These results, together with previous one in free vs guided navigation, motivated us to analyse individuals' exploratory and self-regulatory behaviour while using an educational application.

6.1 Introduction

Visitors of historical sites may face logistical and contextualization constraints during their experiences. We aim to improve visitors' experience by applying fundamental principles of learning and via using advanced interactive technologies such as Virtual- and Augmented Reality. Here we present an application to support visitors' experience to a memorial site and its outcomes on memory recollection.

Historical and Cultural Learning (HCL) is essential for the reflection on identity. Exposure to HCL is mainly viewed as visits to historical sites, with the aim of learning and advancing our understanding about the events occurred in those places. However, users may face logistical and contextualization commitments, required to provide the unique experience of visiting physical sites.

Visitors may find their experiences constrained in spatial, temporal and informational terms, due to the way current paradigms of HCL are expressed in memorial museums. This can be caused by the linearity of the predetermined routes followed during memorial sites visits. Visitors usually encounter highly predetermined paths, both in physical and virtual terms, as well as information that they consume in fixed amounts of time.

a) Recommender systems for HCL

Recommender Systems (RCs) have become common through commercial applications like for online shopping (Amazon), music (Spotify, LastFM), or tourist trips (Bobadilla, Ortega, Hernando, & Gutiérrez, 2013; Borràs, Moreno, & Valls, 2014). Based on a dynamic profile built during previous actions, a system's algorithms direct a user towards an item or place of her supposed interest. The digital application of such systems in the context of informal learning and tourism in HCL (both indoor and outdoor) is of great interest for a range of (self-) guided experiences, especially when information is vast while resources for human expert guides are limited. Similar to a human guide, advanced, context-aware recommender systems might be able to strike a dynamic balance between the interests of an exploring user and that of expert curators and educators.

However, the implementation of RCs in a heritage learning context remains a challenge. Unlike when recommending music or retail products, the recommendation of historical information by a teacher often seeks to gradually construct the understanding of a contextual framework, involving a complex of causal relationships, historical circumstances, traditions and viewpoints (Boxtel & Drie, 2004; Huijgen & Holthuis, 2015). The classification of historical content, as is necessary for an RC to pick from, is problematic and requires expert review of many pluriform sources and interpretations. Paradoxically, the preferred user experience (eg with an app used in a museum or memorial place) demands a sufficient level of knowledge to be reached within relatively few content iterations, while being sufficiently motivating and even fun. RCs in digital heritage apps offer to move away from predefined, 'passively' consumed linear narratives, but their development has to address the challenges of content selection and iteration mentioned above.

b) Virtual and Augmented Reality for HCL

A large part of the informational materials presented in museums is comprised by texts, images or videos. Usually, museums curators use instruction boards or audio/visual guides to merge and transmit all this information. Nevertheless, these methods face some limitations, for example, to show more vivid background information (Narumi et al., 2011). At the same time, the fast advances of technology regarding virtual and augmented realities allow to create visual representations of buildings and objects to provide a richer experience to the visitors (Stone & Ojika, 2000).

The reconstruction of historical buildings and objects carried out by Virtual Heritage follows several motivations (El-Hakim, Beraldin, Picard, & Godin, 2009):

- Reconstructing historic buildings that disappeared
- Creating virtual objects to interact with them without the risk of damaging them
- Documenting historic buildings and objects for reconstruction in case of a disaster
- Generating virtual museum exhibits
- Creating historical and cultural educational resources
- Visualizing scenes from different historical viewpoints

c) Visitors' experience in HCL

The Constructivist and Constructionist (more about it in the next chapter) theories defended that learning is created through the interaction with the world. In this line, VH allows visitors to augment the content by interacting with it in a totally different way to the traditional one (Pujol & Champion, 2012). The use of technology in museums facilitates for the design of learning activities based on exploration and discovery (Wakkary, Muise, Tanenbaum, Hatala, & Kornfeld, 2008), allowing visitors to make new connections between the exhibited content (Bell & Cuevas, 2013).

The process of HCL in museums is closely related to the idea of interpretation, that is, "an educational activity which aims to reveal meanings and relationships through the use of original objects by first-hand experience and by illustrative media, rather than simply to communicate factual information" (Tilden, 1977). The concept of interpretation fulfils three purposes: Learning (as acquiring new knowledge), satisfaction (as making the place more enjoyable) and as a provocation (as facilitating attitudinal or behavioural change) (Rahaman & Tan, 2011). The same author (Rahaman, 2018) added later a fourth purpose: Provide users with multiple perspectives of the past, to have a broader and alternative perspective. Visitors go through a process of meaning-making, that may not only be affected by the media they are facing but also their background, knowledge, interest and ideology, among others (Narumi et al., 2011).

d) Active Learning for HCL

The term *Active learning* has been frequently used in education to refer to the type of instruction focusing on learners' responsibility in their learning process. In it, students seek to construct conceptual meaning and understand the subject at a deeper level, so the teacher moves from teaching to memorise (Harland, 2012). Active learning encourages students' self-regulation by promoting opportunities to take responsibilities and make decisions (for example, of how and to what extent the learning will take place, or the type of help provided). To implement active learning in history education, teachers need to provide students with content very rich in breadth and scope.

Learning requires learners to actively construct meaning, in line with the constructivist theory presented in the Framework. Like this, learners construct meaning from merging the old information they have with the new one they acquire through interacting with the world. Building multiple mental models facilitates what is known as *meaningful learning* (Mintzes, Wandersee, & Novak, 2005) or *learning with understanding* (Simon, 2001). Consequentially, learning will be compromised if the old models contain misconceptions (that is, if they are flawed). It is thus crucial to engage users in questioning their hypothesis by encouraging more exploration to resolve possible contradictions. This revisiting and reformulation of their perspectives can promote better understanding. That is why it is important to assess learners' prior knowledge (or models) to build on them.

Speaking about meaning in cultural heritage education is speaking about *interpretation*. Interpretation refers to an educational activity aiming to present meaning through the use of objects and illustrative media, rather than using mere factual information (Tilden, 1977). One must consider, though, that the interpretation of heritage is highly related to visitors' subjectivity, cultural position and spatial literacy (McCullough, 2004). Consequentially, there can be cases of *heritage dissonance*, as not everybody perceives the value of specific content in the same way (Tunbridge & Ashworth, 1996). This is why it is important to capture various perspectives from various end-users, to ensure content and narrative multiplicity (Roussou, 2002) facilitating users' broader understanding of cultural heritage (Tammaro, 2016).

Allowing visitors to explore freely is highly recommended in real-world heritage interpretation (Brooks & Brooks, 1999; Copeland, 2006). Nevertheless, one needs to consider that users may not have enough time or interest to visit the whole site and exhibits or do all the activities (if any). It is for that reason that interactive maps and external guidance is also highly recommended to support visitors' experience.

Something to also consider is the sociocultural approach of learning. As already presented in the Framework, Vygotsky's Zone of Proximal Development (ZPD) focuses on the key role played by learning in social interaction. Moreover, Mercer (Mercer, 2004), in line with Vygotsky, defended that communication is shaped by cultural and historical aspects.

e) The FutureMemory learning paradigms

The FutureMemory (FM) project was created with the aim of advancing HCL via applying fundamental principles of learning and digital learning and via using advanced interactive technologies such as Virtual- and Augmented Reality applications. As a first instance the FM project goal was to improve and facilitate the access to information on the victims, perpetrators, resisters and bystanders of the Holocaust and the numerous sites of their lives and deaths.

The FM project conceptualization is grounded on Vygotsky's constructivist Zone of Proximal Development (Vygotsky, 1980), through supporting the learner to actively create knowledge by scaffolding the already existing one. It is also grounded on Inquiry-Based Learning (IBL), where an external tutor guides the learner by presenting challenges, which the learner responds to by actively gathering information.

Activity Theory (AT) is a similar perspective emerged in the field of Human-Computer Interaction (HCI). This approach, rather than considering single behaviours, focuses on the use of series of actions as meaningful units (Law & Sun, 2012; Rogers, 2008). The Active Learning in Digitally Enhanced Spaces (ALDES) approach comes from the union between the aforementioned psychology and neuroscience of learning and the new frameworks of HCI.

The scientific grounding of the ALDES paradigm comes from research on human spatial behaviour and memory. Active exploration has been shown to modulate spatial memory and performance in recollection (Chrastil & Warren, 2012). Moreover, previous research in interactive systems and complex information networks, shows its effects on deeper and better understanding of causal structures (Liao, 2005).

The term "active learning" is frequently used in education to refer to the type of instruction focusing on learners' responsibility in their learning process. It is also referred as "deep learning" (not to be confused with the ML approach) or "meaningful learning", in contrast to "shallow learning", where learners just memorise content. In similar lines to Constructivism, in this approach the teacher moves from being an information provider to a "guide in the side", supporting student's learning process. Students reflect upon ideas and how they are using them in the learning process. This requires a regular self-assessment of their understanding and skills. It involves the active construction of meaning by the learner.

e) The FutureMemory application

The Future Memory app caters the visitor with a virtual reconstructions of former concentration camps together with a multi-modal presentation of its history and the experiences of its prisoners. The app provides high levels of graphical details and also the ability to interact with 3D reconstrued virtual buildings and environments.

The application's database of Future Memory allows to organize and interact with historical context, and to associate them to real-world-coordinates. The database includes a searchable structure developed in SqlLite and its content (FM-db) contains digitized multimedia files.

Relevant components of the database organization are the metadata-sheets, which store multiple information fields, providing, like that, the connection between on-site navigation and database search. Other important items contained in the database are the ones delivering contextual connections (for example, introductions and guiding texts).

FM-edit, a dedicated web application, was developed to support experts' selection of content in the platform database, thus facilitating the collaboration between them and the developers. Moreover, it also provided a method for fast development and testing, as its connection to the database allowed for automatic update of the content's presentation.

The application is divided in two modes: Presentation and Master modes. The Presentation mode is application-specific, targeted towards the final user, with suitable style and information level. Through the Presentation mode, users can navigate both in space and content, using a visual representation incorporating the information (spatially associated via metadata). The Master mode includes more detailed information, from available historical content (such as photographs, blueprints, witness descriptions, and drawings) to maps and geographical data.

This platform has been deployed in two applications: an immersive indoor installation (FM-Room, ***Figure 28***) and hand-held version (FM-App) for outdoors. The FM-Room, (located at memorial centers, ***Figure 29***), has an introductory function, with a 180 degrees screen projection of the site reconstruction. The FM-App, is a hand-held version of the platform, and includes a geolocalized visualization of the camp's reconstruction and a content navigation tool.

Figure 28. Setup of the experience. A tabletop application allows users to navigate in space and content and a set of panoramic screens shows an interactive map of the BB camp.

Figure 29. The Augmented Reality effect. Through the hand-held device, users can see virtual reconstruction placed in real-world coordinates.

The aim of the FM-App is two-fold: on the one hand, it enhances the visitor experience at the site; and on the other, it serves for spatial navigation research and to study how it could be used to organize education and storytelling. The hardware used were iPads and the software used to implement it, Game Engine Unity3D[6].

The camp depicted in this implementation is the Bergen-Belsen memorial site. This camp, active from 1940 to 1945, was 250 acres long and was located 35 miles north of Hannover. It was close to the small town of Bergen, adjacent to the rural village of Belsen and to an extensive military complex.

The reconstruction of the former camp was carried out by merging the content (maps, descriptions, photographs and drawings) and the experts' research into master model files. That content was also used as a base for 3D modelling, done with Autodesk Maya[7] and Sketchup[8]. To ensure a reliable positioning of the buildings, maps were matched to old aerial photographs and present-day satellite imagery used for the tablet's global positioning system (GPS).

The placement and presentation of the content are possible via Points of Content (POCs), which associated latitude and longitude coordinates to relevant items; and Points of Interest (POIs), which group POCs to generate area-related groups and basic routes, depending, for example, in their thematic or topography.

POCs were initially hidden in the main platform design. Nevertheless, they were shown under controllable cases to allow for different kinds of interaction with the content. POIs can be presented in two ways: Free and Guided exploration. During Free exploration, the user can see all the POIs regardless of where s/he is in the space and access them when being close enough. During Guided exploration, the POIs are revealed stepwise, becoming visible later following the same proximity condition.

[6] https://unity.com
[7] https://www.autodesk.es/products/maya/overview
[8] https://www.sketchup.com

6.2 Materials and Methods

We carried out a user study to assess the users' experience (UX) with the application. UX can be defined as ''a consequence of a user's internal state (predispositions, expectations, needs, motivation, mood, etc.), the characteristics of the designed system (e.g. complexity, purpose, usability, functionality) and the context (or the environment) within which the interaction occurs (e.g. organisational/social setting, the meaningfulness of the activity, voluntariness of use)'' (Hassenzahl & Ullrich, 2007). Following that approach, we wanted to assess users' initial familiarity with the topic and interest, emotional state (before and after interacting with the installation), how they perceived the presentation, their experience, how it affected their interest, and how they navigated through the information.

a) The Hollandsche Schouwburg Memorial

This study was run at the Hollandsche Schouwburg Memorial (***Figure 30***), in Amsterdam (The Netherlands.) During World War II, the building was used by the German occupying forces as a collection site for Jews, who would be later deported to concentration and extermination camps. Now a place of commemoration, its main part is the memorial, but it also serves as a place to test different approaches to present cultural heritage information, specific about the Holocaust (for more information, see[9]).

[9] https://jck.nl/en/location/national-holocaust-memorial

Figure 30. The Hollandsche Schouwburg Memorial. Photo from the memorial's website.[10]

b) Participants

A total of 19 participants (7 females) interacted with the application. The age range is skewed, with a 40.05 ± 20.52 but a median of 29.0 and a MAD of 8.0. The main nationalities were The Netherlands and the UK; and the most common educational level was Bachelor, followed by a Master or a Doctorate degree. ***Figure 31*** depicts a visual summary of the demographical data.

[10] https://europeanmemories.net/memorial-heritage/hollandsche-schouwburg-national-holocaust-memorial/

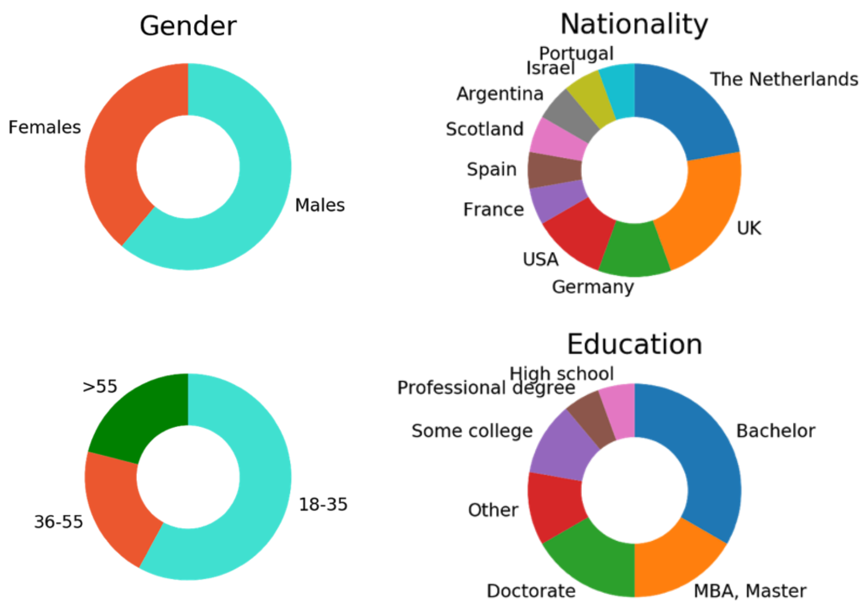


Figure 31. Visual summary of the demographical data

Neverthelesswe only have application data of 8 participants (3 females, age: 52.63 ± 18.86). For that reason, the results of the survey will contain all the participants (n=19) and the ones of the logs, only 8 visitors.

c) Experimental Protocol

Visitors first filled in a pre-questionnaire assessing their initial familiarity with the topic and interest, and their emotional state. Then, they interacted freely with the application as much time as they wanted. When they reported having finished, they filled in a post-questionnaire assessing their emotional state, how they perceived the presentation, their experience, if it changed their plans, how they navigated through the information and their demographical data. To see the full questionnaire, please check the Appendix. ***Figure 32*** presents the experimental protocol.

Figure 32. Experimental protocol. Participants filled in a questionnaire about their interest and previous knowledge on the topic and their emotional state. Then, they interacted with the application. Finally, they would be asked again about their emotional state, their experience with the application and demographics questions.

To assess participants' emotional state before and after interacting with the experience, we presented them with a 32-items list of emotional states (Nawijn, Isaac, Gridnevskiy, & van Liempt, 2018). They were instructed to first select the ones they were feeling and from them, the ones they were feeling the most in comparison to the rest of emotions.

The main tools of this study were the interactive experience (explained above) and the surveys. To create the surveys, we used soSci[11], a fully customizable online platform to create questionnaires. Participants filled in the questions in a tablet.

6.3 Results

a) Questionnaire results

Familiarity with the topic and interest

Given the location of the memorial (the Jewish neighbourhood), we wanted to assess if the users had visited the other museums in the area before (***Figure 33***, top). The museums surrounding the

[11] https://www.soscisurvey.de/

memorial are the Jewish museum, the National Holocaust Museum, the Children museum, and the Portuguese synagogue. Around 70% of users had visited the National Holocaust Museum before visiting the installation. Moreover, 40% of the users visited the installation before any other part of the memorial, from the other 60%, all had visited the memorial before using the installation.

We also assessed users perceived knowledge about the Holocaust and the Second World War (***Figure 33***, bottom). They mainly reporting knowing more about the Second World War than specifically about the Holocaust. When asked for their goal when interacting with the application, 60% of them reported being “only exploring” the application (instead of specifically wanting to learn about Bergen-Belsen or the Holocaust).

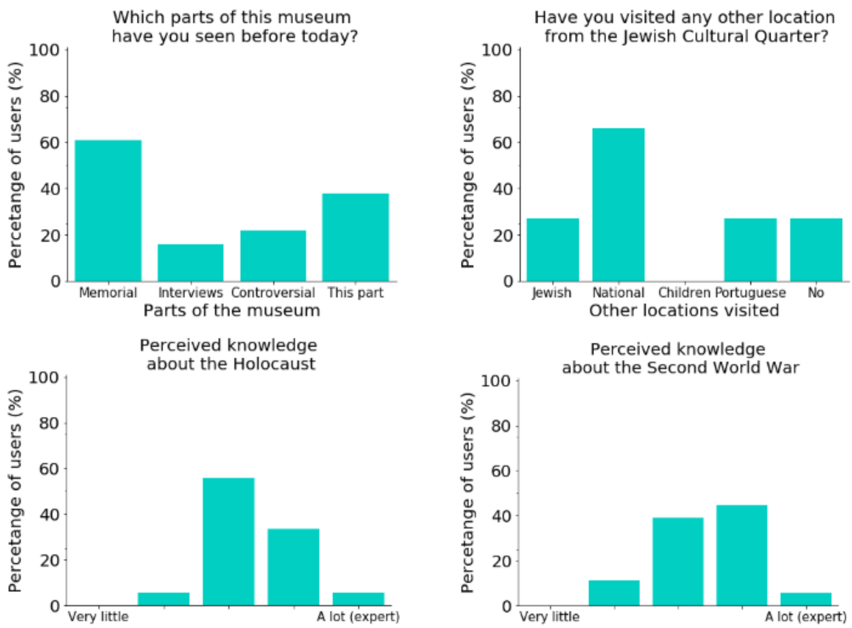


Figure 33. Results on familiarity with the topic and interest

Emotional state

Before interacting with the experience, participants mainly reported feelings of interest (84%), compassion (68%), sadness (63%), shock (47%) and fascination (36%). Afterwards, they showed a decrease in guilt (-16%), fear (-10%), hope (-10%), confidence (-10%) and love

(-5%); and an increase in sadness, embarrassment (+5%), affection (+10%), gratitude, (+6%) surprise (both negative (+6%) and positive (+5%)), annoyance (+10%) and awe (+5%). The results of all the items for the emotional state in the pre- and post-questionnaires are presented in ***Figure 34***.

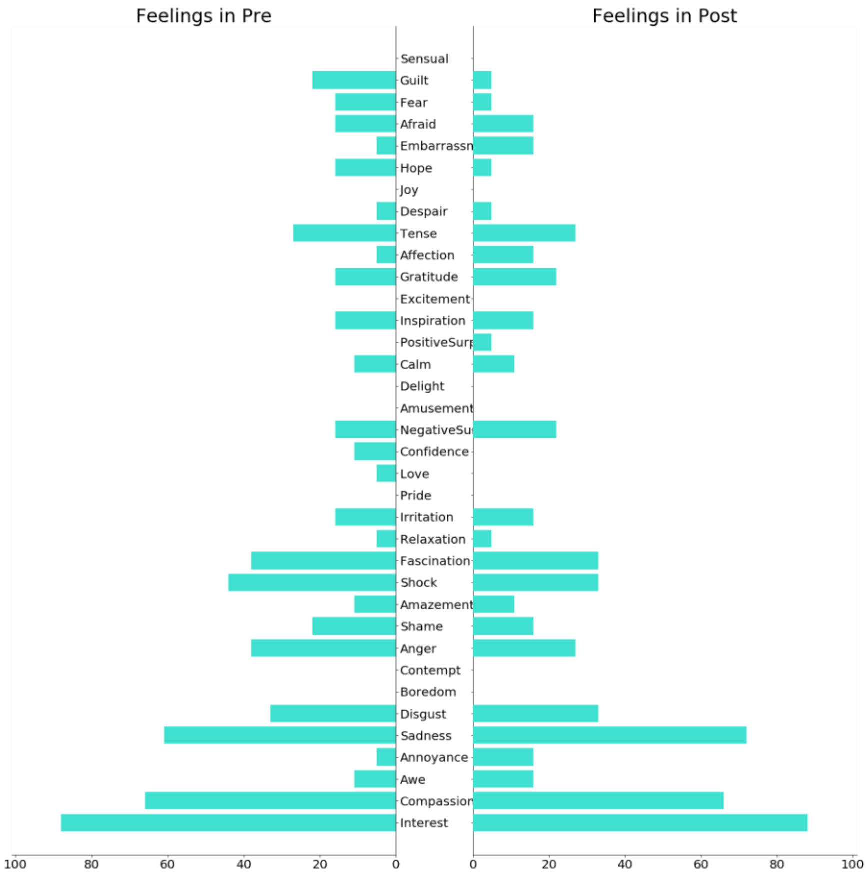


Figure 34. Comparison between self-reported feelings before and after interacting with the application. Interest, compassion and sadness are the most reported feelings, both in the pre- and the pos-surveys.

Presentation

When assessing users' perception of the experience's presentation, they reported that, despite not getting a clear impression of the daily life in Bergen-Belsen, they moderately understood how the camp looked like and could imagine it better ***Figure 35***. Moreover, using the application seemed to moderately influence users' plans about finding more information about Bergen-Belsen or planning to visit more memorials and museums on the subject.

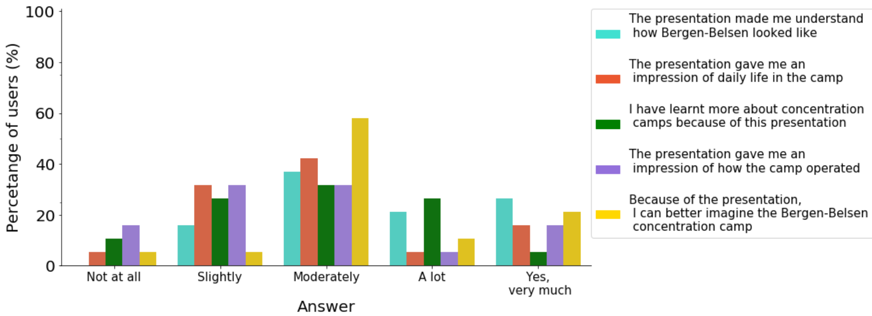


Figure 35. Visitors' perception of the application presentation.

User Experience

Users perceived the experience as interesting, impressive and slightly exciting (***Figure 36***). They reported the experience was innovative and challenging enough and that they would recommend it to others. It was perceived as moderately easy to learn to operate and to navigate in the space (***Figure 37***).

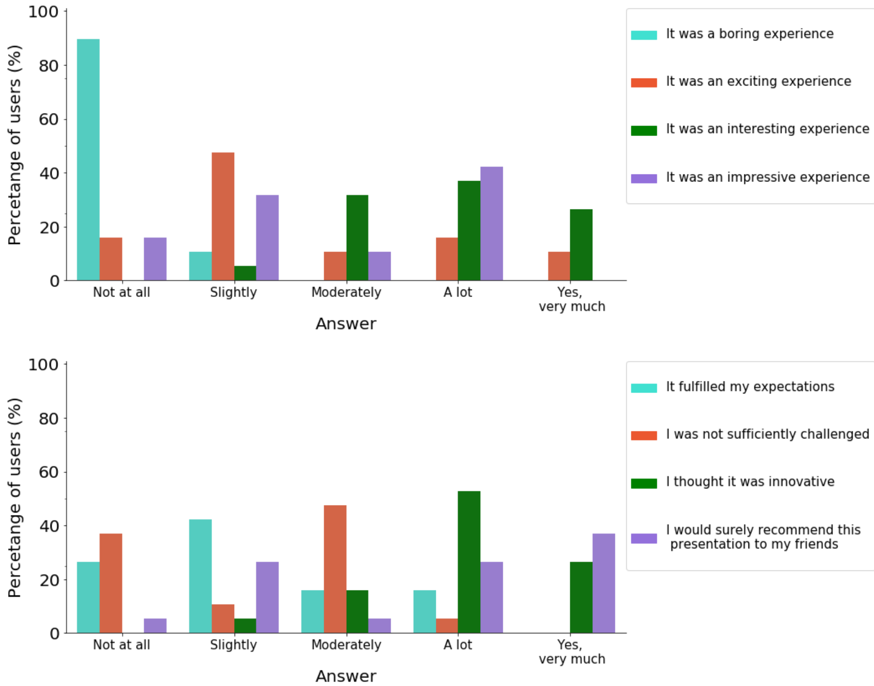


Figure 36. Visitors' perception of the experience

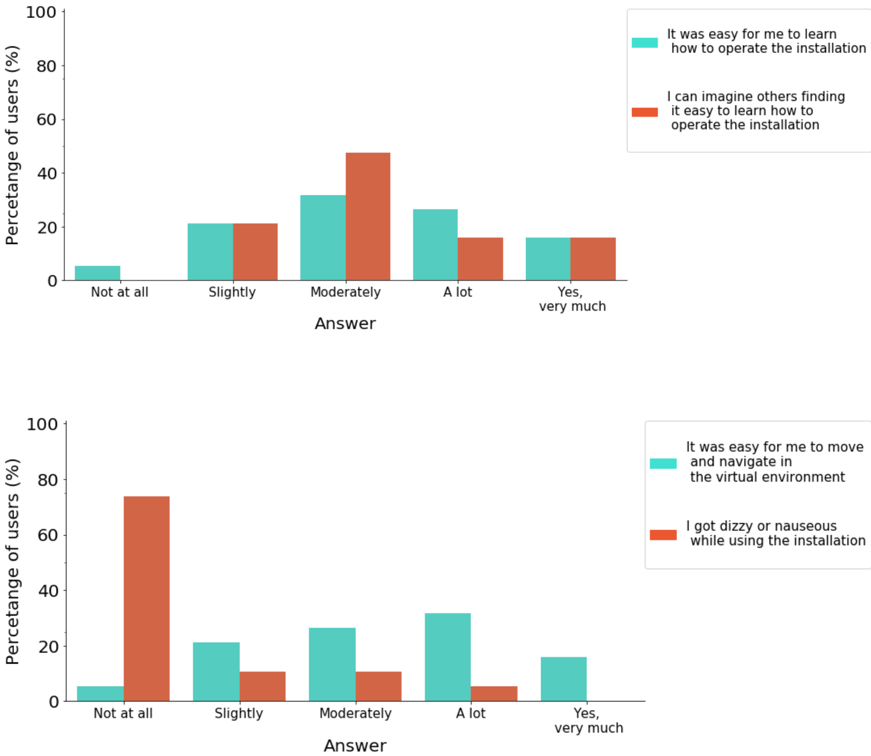


Figure 37. Visitors' perception of the easiness of use

Information

The content perceived as more interesting and surprising was the parts on children and daily life, with a high focus on personal stories ***Figure 38***, top). Moreover, users found it easy to understand the text' content and the way it was written (***Figure 38***, middle) and wanted to know more about Bergen-Belsen after interacting with the experience (***Figure 38***, bottom).

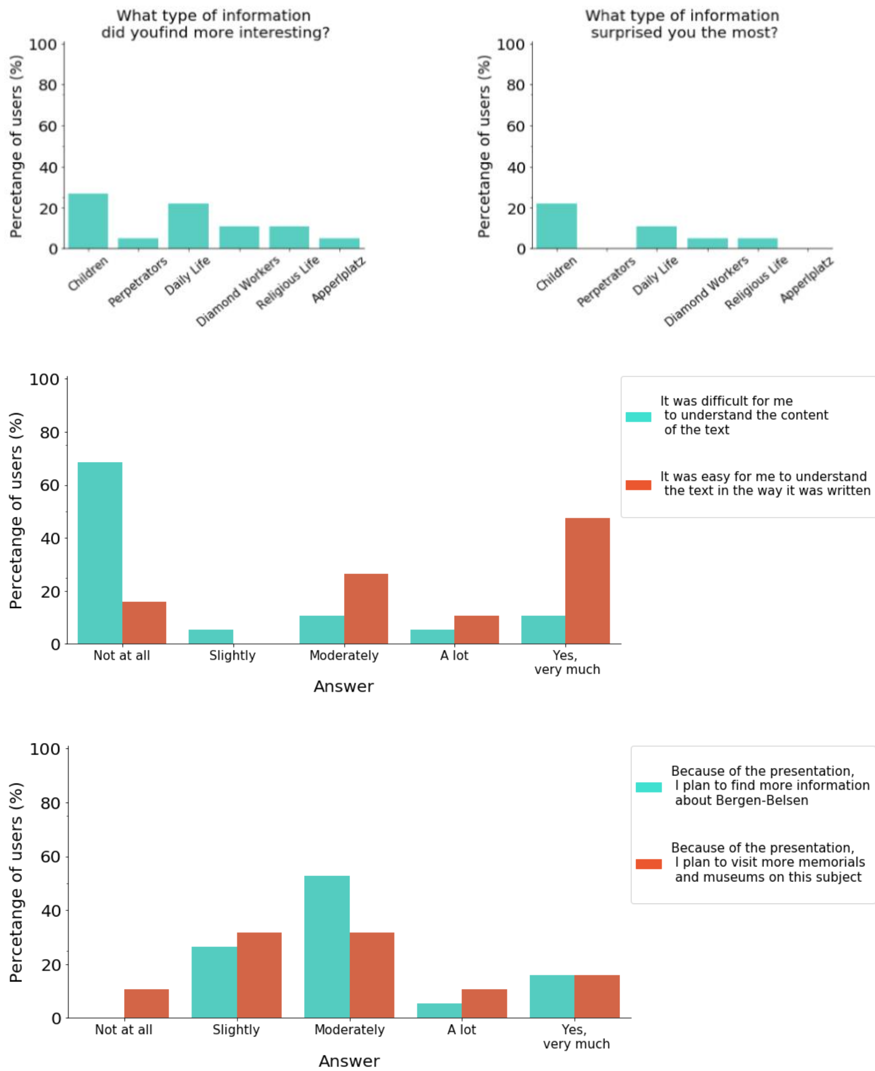


Figure 38. Visitors' perception of the provided information. ***Top:*** The content perceived as interesting and surprising. ***Middle:*** Understanding of the text. ***Bottom:*** Interest in knowing more about Bergen-Belsen.

Users' comments

As possible improvements, users commented they would like to see more data, have overall information of the site, more stories on how the people arrived there, more history of the camp and what was the fate of the people there. Other possible improvements would be to present it in more languages (at the moment, it is only in Dutch and English). Finally, regarding the use of technology, older users perceived the application as more directed to younger audiences and presented their concerns on it looking too much like a videogame, what, they said, could be perceived as "sterilizing". In general, despite missing some functionalities (like being able to walk into cabins), users perceived it as immersive, innovative and useful.

b) Content navigation results

The information coming from the computer logs allowed us to assess the most visit content. In line with the results reporting the "Children" and "Daily Life" exhibits as more interesting (***Figure 38***, top left), users mostly visited the "Children", "Daily Life" content. ***Figure 39*** represents the percentage of users visiting the exhibits.

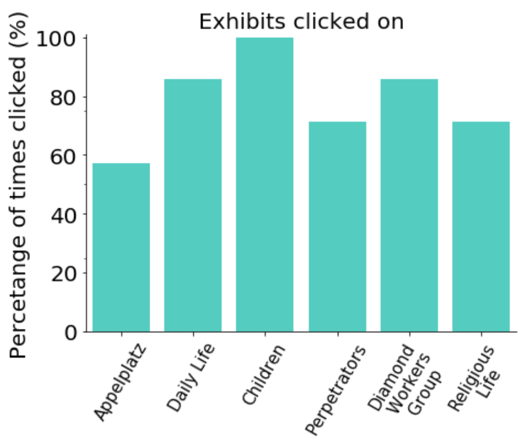


Figure 39. Visited content. All users visited the "Children" exhibits, followed by the "Daily Life" and "Diamond Workers Group" ones.

Figure 40 provides more detail of the content visited inside of each exhibit.

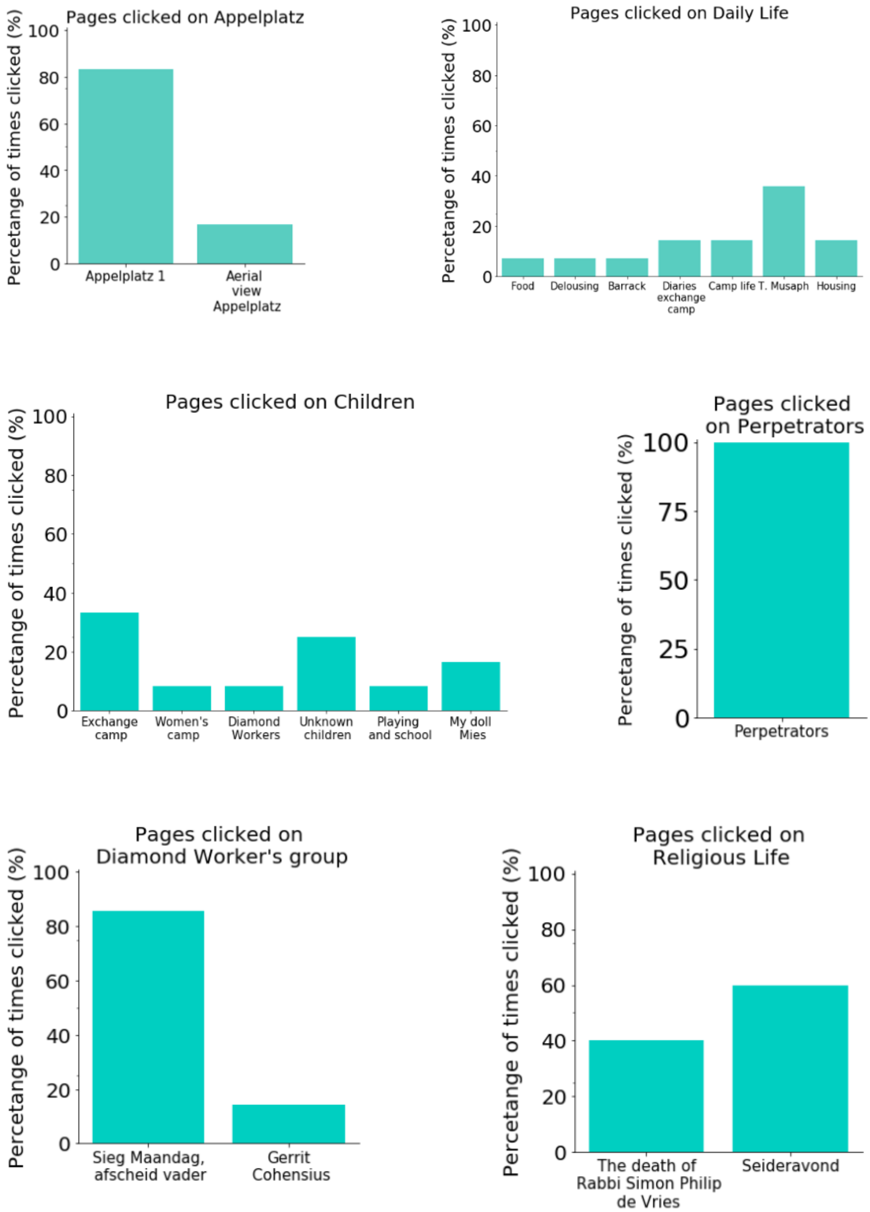


Figure 40. Content visited inside each exhibit.

6.4 Discussion

This paper presents the first steps in the validation of an immersive experience to present information about the Holocaust. The application seemed to induce some changes in users' emotional states. It helped them to understand how the camp worked and increased their interest in knowing more about the topic and visiting similar sites. It was perceived as innovative and interesting.

Moreover, we found a strong focus on content about personal stories. This study shows the possibilities of immersive experiences for presenting information about the Holocaust. The main limitation of this study is the sample size, which we plan to increase in further iterations. Due to time and space limitations, we did not conduct interviews or focus groups, which would give us more detailed insights on users' perception of the experience. Possible next steps would be to implement an introductory video/text, as some users suggested, to introduce visitors to the history of the camp.

Further steps

Experiment data suggested that the above app helped them to better understand different factors like daily life in the historical period, but also functional, and spatial organisation. However, the detailed, continuous evaluation of these and other effects relative to a particular application design remains a challenge, to an important extent because of the particularly complex nature of historical content as stimuli, but also due to the way the app is currently embedded into a museum practice.

For proper evaluation, a practical blend of qualitative and quantitative methods will be necessary (Bryman, 2006). It is important to get better insight into when the found effects work and how they interact, in context of a particular topic context (e.g. cultural biases related to the choice of political or ethical topics like the Holocaust or Soviet suppression), in context of a particular target user, and mode of use (e.g. group collaboration and discussion vs. an isolated user). Such project would focus on a more detailed evaluation of the most important effects suggested by previous research, to systematically improve both the strategies and instruments of continued evaluation, and elements of the application itself.

Other further steps to consider for next studies deal with Tilden's set of principles for effective interpretation practice:

- Generating a more meaningful experience by connecting to visitors' past experience
- Encouraging participants to explore and engage by setting cognitive dissonances through challenges
- Raising interest level by presenting novel, conflicting and surprising content
- Allowing visitors to have more control over their experience by designing a system that facilitates their orientation and navigation
- Supporting participants' needs and interests by providing variety in the presentation
- Enriching the content in an incremental way to facilitate gaining new information

Something to consider is making sure the ratio of content/experience and technology is balanced, as there is the possibility that visitors remember more details about the technology than about the actual HCL experience.

7. STUDY 5: A methodology to teach programming and robotics to children

The previous two studies showed results on individuals' exploration patterns in learning tasks. With that in mind, we assessed the design of a methodology to teach robotics to children, in an inquiry-based learning way.

7.1 Introduction

Teaching computer science (CS) and robotics in primary school should have the goal to provide students with learning tools and skills other than programming or building a robot, to enable them practicing problem-solving, creativity and team-work (Chaudhary, Agrawal, Sureka, & Sureka, 2017). This would allow students to reflect on their solutions (Kalelioğlu, 2015), make confident decisions and develop social, emotional and cognitive skills (Bers, 2017). Thus, these learning tools are consistent with the so-called Four Cs of the 21st Century skills: Communication, Critical thinking, Collaboration, and Creativity.

In 2002, a US-national, non-profit organization was created with the aim of defining the set of skills students needed in the 21st Century. This organization, called the Partnership for 21st Century Learning (also known as P21), established a collective and unified vision for learning, called the Framework for 21st Century Learning. They established what is known as the 4 C's of the 21st Century, that is, the competencies that students need in this century. The 4 C's of the 21st-century skills are:

- **Collaboration:** While collaborating, learners work together to reach a common goal. It helps students to understand how to address a problem and discuss possible solutions with their peers. Through collaboration, children learn from and contribute to the knowledge of others. It also means to be flexible and willing to help support the way to a common goal; and to assume shared responsibility for the work, valuing each team member's individual contribution.

- **Communication:** it is the ability to effectively transfer information in oral, written, and non-verbal forms. It is also to learn to communicate in a diverse range of purposes (for example, to inform, motivate, etc) and in a diverse environment (also considering multi-lingual). It must also be highlighted that communication not only refers to the emission of information but also to listening skills, to be able to communicate in a respectful manner with others. Moreover, considering we are living in a technological age, communication is not only about mastering the process but also the means of it. That is, children need to learn to use the several technological approaches they could use to communicate.
- **Creativity:** children need to learn to try new approaches to solve problems, to promote their innovation. To do so, they can learn a wide range of idea creation techniques (for example, brainstorming). It is also important to promote this idea of failure as an opportunity to learn and to view creation as a cyclical process. The importance of creativity in learning has been previously shown, for example, in Pisa reports (Oecd, 2014), showing the connection between creativity, problem-solving, and academic achievement. Facing situations where there is not a unique correct answer but multiple possible solutions, where children learn to deal with ambiguity and uncertainty.
- **Critical thinking:** it is important that they explore various types of reasoning and they learn to choose the appropriate one for the situation. Children learn to effectively analyse and evaluate evidence and how its parts interact with each other. It is also important to learn to synthesize and identify significant questions. Finally, students learn to interpret information and to rationalize it and draw informed conclusions, while reflecting critically about it. It is related to creativity, as it also concerns looking at problems in a new way.

These four skills reflect the nature of STEM (Science, Technology, Engineering and Mathematics) methodologies: Learning through collaboration means using discussion and production to create a shared output. To do so, it is fundamental to promote the practice of effective communicative abilities between peers. A robotics curriculum can offer students with open-ended challenges to solve, without making them follow step-by-step instructions leading them to the same, preconceived outcome, promoting creativity. Moreover, creative thinking involves learners getting comfortable with taking small “risks". Finally, critical thinking has been compared with “learning how to learn” and metacognition.

It is not surprising, then, that CS and robotics are gaining popularity in school curricula (Karp, Gale, Lowe, Medina, & Beutlich, 2010; Petre & Price, 2004; Varney, Janoudi, Aslam, & Graham, 2012). However, programming can be complicated for young students (Kelleher & Pausch, 2003), as they may face difficulties in learning the syntaxes, coding environments and commands of a programming language (Cockburn & Bryant, 1997). Learning Computational Thinking (CT), defined as the set of thinking skills and approaches essential to solving complex problems using computers, before being introduced to formal programming, would help ease students into the formal languages (Kazimoglu, Kiernan, Bacon, & MacKinnon, 2011; Wing, 2008). Additionally, using tangible environments may help to bypass the need for familiarity with computers (Smith, 2007) and get rid of distraction derived from learning how to use them.

In our studies with primary school children (10-12 years old), we have identified several critical points that need optimization to improve teaching robotics/coding courses. Here we address three of them: 1) excessive focus on theory, 2) preference for personalized hands-on and discovery-oriented projects, and 3) dependency on formal programming languages instead of CT. We thus propose “CREA” as a methodology for teaching CS to this age-group.

a) Educational Robotics

Educational robotics (ER) that is, engaging students in activities focused on building and controlling robots using specific programming tools (Alimisis & Kynigos, 2009), can be a powerful learning approach. ER can help students become active learners, construct new knowledge, and develop essential mental skills by acting as researchers (Gura & King, 2007). Exploring through hands-on experimentation allows children to transform abstract concepts into a more functional real-world understanding (Thuneberg, Salmi, & Bogner, 2018). Previous studies have shown how teaching robotics can positively impact on students' motivation (Arís & Orcos, 2019) promote problem-solving and cooperative learning (Nourbakhsh et al., 2005) and increase learners' interest in STEM careers (Robinson, 2005; Rogers & Portsmore, 2004).

Typically, the type of robotic toolkits used depends on the students' age and ranges from Bee-Bots, LEGO Mindstorms, OWI 535 to NAO and Arduino. The most common coding environments for teaching programming and robotics are Alice (Cooper, Dann, & Pausch, 2000) and the well-known Scratch (Resnick et al., 2009). Scratch is based on graphical programming blocks that when assembled, create a program that provides children with a so-called "open sandbox" environment. This environment includes blocks representing code that can be placed on a scripting area desktop to promote tinkering. Another "open sandbox" environment for novices is the MIT App Inventor, which also promotes learning gains (in this case, more suitable for high-school students) (Papadakis, 2019; Papadakis & Orfanakis, 2018).

Most pedagogical models used in ER tend to focus on tools rather than including the aforementioned 4 C's. It is necessary, then, to examine the predominant pedagogical models currently used in educational activities with robotics, identifying and analysing their advantages and disadvantages (Malinverni, Schaper, & Valero, 2020). The three most common approaches to ER as reported in the literature are project-based, goal-oriented and thematic (Eguchi, 2010).

Project-Based Learning (PBL) organises learning around projects based on challenging questions or problems, that involve students in design, problem-solving, decision making, or investigative activities (Thomas, 2000). Students work in teams, through a process of research and creation, in a relatively autonomous way. When applied to ER, the teacher or the students suggest the main topic of the project, and the goal is to build a robotic device that will solve the proposed problem. Based on this model, different learning approaches depend on the objective to be achieved. However, there is a lack of a universally accepted model or theory of PBL, resulting in various PBL research and development activities.

One example of using PBL methodology to design teaching material is Lego's "Four C" Learning System[12]. Lego divides the learning experience into four phases: Connect, Construct, Contemplate and Continue. Along the process, students are presented with a challenge and instructions on how to build and program the corresponding model. Upon model completion, students are asked to look for new solutions on the same challenge. The provided material represents a straightforward methodology to be implemented in a classroom, and the projects are very defined. As a result, the "what and how" approach to each project is described in a restricted way, and the ability to create and think during the project is limited.

The risk of simply following instructions is that students might fail to internalise the operation required for building or programming the robot. Indeed, in a study where participants had worked for three years with Lego's Four C learning system, when asked to do a project without instructions, several of them had trouble connecting the motors (Botelho, Braz, & Rodrigues, 2012). Examples of the goal-oriented approach are the international competitions of the First LEGO League[13], the World Robot Olympiad[14], the Eurobot[15] and the RoboCupJunior[16]. This approach involves the students as a part of the game, to augment their motivation and improve their attention capacity. The ideation, design, assembly and setting up of the prototype are part of the learning process. The theme-based approach focuses on activities based on a specific topic, and learning occurs through inquiry and communication (Detsikas & Alimisis, 2011).

To explore creativity and socialisation skills (Botelho et al., 2012) proposed a new methodology that does not offer step-by-step building instructions nor solutions for the project. Instead, it focuses

12 https://education.lego.com/en-us/about-us
13 https://www.firstlegoleague.org/
14 https://wro-association.org/
15 https://www.eurobot.org/
16 https://junior.robocup.org/

on robotic design and contains five main steps (Virtual Sketch, Functional Sketch, Concrete Sketch, Construction Prototype and Presentation), where, by creating sketches with different levels of detail, children develop the design of their robot to prototype it. This methodology allows students to create robots in teams without following strict guidelines, highlighting the value of students, while offering great versatility, as it is not limited by the technological tools to use (Frangou & Papanikolaou, 2008).

The new pedagogical models in ER respond directly to some of the gaps of the traditional models (for example, the autonomy and active role of the student, the promotion of creativity and the ability of problem-solving, etc.). At the same time, they introduce key concepts and elements to facilitate significant learning processes. Some relevant examples are the use of the design process (draft) as a way of thinking, games as a way of approaching the materials and narrative as a way of using fantasy in the robotics field. Although there is no consensus on the model employed to teaching robotics, project-based, and goal-oriented approaches are the most predominantly practised in schools.

b) Computational Thinking

One of the approaches to teaching how to program is through Computational Thinking (CT). CT (Wing, 2006). comprises the set of thinking skills and approaches essential to solving complex problems using a computer. CT comprises the procedure of defining a problem and articulating its solution in a way that can be understood by a computer. The original definition of CT came from Papert (we will see more about him in the next section) and was later popularized by Jeanette Wing (Wing, 2006). Nevertheless, there is not a clear consensus for the definition of what CT represents. In general, it symbolises the analytic and problem-solving skills and approaches used in computer science.

CT can be divided into several steps: decomposition, pattern recognition, abstraction, and algorithm design. Decomposition represents the process of breaking a complex problem into smaller and more feasible sub-problems. Patterns, i.e. similar characteristics between parts, are found in sub-problems, and allow finding a common solution for several problems. In abstraction, the key components are selected, removing what is unnecessary and finding one solution for multiple problems. In fact, for Wing, CT mainly focuses on abstraction (Wing, 2006). Finally, in algorithm design, ordered step-by-step instructions are created on how to solve a problem.

Some studies defend the importance of teaching CT to students before they are introduced to formal programming (Kazimoglu et al., 2011; Perković, Settle, Hwang, & Jones, 2010; Wing, 2008). One of the main ways is to teach students how to write algorithms without using code. To do so, there are two main ways to represent algorithms: pseudocode and flowcharts. The first one, pseudocode, comprises a representation of the instructions given to the computer (we should consider that pseudocode is not a programming language in itself, but more similar to spoken language). The most common way is to do it in plain English, although, if English is not your students' first language, other languages can be considered to facilitate the thought process. Flowcharts are a more graphical way to represent the algorithm. They use specific visual representations (as diamonds, boxes, etc) for each part of the code. In any of the cases, it can help students to plan their solutions before facing possible mistakes when writing code. To do so, students use keywords and the variables names they will use later, to facilitate the consequential translation to code.

Some researchers advocate for the idea that to form a solid base for further skill development, CT should already be taught in primary school (in the form of teaching vocabulary, symbology) (Lu & Fletcher, 2009). This is because this idea of taking a large problem and dividing it into smaller, more solvable, steps is a skill that can be extrapolated outside the field of CS. That is, students, by taking, for example, a mathematical problem or an essay assignment, and dividing each step into smaller parts, can help them to work on more manageable problems that sum up to a whole. In the end, this skill comprises executive planning and self-regulation skills, as children need to identify the steps and their order and evaluate what they will need to complete them.

These CT steps can also be useful outside any STEM-related subject. For example, if we consider children writing an essay for a History class. They would first need to break down the essay into smaller parts (decomposition): What can I write about? Where do I take the information from? What writing style do I want to follow? After knowing this, they can relate their work to what they have down in class, identifying similarities (pattern recognition). They will then write a draft, research about the sources, identify the key points of their work, define a structure (abstraction). Finally, they think of the steps needed to write the essay (algorithmic design), for example, they can create a "to do" list or a flow diagram of the structure of their work. As we have seen, CT can be a powerful tool to teach children and teenagers about problem-solving, promoting a more active student.

c) Metacognition in coding and educational robotics

Through the process of debugging, children learn to code by testing, by failing and learning from their mistakes. Coding is an iterative process where the learner can receive constant feedback while debugging their code. That constant contact with failure (despite needing to be controlled to avoid dropping out) promotes resilience and persistence and can teach students how success comes from hard work rather than immediate mastery. This kind of skills can help students persist in the face of challenges and failure. Nevertheless, students require some skills to make good use of feedback, as it can be ignored (if students think they do not need it) or overused (if they do not want to put effort). It is undeniable thus, that in CS and robotics, students need a problem-solving mindset. Programming processes involve problem-solving steps: problem identification, problem definition, strategy formulation, organisation of information, allocation of resources, monitoring and evaluation. In order to do so, children need to plan, control and evaluate their cognitive process.

Planning, designing and debugging have been considered among the most difficult tasks in programming (Sivasakthi & Rajendran, 2011). Specifically, designing a program for a given problem and to divide functionality into procedures. Moreover, problem-solving may be complex for novices, as they may not have experienced that coding issue before. It has been shown that using metacognitive strategies when taking a decision relates to better problem-solving skills (Tan, Ting, & Ling, 2009), and CS can help to activate metacognition by promoting critical thinking. Having more than one correct answer to solve a problem allows students to exercise metacognition, as they can think about several possible solutions and choose the one that fits their task better. Thus, giving students opportunities to reflect on what they are doing and to take responsibility for their own learning are keys to building metacognition. At the same time, it has been shown that high-performers in programming courses use more metacognitive strategies than low-performing students (Teong, 2003).

d) Metacognition and collaboration

Previous research (Jia, Li, & Cao, 2019) has presented the benefits of collaborative settings for learning and the use of metacognitive skills. During any collaborative process, learners need to reflect, focusing on monitoring their comprehension and problem-solving processes.

Metacognition also plays a central role in suprapersonal decision-making (Dunstone & Caldwell, 2018; Heyes, 2015; Shea et al., 2014). It not only enables individuals to monitor their own cognitive processes, but it also enables broadcast and sharing of otherwise private mental states with others. Cognitive offloading often involves depositing information with, or soliciting information from, other agents (Goupil, Romand-Monnier, & Kouider, 2016). For example, when coordinating complex actions in team sports, people use metacognitive representations to decide the contribution of each team member.

7.2 The Framework

Our primary focus is that students learn how to build and program a robot that performs a certain task. Our methodology, "CREA" (accounting for "Coding Robots through Exploring their Affordances"), proposes collaborative inquiry-based activities for learning CS and robotics where the students are at the center of the learning process and where the teacher's role is to guide rather than to provide information. To do so, the process of building and programming a robot can be decomposed into smaller, manageable components. These components, which are shown to benefit students when learning to program (Zhang, Liu, de Pablos, & She, 2014). make use of visual representations, require verbal explanations and promote discovering on their own. In the following sections, we present the technology and the robotic platform employed in our study, as well as a description of the methodology employed to teach young children how to build and program a robot.

a) Tools

The basic component of our robotic platform is an Arduino Uno[17] with a Grove[18] board (for safety and ease of use) on top of which children can connect a variety of sensors and actuators, such as potentiometers, buttons, light sensors and LEDs (***Figure 41***). To program the Arduino, students used Visualino[19], a multiplatform visual programming environment similar to Scratch. Visualino not only provides graphical programming blocks that can create a program when assembled and promote tinkering but also generates native code for the Arduino, thus no computer is required for the Arduino to work.

Students were also provided with worksheets to support some sessions. The purpose of these worksheets was for children to write down the robot's behaviour and later use them to identify the specific components of the robot. After that, these worksheets were used as bases for the robot's programming. In each session, children

[17] https://www.arduino.cc/

[18] https://wiki.seeedstudio.com/Grove/

[19] http://www.visualino.net/index.html

received a different robot (and its related worksheet, filled in by their peers) than the one they worked with the previous session. This exchange was to make sure all the groups get to know all the robots used and to promote collaboration between groups.

Figure 41. Technological devices used in CREA. Left: Arduino Uno board (image from the Arduino website). Right: Grove starter kit (image from the Grove System website).

b) ASSURE Methodology

Instructional design is the practice of developing educational experiences to facilitate learning. It is an iterative process, as it is in constant evaluation. In our case, CREA used an approach similar to the ASSURE model (Smaldino, Lowther, Russell, & Mims, 2008) to develop the proposed methodology (Table 1 shows a comparison between the steps and the authors' case). This model focuses on the importance of learners actively participating in their learning and interacting with their peers. Based on a constructivist approach, it is thought for teachers that want to develop methodologies that integrate technology. The name ASSURE is an acronym representing its main principles:

1. **Analyse learners:** The teacher analyses the needs and characteristics of their learners, to decide the specific strategies and resources to scaffold their learning process. In CREA, we evaluated the students' previous experience with robots with a short test. Groups were divided based on the students' interests.
2. **State objectives:** The teacher decides what should be the learners' outcome, and the knowledge learners will have

obtained as a result of the methodology. In CREA, the objective was to teach children how to design and build their robotics project, using the concepts and tools (sensors and actuators) as specified above.

3. **Select media and materials:** The teacher selects the kind of strategies, technology and materials that will be used in class. For example, what software/hardware will they use to program and build their robot. In CREA, the students would receive general instructions of what step would the class focus on and be provided support as needed through each process. The software used was Visualiano and the hardware, Arduino (with a Grove board on top to facilitate its use and make it safer).
4. **Use of media and materials:** Here, the teacher plans how students will use the selected material. To do so, teachers need to test the technology themselves, make sure it works properly, and think about how to present the material to the students. In CREA, the technology used was tested in advance to optimise the content and learning material. The steps and material were updated based on previous' class observations.
5. **Require learner participation:** This step relates to how the teacher will actively engage students in the use of material, both as a class and at the individual level. This is one of the main strategies for learners' participation in class discussions/presentations. In CREA, students were informed at the beginning of each class how each session would help them for their final project. Figure 3 shows the children working on their projects
6. **Evaluate and revise:** Finally, teachers evaluate the teaching and learning processes. For example, how can this lesson be improved? What were its strengths and weaknesses? Was the material the correct one? Did the class arrive at the planned learning objectives? In CREA, we carried out interviews and a short test with the children after the last session. This allowed us to assess students' perception of the course, their learning process and gains obtained.

c) “CREA” - A methodology to teach programming to children

Our Program consisted of six sessions (units) lasting one hour and a half each (represented in ***Figure 42***). In the **first session** (*Robot Observation*, RO), children formed groups of two or three people and remained in the same group throughout all sessions. Here, each group was randomly introduced to a robotic platform and was asked to observe it, interact with it, identify its behaviours and write them down in a worksheet. The first session served as an introduction to the robots and the following sessions guided the students from initial robotic components observations to the programming and construction of the robot. Thus, students were not only presented with a possible outcome (a fully built robot) before engaging in the task, but they also could decompose it in discrete parts and reflect on how these parts worked in isolation.

For the **second session** (*Identification and Analysis*, IA), The aim was to identify the key elements that compose a robot (inputs, outputs, and pins) and to relate them to the components of programming (functions). To do that, students were organized in groups and each group’s worksheet was transcribed in a printed version and randomly assigned to another group. Students were then asked to analyse the information on the sheets and to identify and classify the inputs, outputs, functions, variables, and pins (the position of the board where an item is connected) and colour-code them similarly to the colours used in Visualino (Figure 1-2). Finally, students were asked to relate the inputs with the corresponding functions and outputs (for example: Button -input- → when pressed, turns on -function- → LED -output-).

In the **third session** (*Paper Blocks*, PB), the worksheets generated in the previous session were again randomly distributed to the groups along with a paper version of the blocks used in Visualino to program each robot. Groups had to first read the instructions from the worksheets, find the corresponding blocks of Visualino and connect them accordingly, almost like assembling a puzzle (Figure 1-3). This session allowed students to familiarize with the Visualino blocks and the programming units, without being exposed to the actual interface. At the end of sessions IA and PB, groups received feedback on their work separately.

In **session four** (*Arduino Pairing*, AP), students familiarize with the software interface. The groups were given papers with code in C (Figure 1-4) and groups had to pair each paper with the corresponding robot. In this way, students are introduced to formal programming and how code is translated into action or behaviour. Afterward, each group was given a random picture of the assignment of PB (an image of Visualino blocks) and were asked to translate it into actual Visualino code. In the **fifth session** (*Code in Visualino*, CV), each group chose a robot and decide on a new design and task to be implemented in the robot, by making changes in the Visualino code. This is in agreement with previous work showing that the ability to choose one's project increases the motivation to perform a task. Finally, in the **last session** (*Robot Construction*, RC), each group applied what they had learned in the previous sessions to program their proposed robot.

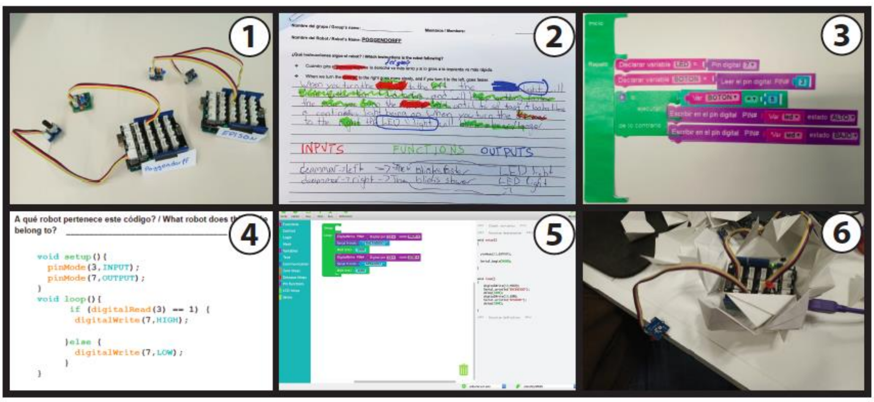


Figure 42. Examples of the six sessions of "CREA". 1) Robot Observation (RO), 2) I Instruction Analysis (IA), 3) Paper Blocks (PB), 4) Arduino Pairing (AP), 5) Code in Visualino (CV), 6) Robot Construction (RC).

7.3 Materials and Methods

a) Overview and Participants

We applied our methodology in an extracurricular course of "Robotics and Programming" in an international elementary school in Barcelona. Our participants, the whole class, consisted of 10 students with 9 boys and 1 girl (age: 11.11±0.60).

b) Experimental Protocol

To evaluate the effectiveness of our methodology, before the first session the students were given a pre-test and at the end of the last session a post-test for their knowledge about CS and robotics. Both tests contained 12 questions of multiple-choice (four answer options) regarding hardware and coding elements that would be covered during the sessions: inputs and outputs used with the Arduino board and coding with Visualino.

Additionally, at the end of the last (sixth) session, we performed a semi-structured interview with each student. ***Figure 43*** depicts the procedure of the interview. The goal of the interview was to acquire insights on how the students evaluated the class. More specifically, we presented the students with images (see Figure 1.) that depicted each of the six sessions and asked them to evaluate them based on the following criteria: likeability (if they enjoyed each session), perceived difficulty, if they would do it again, if they worked alone or in collaboration and if they would remove any part. The evaluation was done by ordering each image according to each question (1-6 scale). The perceived difficulty of the exercises served not only for the students to assess their performance, but also identify what they found challenging (Guzdial, 2008). All the content of the class was presented in Spanish and English.

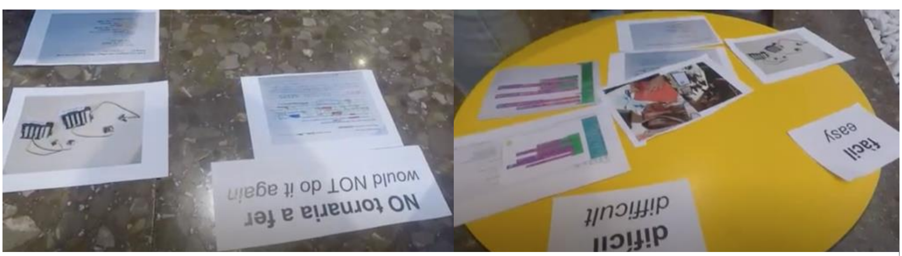


Figure 43. Examples of the procedure carried out during the interview. ***Left:*** Children were asked to order the sessions from "I would not do it again" to "I would do it again". ***Right:*** Children were asked to order the sessions from "Easy" to "Difficult".

7.4 Results

To quantify the learning progress between the different sessions, we used a Wilcoxon signed-rank test. Our results (***Figure 44***) show improvement in programming and robotics-related knowledge after each session ($p = 0.016$), with a mean performance of 33.33%±9.32 in the pre-test and a mean performance of 52.08%±13.01 in the post-test. Regarding the short interview, children highlighted the discovery-oriented approach of the RO session; additionally, they reported enjoying guessing the robot's behaviour. Children's perception of the IA session was conditioned by their writing skills. Nonetheless, it is worth noting that in their descriptions they have used most technical words learned during the session.

The PB session seemed to generate confusion, as the setup could be easily dismantled. Here, we found a negative correlation between how much the students enjoyed the session (likeability) and the perceived difficulty ($r = -0.93$, $p < 0.001$). AP was well received by the students, as it allowed them to move freely around the class and promoted exploratory behaviour. However, students encountered difficulties in understanding the Arduino code. Students also enjoyed the CV session as it allowed them to use the computer. Finally, the last session (RC), had the highest score in likeability as it involved crafting, creativity and allowed them to choose their project (***Figure 45***). A summary of children's evaluation of the sessions is provided in ***Table 10***.

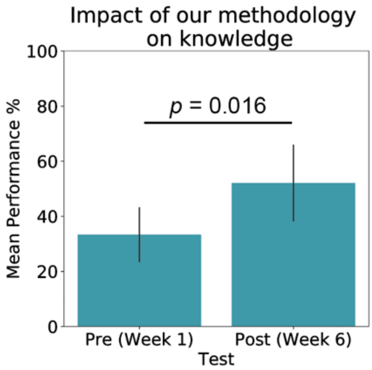


Figure 44. Mean performance at the pre- and post-tests represent improvement in programming and robotics knowledge after the intervention. Lines represent standard deviation.

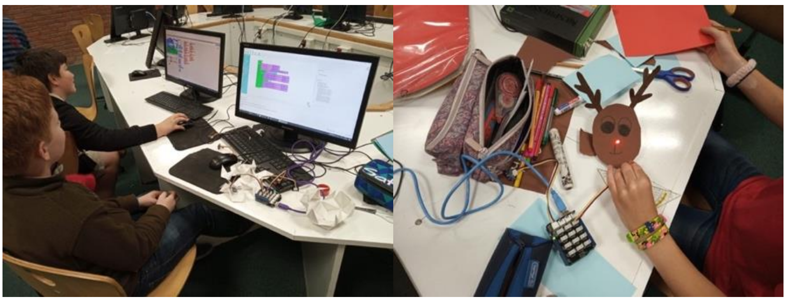


Figure 45. Children working on their projects. ***Left:*** working on the code in Visualino. ***Right:*** Working on the robotic device with Arduino.

Table 10. Children's perception of the different sessions. RO: Robot Observation, IA: Instruction Analysis, PB: Paper blocks, AP: Arduino Pairing, CV: Code in Visualino, RC: Robot Construction. Perc. Diff.: Perceived Difficulty. Asterisk represents significant correlation between likeability and perceived difficulty.

Unit	Likeability (1-6)	Perc. Diff. (1-6)	Again (%)
1. RO	4.28±0.75	4.39±1.36	100
2. IA	2.50±1.37	3.00±1.22	22.22
3. PB*	2.54±1.59	2.55±1.67	44.44
4. AP	3.72±1.30	3.39±1.11	66.66
5. CV	2.83±1.12	3.11±1.96	77.78
6. RC	5.22±0.97	4.56±1.67	100

7.5 Discussion and Conclusion

This study proposes a methodology, "CREA", to teach programming and robotics to primary school children and it consists of six learning stages that aim at scaffolding the students' learning process. We applied and evaluated the proposed methodology in an extracurricular course for primary school students (10-12 years old). Our results show an increase in performance between the first stage and the last stage of the course suggesting that our method was effective as a teaching strategy. Nevertheless, due to the relatively small sample and having no control condition to compare with, we cannot make any general claims.

Regarding the students' evaluations of the learning activities in the course (programming, design and building), children enjoyed more those that involved crafting, creativity and the usage of computers, as well as those that promoted teamwork. When referring to the RO and IA activities, children used verbs like "discover" and "guess"; however, they found the writing part of the IA not interesting and more challenging. When comparing programming using paper blocks or the computers, opinions were divided: some children preferred the constrained possibilities offered by paper blocks while others enjoyed exploring their options using the Visualino software. For future versions of the methodology we will consider a larger and harder version of the pieces in the PB part (even stickers or magnets), so they are easier to assemble.

In comparison, other methodological studies to teach CS and robotics to an age range similar to ours include more (up to 28 sessions in (Barker & Ansorge, 2017)) or longer sessions (up to 150 minutes in (Chambers, Carbonaro, & Murray, 2008)). In the first case, the results show a significant increase in knowledge on Science Engineering technical concepts in the experimental condition compared to a control one. In the second case, results suggest a variability among students' conceptual development. In other cases, though, learners were older than in our study (for example, between 10 and 15 years old (Karahoca, Karahoca, & Uzunboylu, 2011)). Moreover, both of these studies contained sessions focused on learning theory first, something we have avoided to follow an inquiry-based learning approach.

Overall, our results suggest that our inquiry-based methodology can provide the first steps to effectively teaching instructional design of robotics and programming. For example, as children seemed to prefer crafting-related activities and computer-mediated environments, the course activities could be more hands-on oriented but include the use of computers. Additionally, given that students reported having difficulties in understanding the Arduino syntax (C code), to propose alternative ways to teach CS, for instance visual programming may be preferable for such age groups. In our future steps, we will strengthen the methodology to use it in robotics with more components and more complex behaviours. More studies will be needed to validate our methodology, including controlling for factors such as multilanguage background, gender and different learning needs.

In our case, we could not study the possible gender differences in this study, as we only had one female students. Nevertheless, it is an important topic to consider, as the STEM-related subjects are highly affected by gender stereotypes. Studies have repeatedly reported that math and science are perceived as male domains, and scientists as predominantly male (Makarova, Aeschlimann, & Herzog, 2019). Male students show higher levels of confidence and self-efficacy towards STEM subjects than their female peers (Atmatzidou & Demetriadis, 2016; Johnson, 2003). These feelings not only relate to perceived ability but also perceived belongingness to a field (Gomoll, Hmelo-Silver, Šabanović, & Francisco, 2016; Master, Cheryan, & Meltzoff, 2016). Possible reasons behind these differences in self-perception could come from male students tending to be more familiar with technology (Papastergiou, 2009) and have significantly more prior programming experience (Witherspoon, Higashi, Schunn, Baehr, & Shoop, 2017). Some studies have shown no gender differences in performance but in each gender's accessibility to the learning material. Moreover, these differences in self-perception seem to arise from prevalent negative stereotypes of female students' abilities in CS.

Our next steps will focus on further developing the CREA methodology taking into account the feedback obtained from the children's interviews. Another possible way of teaching about the peripherals (inputs and outputs) of the robotic platform would be to, in line with the Session 1 of CREA, promote students' observation through a matching-pairs game, where students have to guess the functionality of each peripheral (***Figure 46***). In this activity, students are provided with Arduino/Grove peripherals and papers with their corresponding names (colour-coded depending on the type of peripheral (sensor or actuator). They have to observe them, classify them between the type of peripheral and try to guess to which paper they correspond to.

Figure 46. Example of the activity on peripherals observation.

8. STUDY 6: Prediction and Collaboration in Autism Spectrum Disorder

During this thesis, we have assessed the differences between perceived and actual performance and the use of explorative and collaborative methodologies in learning. Given the importance of inclusivity in education, we wanted to explore collaborative tasks in students outside the typical neurological paradigm. In this case, we focused on autism spectrum disorder, considering its implication in social-related settings and the increase of diagnostics in the last years.

Blancas, M., Maffei, G., Vouloutsi, V., Sánchez-Fibla, M., & Verschure, P. F. M. J. (2020) *Collaboration variability in Autism Spectrum Disorder*. In Frontiers in Human Neuroscience

8.1 Introduction

a) Autism

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder, whose main impact falls in two domains: persistent deficits in social communication and restricted, repetitive patterns of behaviour or interest (DSM-5 American Psychiatric Association -APA-, 2013). It has been linked to a deficit in prediction abilities and to the fact that feedback is more prominent compared to feed-forward anticipatory control (Schmitz, Martineau, Barthélémy, & Assaiante, 2003; Sinha et al., 2014; S. S. H. Wang, Kloth, & Badura, 2014).

Recent research (Sinha et al., 2014) suggests that a prediction deficit present since early development (Prediction Impairment in Autism, PIA hypothesis, in Sinha et al., 2014) could cause the diversity of expression of the autism syndrome. This theory divides the prediction difficulties among insistence on sameness, sensory hypersensitivities, interacting with dynamic objects, theory of mind, and islands of proficiency. Insistence on sameness represents repetitive actions and thoughts, inflexible adherence to routines, resistance to change, and ritualized patterns of verbal or nonverbal behaviour. Sensory hypersensitivities refer to the sensory abnormalities (like hypersensitivity to bright light) experienced by individuals in the spectrum; however, these abnormalities are not caused by abnormally enhanced sensation. Individuals in the autism spectrum also have difficulties with *Theory of mind* (that is, inferring mental states to others and ascribe causes to observations about a person through the connection of previous with current behaviour), which can cause deficit-adjusting behaviour to suit different social situations. Finally, they can exhibit enhanced abilities in strongly rule-based domains (known as islands of proficiency). These domains, like mathematics, musical performance, or calendar calculations, are strongly rule-based, which minimizes uncertainty.

Individuals with ASD show attenuated top-down prior expectations, which leads them to rely more on bottom-up sensory signals. They thus experience hypersensitivity, enhanced perception and sensation, and sensory overload (Mitchell & Ropar, 2004). Consequently, this dependence on perceptual evidence merged with impairments in contextualizing sensory evidence, impedes understanding actions, and predicting social intentions. Nevertheless, individuals with ASD do not show difficulties in perceiving social stimuli, but rather in using them to update internal models of social interaction, what leads to impairments in social abilities (D'Cruz et al., 2013; Robic et al., 2015; South, Newton, & Chamberlain, 2012).

b) Metacognition in autism

As introduced in the previous section, individuals in the spectrum seem to show diminished *Theory of mind* (also termed *mindreading*), that is, they show lower ability to attribute mental states to others to explain and predict their behaviour. This ability is thought to rely on the same metarepresentational ability as metacognitive monitoring (D'Cruz et al., 2013; Robic et al., 2015; South et al., 2012). For example, previous research has shown diminished accuracy in confidence reports in children with ASD when compared to neurotypical participants (Carruthers, 2009).

It has been proposed that this self-referential metacognition can be related to other-referential metacognition, both of which may have similar cognitive and neural processes. It is important, then, to distinguish between self-referential metacognition (termed 'metacognition') and other-referential metacognition (termed 'mindreading'). Commonly, research on ASD has studied mind-reading in individuals in the spectrum, exploring their abilities to infer the mental states of other people (Grainger, Williams, & Lind, 2016; McMahon, Henderson, Newell, Jaime, & Mundy, 2016; Wilkinson, Best, Minshew, & Strauss, 2010). Nevertheless, less is known about metacognition in people with autism (Baron-Cohen, 1997; Baron-Cohen, Jolliffe, Mortimore, & Robertson, 1997).

c) Sensorimotor integration

Sensorimotor integration can be defined as the brain process allowing to respond to specific demands of the environment by executing voluntary motor behaviour (Machado et al., 2010). Planning and executing a simple movement require sensory feedback, to effectively coordinate movement while acting. Thus, sensorimotor approaches consider perception and action as a united process. This interaction between action and perception must be highlighted in sensorimotor approaches, as they are not seen as separate processes. On the contrary, actions are conferred an integral function for perception to explain cognitive functions.

To consider an anticipatory effect as reflecting prospective sensorimotor control, an action has to differ depending on the subsequent one (Ansuini, Cavallo, Bertone, & Becchio, 2015; Rosenbaum, Chapman, Coelho, Gong, & Studenka, 2013). Sensorimotor contingencies (SMCs) can be seen technically as forward models that predict the expected sensory changes given a certain set of movements. Knowledge of SMCs allows an agent to simulate potential outcomes of behavioural alternatives. Impairments in sensorimotor integration could lead to ineffective use of sensory feedback in, for example, movement correction. As a result, the individual could face difficulties in coordination and sensory reactivity.

The main brain areas associated with sensorimotor integration are the cerebellum (Glickstein, 1998; Paulin, 1993) and the basal ganglia (Chukoskie, Townsend, & Westerfield, 2013; Nagy, Eördegh, Paróczy, Márkus, & Benedek, 2006). It is not surprising, therefore, the significant differences found in these specific areas of autistic patients. For example, previous research showed a lower number of Purkinje cells in the cerebellum (Amaral, Schumann, & Nordahl, 2008; Bauman & Kemper, 2005) and a decreased volume in the basal ganglia (Estes et al., 2011) in ASD individuals as compared to typically developed ones.

The cerebellum is suggested to control the anticipatory and predictive adjustments of motor programs (Koziol, Budding, & Chidekel, 2012). Its pathways link sensory signals to motor areas in the brain (Glickstein, 1998), which have a pivotal role in controlling and coordinating movement (Paulin, 1993). Research on autism has provided ample evidence that the cerebellum is among the most frequently disrupted brain regions in ASD (Courchesne, Redcay, Morgan, & Kennedy, 2005; Palmen, Van Engeland, Hof, & Schmitz, 2004), with persistent differences in volume emerging since the first two years of life (Hashimoto et al., 1995; Stanfield et al., 2008). Studies suggest that ASD is characterized by alterations of the brain's inference on the causes of socially relevant signals, and this lack of ability to predict actions of other individuals stems from cerebellar dysfunctions (Schmitz et al., 2003; Sinha et al., 2014; S. S. H. Wang et al., 2014).

The basal ganglia play a functional role in sensory integration and motor control (Nagy et al., 2006). This area, reciprocally connected to the cerebellum (Chukoskie et al., 2013), has previously been claimed to be different in individuals with autism. For example, it has a lower volume than typical brains (Estes et al., 2011), and one of its areas, the striatum, shows larger functional connectivity in individuals with autism (A. Di Martino et al., 2011). Previous research has shown weak connectivity between sensory and motor brain areas in individuals with autism (Oldehinkel et al., 2019). These findings are consistent with the sensory symptoms (such as hypersensitivity) experienced in ASD. They are also in line with work showing out of sync interactions between visual and motor regions in individuals in the spectrum.

The aforementioned alterations in sensory input and motor execution could play a pivotal role in autism. The available evidence seems to suggest that autism shows wide-spread disturbances in sensorimotor behavior (Cook, Blakemore, & Press, 2013; Gowen & Hamilton, 2013; Haswell, Izawa, R Dowell, H Mostofsky, & Shadmehr, 2009; Rinehart & Mcginley, 2010; Thompson et al., 2017). Along similar lines, self-reports about sensorimotor behavior coming from people in the spectrum provide further evidence on sensory alteration and over-responsivity (Ben-Sasson et al., 2009; Kern et al., 2006; Tavassoli, Hoekstra, & Baron-Cohen, 2014).

Some examples of sensorimotor alterations in ASD comprise impaired motor processing and higher detection of unattended changes compared to neurotypical individuals. There is support presenting these impairments in movement and sensory responsivity not as a peripheral feature of autism, but as a fundamental cause of the social and communicative impairments seen in the condition (Hilton, Graver, & LaVesser, 2007; Leary & Hill, 1996; Matsushima & Kato, 2013; Reynolds, Sammons, De Fraine, Townsend, & Van Damme, 2011). Sensorimotor difficulties in autism are associated with the development and maintenance of social impairments characteristic of the disorder. Integrating sensory information from the environment is required to plan and execute movement effectively, to, altogether, carry on proper social reciprocity.

d) Prediction and collaboration

The so-called social symptoms encompass deficits in social interaction and communication. These poor “social-specific” priors compromise their interaction with others, as ASD individuals have difficulties in coping with the uncertainty that comes with social behaviors (Chambon et al., 2017). Acting together with another partner requires considering and integrating both one’s own and the partner’s next action. This planning of cooperative actions, although less studied, is also considered an aspect of sensorimotor control (Sebanz, Bekkering, & Knoblich, 2006).

The relation between sensorimotor impairments and social deficits in autism suggests impairments in the coupling of perceptual and social cues. More specifically, ASD individuals may encounter difficulties using the sensorimotor contingencies exhibited by another agent to predict the agent’s behaviour. Thus, this work focuses on the evaluation of the coupling of perceptual and social cues based on sensorimotor interaction and the ability to predict another agent’s behaviour. More specifically, we aim to assess how predictive abilities affect collaborative interaction and how they differ between ASD and Typically Developed (TD) individuals.

To do so, we devised a predictive game task where participants collaborate with a synthetic agent that displays different behavioural patterns (sensorimotor contingencies). This task is an adaptation of the game of Pong where players in collaboration with a synthetic agent need to intercept a falling target (see the following section for more information). To succeed in this task, players need to identify and learn the social characteristics of the agent. Doing so will allow them to use this information both during the interaction and in the last part of the task (where each trial has a random agent). As the agent’s behaviour is based on sensorimotor information, we hypothesise that ASD individuals will show deficits in successful social predictive/anticipatory skills. To assess the differences in prediction between ASD and TD players, we look at aspects of adaptive collaborative skills by analysing the interaction of the players with the AI agent of the game, how the interaction evolves during the task and how it relates to the participants’ understanding of the other agent’s characteristics. More specifically, we study partner monitoring and how it affects the covered space and look at the mutual influence between the player and the AI-controlled agent.

e) Hypotheses

- Participants in the autism spectrum will show slower and less adaptation to the other agent than neurotypical ones
- Participants in the autism spectrum will show less adaptation to the other agent when the task becomes more uncertain
- Participants in the autism spectrum will show more variable behaviour than neurotypicals

8.2 The Scenario

a) The Task

The purpose of the current study is to evaluate how goal-oriented coordination between partners could be achieved through sensorimotor adaptation. To do so, we designed a collaborative multiplayer version of the game of Pong: a computer version of a 2D tennis where two players try to intercept falling targets from the top before they hit the ground by moving their paddles at the bottom of the screen. The paddles move on the same horizontal line and can push each other but cannot switch sides. In this game, one player is AI-controlled, and the other is a human. ***Figure 47*** represents an example of the proposed scenario.

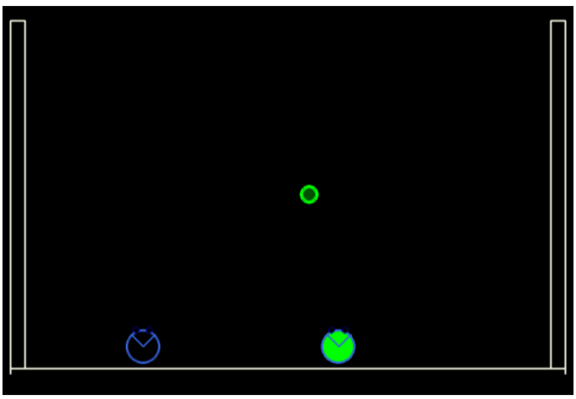


Figure 47. Example of a trial during the task. Targets fall from the top of the screen and players need to intercept them before they hit the ground. The player on the left (blue) is controlled by a human and the one on the right (green) is the synthetic agent. This example represents interaction with the “Middle” agent.

For this task, we considered a collaborative team task like playing tennis doubles, where each player should cover a maximal part of their field so that all targets return to the opponent’s side. Targets sometimes fall in the middle part of the field, thus in a zone where both players could intercept the target. The location of the target was randomly selected from a uniform distribution of possible angles, and the pace of the target drop was uniform across all trials. The velocity of the artificial player was controlled and the same across all trials and the velocity of the participant depended on their motion on the trackpad. A player can be characterized by the area they cover and intervene, given the target's direction. Typically, in a game of two, the area covered by each player is half of the playable area. However, more active players may sometimes overpass their area to try and catch ambiguous targets directed towards the middle area. Collisions with the other agent were penalized by subtracting a point, and participants were informed about the penalty before beginning the task. To evaluate whether the synthetic agent's behaviour and predictability can influence the humans' behaviour, we varied the playing styles of the agent.

The AI-controlled player differs in the way it approaches the target and the area in which it will intervene, resulting in three different agents: "Wider", "Narrower" and "Middle". A "Middle" agent will try to intercept any targets that fall within its half of the space and has a 0.5 probability of intercepting an ambiguous target that falls in the middle. A "Wider" agent will try to intercept the target and overpass its area to try and catch a target even if the target's position is not ambiguous. In contrast, a "Narrower" agent would try to intercept the target without overpassing its area; in fact, it would cover a space that is smaller than half of the overall space. The next section explains in more detail how the agent's behaviour is obtained.

b) The Point of Social Subjective Equality

To measure the collaboration between the human player and the AI player, we introduce the Point of Social Subjective Equality (PSSE). The PSSE can be computed for every two players and all possible target trajectories. This measurement is an analytical measure of collaboration (i.e., social affordance gradient) that defines the probability of going for the target depending on the target's position (***Figure 48***, left). Therefore, the PSSE is the point where each player has the same probability of going to intercept the target (***Figure 48***, right), and is an extension of the Point of Subjective Equality (PSE) (Stoloff, Taylor, Xu, Ridderikhoff, & Ivry, 2011) to a socially collaborative task. PSE represents the point where there is an equal probability of using any of the two hands to reach a target (presented from left to right circularly in front of the participant). Thus, the Point of Social Subjective Equality indicates how a player is relying or not on the partner, invading, or not the partner's area of the field while intercepting targets in the horizontal range. In short, it is the point where a player has an equal probability of intercepting the target or letting their partner intercept it.

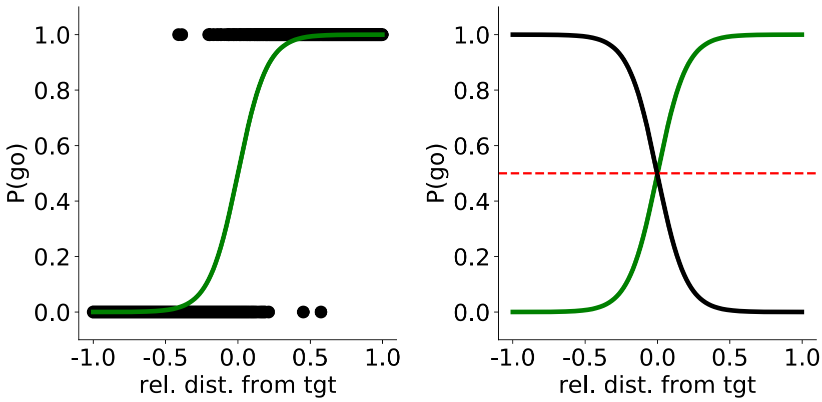


Figure 48. Probability to go depending on the relative distance to the target and PSSE. The left image represents the probability to go for the target depending on the player's relative distance to the target. The right image depicts the Point of Social Subjective Equality (PSSE). Here, the green line represents the AI agent (in this case, the "Middle" one) and the black line, a simulated perfectly matched participant. The red dashed line represents the moment when both agents have the same probability of going for the target. The x-axis represents the relative distance of the agent from the target; and the y-axis, the probability it has to go for the target.

To calculate it, we first calculate the relative distance from the player to the target (Equation 1), that is, the difference between one player to the target and the other player to the target. After that, we fit a sigmoid function with the distance to the target (*rel_dist*), a constant factor (*k*), and a bias value (*b*) representing the behaviour of each of the agents (Equation 2). We estimated the parameters of the PSSE (*k*, *bias* and *rel_dist*) by running a Logistic Regression using sklearn[20]. To our knowledge, this is the first time such a direct behavioural measure of collaboration is introduced.

$$rel\_dist = \frac{(|p_t - p_{p2}| - |p_t - p_{p1}|)}{w}$$

Equation 1. Representation of the relative distance (rel_dist). $p_t p_t$ represents the position of the target, $p_{p2} p_{p2}$ represents the position of the other agent, $p_{p1} p_{p1}$ represents the position of the participant, and $w w$ represents the width of the (game) screen.

$$PSSE = \frac{1}{1 + e^{-(k*rel\_dist+b)}}$$

Equation 2. Representation of the Point of Social Subjective Equality (PSSE). *b* represents each partner's bias, *rel_dist* is the relative distance from the target, and *k* is a constant factor (k=20).

Based on the PSSE, two complementary partners would intercept the target with the same probability (P = 0.5, ***Figure 48***, right), whereas any shift would indicate a lack of balance between the partners. As mentioned previously, participants play with three different AI agents, and we modulated their behaviour based on this shift of the interception point. Our three proposed agents, namely "Middle" (M), "Wider" (W), and "Narrower" (N), have therefore different probabilities of intercepting the target. More specifically, the "Middle" agent has a 0.5 probability of going for an ambiguous target (when the target falls in the centre of the arena). A "Wider" agent is more prone to invade the space of the participant; therefore, the curve of the probability to intercept the target based on the target's location would fall towards the left part of the space. In contrast, the "Narrower" agent is more prone to stay in its half of the space and allow the participant to enter the AI agent's space to catch the target. Consequently, the curve would fall toward the right part of the space. Thus, if we split the playable area into two equally sized sides, one for the participant and the other for the synthetic agent, a "Middle" agent would cover only its 50% of the space, while the

[20] https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html

“Wider” would cover more than 50% and the “Narrower”, less. ***Figure 49*** provides an example of the representation of the curve for each AI agent.

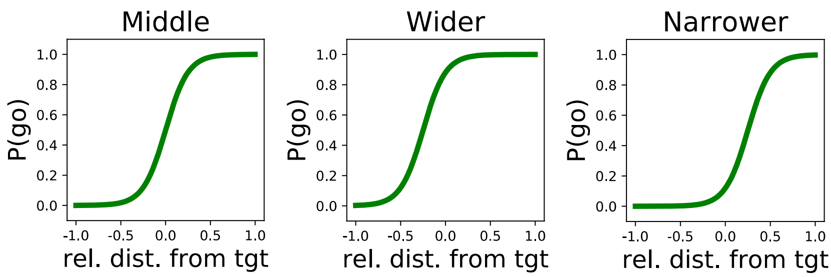


Figure 49. Representation of the curve of the probability to intercept the target based on the target’s location for each of the three proposed agents. From left to right: the curve of the “Middle” agent lies in the middle as both the synthetic agent and the human player have the same probability of intercepting the target. In contrast, the curve of the “Wider” synthetic agent is slightly skewed toward the left, as this agent will enter the space of the human participant. In contrast, the “Narrower” agent’s curve is skewed toward the right; this agent has a higher probability of staying toward its half of the space and allowing the human participant to intercept the target.

The agents were programmed to catch the target following a pre-defined strategy (M, W, or N). Consequently, if a participant decided to leave the target to the artificial agent, the agent’s behaviour would depend on the predefined strategy and therefore the position of the target and the relative positions of the two players. Thus, there would be cases where the ball would be intercepted by the artificial agent and others that it would be missed, however, the PSSE sigmoid function would not be affected by the movement of the human player. The coefficient and the intercept of this curve will allow us to assess participants’ adaptation to the other agent.

8.3 Materials and Methods

a) Participants

The ASD participants recruited for the study had previously been diagnosed as autistic meeting the DSM-5 criteria for level 1 of Autism ("Requiring support", American Psychiatric Association, 2013) (N=15, 1 female, age: 18.67±2.4). This criterion comprises difficulties in initiating social interactions and switching between activities. This group was recruited in the Educa Friends centre[21], an educational support service part of the Friends Foundation, focused on providing support to high-level functioning ASD individuals. All participants had normal or corrected-to-normal vision and were not colourblind. The typically developed participants were recruited in a high-school of Barcelona and the campus of the Polytechnic University of Catalunya, and their age-matched those with ASD (N=15, 2 females, age: 18.38±1.06). Written informed consent was obtained for all participants (for the ones under the age of 18, parental written informed consent was obtained too). The study was approved by the local ethical committee (Parc de Salut del Mar).

b) Experimental Protocol

All participants were provided with an information sheet that contained the explanation of the task and a consent form they had to sign before beginning the experiment. For the minor participants, information sheets were given to both participants and their parents / legal tutor, together with a consent form for the parents / legal tutors and another for the participants. Additionally, participants filled in a small questionnaire with demographics and the frequency of playing video games. Once participants completed the questionnaire, the game-like scenario would commence. As mentioned earlier, the task is a Pong adaptation and the goal is to intercept falling targets. The task was performed in a computer using a touchpad. The task consisted of three main phases. In "Phase 1", participants played alone for one block. Each block consisted of 150 trials. In "Phase 2" participants played for one block with each of the three AI players (in total three blocks) and finally, in "Phase 3" participants played for one block with all agents. The order of the three AI players in "Phase 2" was randomised and for each type of the three AIs the player would play for 150 trials. "Phase 3" was used to assess the

[21] https://fundaciofriends.org/es/servicios/educafriends/

social predictive abilities of the participants, as they had to interact with a random agent in every trial (counterbalanced so there were 50 trials with each agent). Each of the three agents was depicted in a different colour. Colour choices were made arbitrarily.

In this study, we used behavioural data, questionnaires, and interviews as instruments to collect information about the subjects' behaviour and perception of the task. Between each of the blocks, participants had to answer questions in a tablet. The questions involved perceived collaboration and predictability of the target and the other agent and engagement. To answer, participants had to rate each of them on a Likert scale from 1 to 5. At the end of the task, a semi-structured interview was carried out, to assess the perceived differences between agents. ***Figure 50*** represents the experimental protocol. In total, the whole experiment took around 30 minutes.

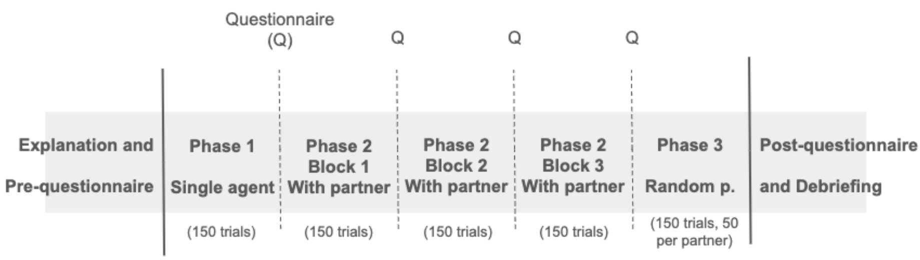


Figure 50. Representation of the experimental protocol. First, participants are introduced to the task and fill in a short questionnaire. The task comprises three phases, which contain one, three, and one blocks (of 150 trials per block), respectively. In "Phase 1", participants play alone. In "Phase 2", participants randomly play with each of the AI agents for one block (three blocks in total). In "Phase 3", participants are presented with a random agent in each trial (50 trials per agent). Self-reports on perceived collaboration, engagement, and agents' and target predictability are presented between phases/blocks. Finally, participants fill in a short questionnaire and undergo an interview and debriefing.

c) Measures

The data gathered for this study comes from three different sources: the behavioural data coming from the task, the questions between the blocks and an interview at the end of the task. The first source, the logs of the game, is obtained in a trial by trial basis and is comprised of the *performance* for that trial (one point if either the participant or the other agent reach the target), the *identifier* for the other player (to know which other agent they were playing with), and the *positions of the player*, the *target* and the *other agent*. Those last three allowed us to obtain the *PSSE* measure, as explained in Equation 1.

The data from the questionnaire allowed to assess participants' perception of the other agent and the task. Thus, the variables that arise from those questions are *agent predictability* and *collaboration* (asked after all the parts where there was another player; that is, all but part 1) and *engagement* and *target predictability* (asked after all the parts). They were all reported in a 5-point Likert scale. Moreover, at the end of the task, they were requested to answer if they thought the other player was another human or a computer, also adding "I don't know" as a possible answer.

Finally, the interviews allowed us to assess participants' thoughts about the task and more detailed information about their perception of the other agents. Participants answered, in this order, how difficult they perceived the task, what they thought about the other player, their description of each of them (if they reported finding differences). They also reported how they perceived the random parts in terms of difficulty and if they followed any strategy. After finishing, the experimenter asked if they had any question about the task and debriefed them by explaining them the study.

d) Tools

Participants sat at a viewing distance of (approximately) 50 cm from a 27-inch monitor that operated at a resolution of 1920x1080. The monitor was part of an All-in-One desktop computer connected to a touchpad and a keyboard. The task was generated using Python and the PyGame library. ***Figure 51*** depicts the setup used.

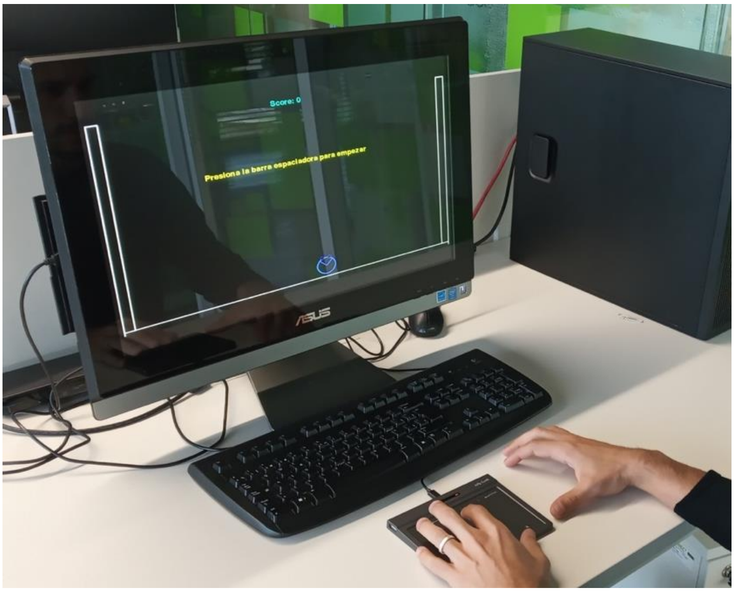


Figure 51. Representation of the setup used in the task. Participants sit in front of a computer screen where the game was displayed. Participants controlled the motion of their avatar using a touchpad.

8.4 Results

The following results have been analysed in Python, using the following libraries: NumPy, JSON, math, scipy, and sklearn. In order to choose the statistical tests used in this analysis, we ran normality tests in the variables. The intercepts of the PSSEs during “Phase 2” had a normal distribution, so parametric tests were used (One-Way ANOVA, in this case). The mean squared errors in “Phase 2” did not show a normal distribution, so non-parametric tests were used (Mann-Whitney U).

We defined performance as the ratio of caught targets out of the 150 of each phase. A Mann Whitney U test (U = 892448.0, p = 0.433) showed that there were no significant differences between the two groups (ASD: median: 1.0, MAD: 0; TD: median: 1.0, MAD: 0) in *performance* during “Phase 1” (when participants played alone). Thus, possible differences in “Phase 2” and “Phase 3” should not be related to their performance when playing alone.

a) Participants in the autism spectrum showed more variable behaviour than neurotypicals

To assess participants' adaptation to the artificial player, we calculated the Point of Social Subjective Equality (PSSE). First, we look at the two groups' behaviour in "Phase 2", where we take into account all trials with each agent per block. To analyse the differences between groups and agents, we calculated the differences between the coefficients among groups for the same agent, and among agents for the same group. There were no differences between agents in their coefficients in none of the groups (***Figure 52***). In terms of intercepts, there were significant differences between agents in both the ASD (One-way ANOVA(6749) = 5.68, p = 0.007) and TD group (One-way ANOVA(6749) = 10.83, p < 0.001). More specifically, an independent samples t-test showed differences in the ASD group were between the Middle and Narrow agents (t(4499) = -2.11, p = 0.04) and the Narrow and Wider agents (t(4499) = 3.46, p = 0.002); and in the TD between the Middle and Narrow agents (t(4499) = -5.17, p < 0.001) and the Narrow and Wider agents (t(4499) = 4.01, p < 0.001).

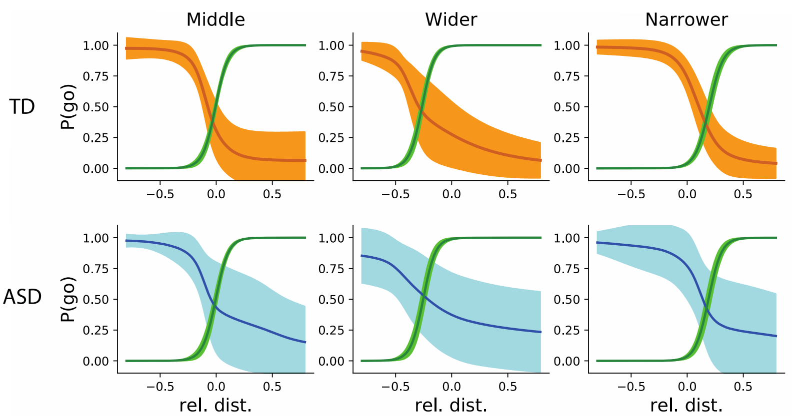


Figure 52. PSSE intersections from the TD and ASD groups (top and orange and bottom and blue, respectively). From left to right: the PSSE curve of the AI agent (depicted in green) and that of the participants, depicted in orange (TD) and blue (ASD), when playing with a "Middle" (N), "Wider" (A) and "Narrower" (S) agent respectively. The bright colours represent standard deviation, while the darker and thicker line represents the mean.

The lack of significant difference between slopes could mean that, generally, both groups adapted in a similar way. Nevertheless, as we can see in ***Figure 52***, participants in the ASD group showed a higher probability of going (approximately 0.25) with a relative distance larger than 0. This means that they had more tendency to go towards the target than the TD group (which probability at that time was around 0.07), even when they should not. The differences in intercept represent the adaptation of each group to the specific agent they were playing with. In the next sections, we will quantify the variability of each group and their behavioural changes with respect to the other agent.

To assess the variability among participants in each group, we calculated the mean squared error between each participant and the general mean. To do so, we first calculated the general mean of the coefficients extracted from the data points obtained in all trials in all blocks from “Phase 2” in both groups. From that, we calculated the average of those data points and obtained a representative mean squared error per group (ASD: 0.08 ± 0.08; TD: 0.001 ± 0.001). A Mann Whitney U test was used to analyse the differences between groups against the general mean ($U = 9.0$, $p < 0.001$). Moreover, when assessing the variability inside of each group (that is, the variability compared to the mean of their group), the difference is even higher ($U = 0$, $p < 0.001$). The U equal to zero signifies that all the mean squared errors in the ASD group are greater compared to all the ones in the TD group.

b) Participants in the autism spectrum showed slower adaptation to the artificial agent than neurotypical ones

To further understand the two groups’ adaptation, we then looked at a possible evolution in time of the PSSE, and more specifically, whether early (50 first trials) and late (50 last trials) trials differed between the groups in “Phase 2”. To do so, we analysed the shift in PSSE for each of the agents. We used the “Middle” agent as a baseline and subtracted from it the shift for the “Wider” and “Narrower” agents. Like this, we could calculate how much the participants’ behaviour changed when encountering the “Wider” and

"Narrower" agents. As we can observe in ***Figure 53***, we found statistically significant differences between the two groups for the "Wider" agent in the early trials (U= 65.0, p = 0.042) but not the late trials. We did not find any statistically significant differences between groups for the "Narrower" agent in both early and late trials.

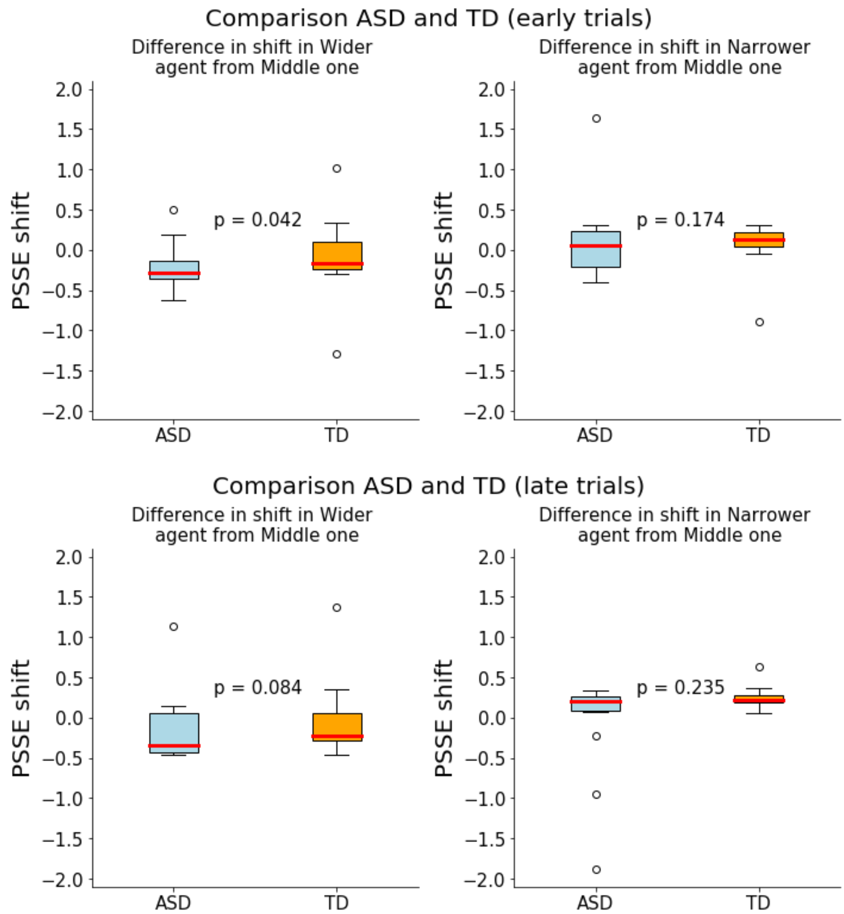


Figure 53. Differences between groups in the difference in shift per agent (compared to the "Middle"one). The upper plots represent early (0-50) trials and the lower plots, late (100-150) trials. Blue represents the ASD group; orange, the TD; and the red line represents the median.

c) Participants in the autism spectrum showed less adaptation to the other agent when the task became more uncertain

During “Phase 2” we have shown that healthy subjects acquired an ad hoc behavioural strategy (i.e. PSSE shift) from the interaction with each individual agent and that the adaptation process was more pronounced in healthy subjects compared to control. During “Phase 3” we aim at assessing whether this strategy can be correctly retrieved when the subjects interact with each agent in a randomized order. We hypothesise that the ASD group will be less able to retrieve a correct strategy, potentially due to the reduced ability to form an internal model of the partner. To do so, for every subject we compute the PSSE associated to each agent during “Phase 2”.

In “Phase 2”, participants played for one block with each of the three AI agents. In contrast, in “Phase 3”, participants encountered a random AI player in each trial for one block. As mentioned earlier, the characteristic that distinguishes the agents’ behaviour is the colour, and if players have not made the colour association with the agent’s behaviour, “Phase 3” becomes more uncertain. Here, we wanted to assess how much the players’ behaviour in “Phase 3” matches that of “Phase 2” when playing with the same agent during each of the blocks. To do so, we ran a logistic regression using participants’ behaviour during “Phase 2” as our “training data”, and compared against their behaviour during Phase 3, which was used as “testing data”.

PSSE for each agent is described by a logistic function with constant k and intercept i. We further group the trials from “Phase 3” according to the agent type and extract, similarly to “Phase 2”, the probability of the subject to go for the target or to let the partner go (p=1 and p=0 respectively). Finally, we compute for every agent how accurately the parameters of the PSSEs from “Phase 2” describe the behaviour (i.e. probability of going for the target) observed in “Phase 3”. The rationale is that high accuracy of the model from “Phase 2” in describing the behaviour of “Phase 3” would confirm the hypothesis that a behavioural strategy tight to each individual agent has been learned and can be correctly retrieved.

In the left panel of ***Figure 54***, we show the mean accuracy matrix for the control group and the ASD group. This is obtained by computing for every subject the accuracy of each PSSE agent model (predicted) in describing the data of each agent during "Phase 3". This generates a set of 3x3 matrices that are further averaged for each group. This result suggests that the participants in the neurotypical group (right, orange) behaved in the same way in both phases, with a mean accuracy score of 0.97). However, the participants in the ASD group (left, blue) did worse in properly matching their behaviour to the one in the previous phase (mean accuracy score of 0.74). A Mann Whitney-U test showed significant differences between the accuracy for both groups in matching Phases 2 and 3 behaviour ($U = 0$, $p = 0.03$) These results could suggest that participants in the TD group developed a model of the other player during "Phase 2" that they later used to adapt their behaviour in "Phase 3"; participants in the ASD group failed to do so.

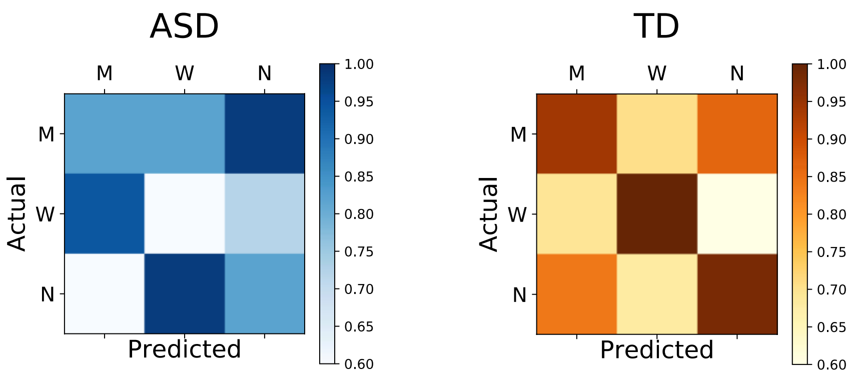


Figure 54. Matrix representing the relationship between Phase 2 (actual) and Phase 3 (predicted) behaviour. The blue matrix represents the group in the autistic spectrum, while the orange one represents data from the neurotypical group.

d) No differences in perception of the task between groups, only by perceiving the other agent as human or synthetic

As previously mentioned, participants had to answer a short questionnaire between blocks. More specifically, participants evaluated target predictability, engagement (in all blocks), as well as agent predictability and collaboration (in the blocks where the AI agent was introduced). There were no significant differences in engagement or target predictability between "Phase 1" and the rest of the task. In "Phase 2", participants rated each agent at the end of each respective block (***Figure 55***). Results suggest no statistically significant differences between groups in any of the dimensions. Nonetheless, participants in the ASD group seemed to feel more engaged with the task than the neurotypical group. Additionally, we observe higher variability in the ASD group when evaluating target predictability. In contrast, the TD group evaluated the target's predictability similarly in all three blocks. Regarding collaboration, both groups reported the "Middle" agent as the most collaborative one. Finally, we could observe differences in the perceived agent predictability, where the "Narrower" agent was perceived as less predictable by the ASD group than the TD one.

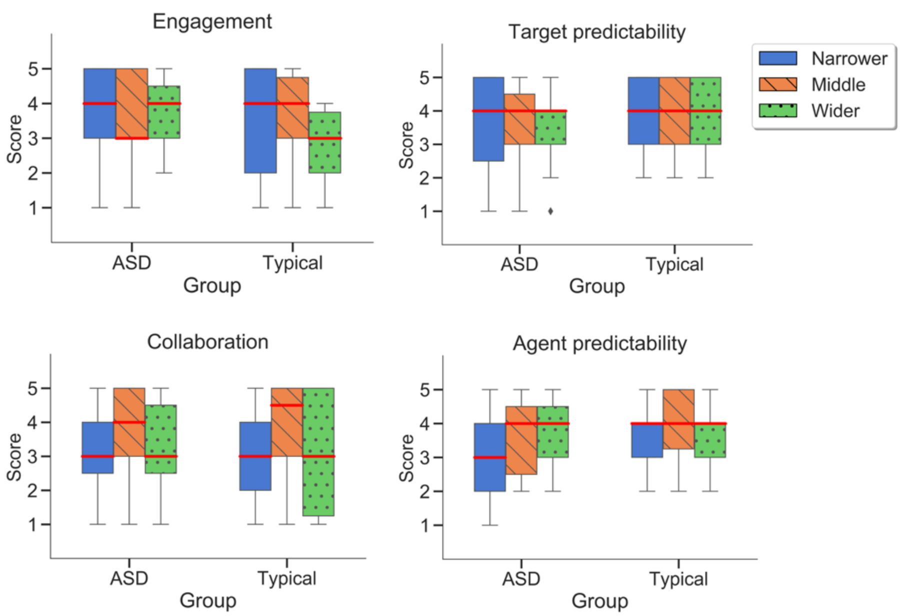


Figure 55. *Differences between agents and groups in perceived agent predictability, target predictability, engagement, and collaboration. Blue represents the results for the "Narrower" agent; orange, for the "Middle" agent; and green, from the "Wider" one. Red lines represent means.*

Finally, at the end of “Phase 3”, participants reported if they were interacting with a human or a computer. There were no significant differences between the ASD and TD groups as to how many participants thought they were playing with a human or a computer. Interestingly, if we divide participants into two new groups (those that thought the other agent was a human and those who thought it was a computer), we observe differences in perceived collaboration (***Figure 56***). Participants that thought the other agent was a human perceived it as significantly more collaborative (Human: median 4.0, MAD: 0.0; Synthetic: median: 3.0, MAD: 1.0; Mann-Whitney U: 37.0, $p = 0.01$). More specifically, when comparing among agents (by assessing the answers during Phase 2, where participants provided self-reports for each of the agents separately), the agent that was perceived as more collaborative was the “Middle” (Human: median 5.0, MAD: 0; Computer: median: 3.0, MAD: 2.0; Mann-Whitney U: 33.0, $p = 0.009$); followed by the “Wider” (Human: median 4.0, MAD: 1.0; Computer: median: 3.0, MAD: 1.0; Mann-Whitney U: 43.0, $p = 0.04$) and the “Narrower” (Human: median 4.0, MAD: 1.0; Computer: median: 3.0, MAD: 2.0; Mann-Whitney U: 43.5, $p = 0.04$).

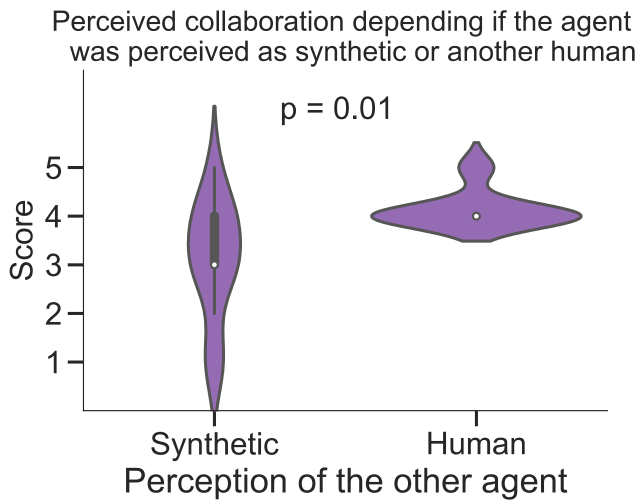


Figure 56. Differences in perceived collaboration between the participants that perceived the AI agent as synthetic or as a human player. The white dot represents the median.

e) Participants in the ASD group focused on movement to describe the other agent

After the last questionnaire in “Phase 3”, participants underwent a short-structured interview, which lasted approximately 10 minutes. First, participants were asked to report the difficulty of the task and how well they performed. We later asked them to comment and describe the other agent they interacted with. In the case where participants reported differences between the agents, we also asked them to provide a short description for each agent.

When participants described the agent(s) they interacted with, we identified the following common features: movement (how fast/slow the agent was perceived), field use (how much of the field the agent was using), colour (the colour of the agent), and collaboration (how collaborative the agent was perceived). ***Figure 57*** depicts the frequency of use of these characteristics to differentiate between the agents (sometimes, more than one per subject). In the ASD group, the most commented characteristic was the agents’ movements (53%), followed by their field use (33%), perceived collaboration (13%), and colour (6%). In the TD group, the frequency of use of the characteristics is more homogeneous. Here the most frequent characteristics are collaboration and field use (38%) followed by colour and movement (30%). Moreover, one participant in the TD group, differentiated between the agents by their perceived performance.

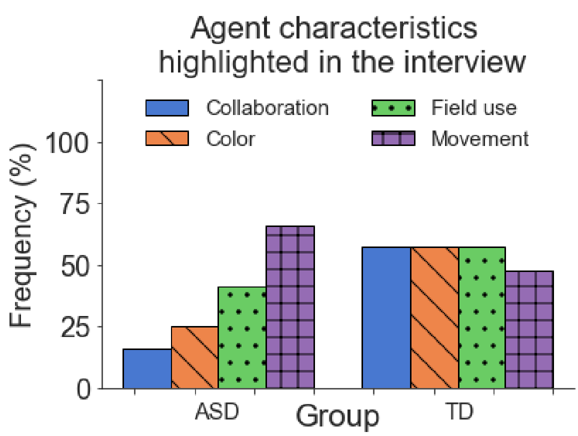


Figure 57. Frequency of characteristics commented about during the interviews. Blue represents collaboration-related characteristics; purple, colour-related ones; green

represents characteristics related to field use; and orange, characteristics related to movement. The sum of the frequencies inside each group overpasses 100% because some subjects highlighted more than one characteristic.

The two groups mainly differed in the type of characteristics they used to describe the other agent. Participants in the ASD used more personality-related terms to describe the behaviour of the other agent ("it's Narrower", "it's more selfish" …) than neurotypical participants, who used a more performance-related vocabulary ("it was playing well", "it was taking my targets"…). When asked about "Phase 3", in which they played with a random AI agent in each trial, participants in the ASD group communicated an added difficulty caused by larger uncertainty. Some reported not knowing if the other agent would go or not for the target; others reported that the task required more focus ("You never know what can happen or how will the other player react. You had to be more focused"). Only one subject reported a relationship between the agent's colour and its behaviour and using it to decide to go or not for the target. In contrast, in the TD group, more participants reported using colour to identify the agent and act accordingly. For both groups, "Phase 3" was perceived as more complicated and confusing than the others.

8.5 Discussion

The main purpose of this study was to evaluate the ability to predict another agent's behaviour based on sensorimotor interaction and how these predictive abilities affect collaborative interaction and differ between ASD and Typically Developed (TD) individuals. We created a task where participants had to learn the behavioural characteristics (as exhibited by sensorimotor contingencies) of a synthetic agent and collaborate with the agent to maximize reward. Each player controlled an avatar, and the goal was to intercept falling targets. To assess collaboration, we developed the Point of Social Subjective Equality (PSSE) that calculates the probability of a player of going for the target given the target's position. Finally, we examined possible perceptual differences regarding the task between the two groups.

a) Discussion on differences in behaviour

As we observe larger individual differences between participants in the autism spectrum (compared to neurotypicals), we hypothesized that participants in the autism spectrum would show more variable behaviour than neurotypicals during the task. Our analysis of the differences in variability between the ASD and TD groups suggests that, indeed, the ASD group showed larger variability compared to the TD individuals.

Social impairments associated with sensorimotor difficulties are a characteristic of the disorder, and we assumed that ASD individuals would encounter difficulties in predicting the AI player's behaviour. Thus, we hypothesized that participants in the autism spectrum would show slower and less adaptation to the other agent than neurotypical ones. To assess this, we analysed the differences between groups in adaptation to the other agent during Phase 2, and we showed differences in adaptation between groups in early trials but not in late ones, showing differences in adaptation timing. Our results show differences in the behaviour of neurotypical and ASD individuals when playing with the three different synthetic agents. We observe the ability to converge to a complementary PSSE in the case of the control group. However, we do not observe the same with the ASD participants. Furthermore, we assessed the online adaptation to the artificial player by looking at the differences in errors between early and late trials among groups. Our results seem to reflect a more accurate adaptation in the neurotypical group than in the ASD.

Finally, as ASD individuals seem to find difficulties when a task is uncertain, we postulated that they would fail to predict the behaviour of the AI agent correctly and, therefore, adapt less to the AI agent compared to the typically developed group. By comparing the participants' behaviour with each agent in "Phase 2" and "Phase 3", we can assess whether they applied previously acquired information from the sensorimotor contingencies of the AI agent ("Phase 2") to a more uncertain task ("Phase 3") and predict the agent's behaviour. Our results suggest differences in the prediction of the agent's behaviour. More specifically, TD individuals were able to develop a better model of the artificial player in "Phase 2" and apply that information to adapt their behaviour in "Phase 3", while participants in the ASD failed to do so.

These results suggest that neurotypical individuals can adapt their behaviour according to the AI player and converge to an optimal game strategy by observing the sensorimotor patterns of their partner. In contrast, ASD patients seem to lack this ability, suggesting an impairment of socSMCs, possibly due to their lower predictive skills (Schmitz, et al, 2003; Sinha et al., 2014; Wang, Kloth, & Badura, 2014).

b) Discussion on differences in perception

To understand possible perceptual differences between the two groups, we looked at the questionnaires provided to the participants after the completion of each block, and the short-structured interview at the end of the task. Participants at the end of each block reported how they evaluated the task, the target, and the other player in terms of engagement, predictability, and collaboration. Although we did not find any statistical significance in any of the items, participants in the ASD group seemed to perceive the task as overall more engaging than the TD group. Participants in the TD group perceived the task as less engaging when interacting with the "Wider" agent. When evaluating the agent's predictability, the "Narrower" agent was perceived as less predictable by the ASD group compared to the TD. Playing with the different agents did not seem to affect target predictability in the TD group. However, we observe higher variability in the reported target predictability in the ASD group when playing with the "Narrower" and "Middle" agents. In terms of collaboration, both groups rated the "Middle" agent as more collaborative than the "Narrower" and "Middle". Despite a lack of significance, our results provide possible insights on perceptual differences regarding the tasks' characteristics with respect to the agent's behaviour. However, more data needs to be collected.

At the end of the task, we asked participants to report whether they thought they interacted with another human or a computer. We found no significant differences between the two groups. The agent was perceived as significantly more collaborative by participants that thought they were playing with a human instead of a computer. More specifically, the “Middle” agent was rated 50% more collaborative when participants thought it was another human. Indeed, according to Turing’s test (Turing, 1950), the behavior of a machine can be confused with that of a human.

Finally, the short-structured interview allowed us to assess further the perceived differences of the agents between the ASD and TD groups. The main differences arise from the characteristics used to describe the agents. Participants in the ASD group mainly commented on the agents’ movements, followed by their field use. The agents’ colour was the characteristic less commented about. In contrast, the TD group differentiated between the agents by almost equally exploiting all three characteristics. The focus on movement as the main differentiating characteristic is something to be expected from the ASD group, as individuals in the spectrum tend to focus on moving objects. Moreover, the fact that almost no subject in the ASD group commented on the agents’ colour as a significant trait could support the idea of the lack of model generation. If the agent’s colour was a characteristic that could help participants predict its behaviour, it would be unnecessary to consider it if no model was being created.

c) General discussion

The contributions of this study are two-fold. On the one hand, we formulated and introduced the Point of Social Subjective Equality (PSSE), a concept that allowed us to model the behaviour of both humans and artificial agents in a collaborative task. By observing the PSSE, we quantified the degree of behavioural adaptation and how it can be modulated based on the variation of sensorimotor contingencies of the synthetic agent. On the other hand, this study demonstrated how collaborative behaviour could implicitly emerge and be modulated through the observation of sensorimotor patterns of the partner.

Our behavioral analysis showed lower and slower adaptation to the artificial player by the ASD group. Similar results were found in (Lieder et al., 2019), where participants in the autistic spectrum showed lower and slower adaptability in the task than their neurotypical counterparts. However, previous studies examining sensorimotor planning in individuals with ASD have yielded conflicting results. Some studies indicate an impairment in prospective control in ASD (Hughes, 1996; Scharoun & Bryden, 2016). In contrast, other studies showed no significant differences (Hamilton, Brindley, & Frith, 2007; van Swieten et al., 2010).

The larger variability in behavioural results of the ASD group is also present in the self-reported data. Nevertheless, the perceived predictability and collaboration during the task showed no differences between groups in these measures. Interestingly, the differences in the behavioural data but not in the self-reports raise the question of self-awareness. Could that be due to a lack of metacognition or due to a coping mechanism? Unfortunately, our current data do not allow us to answer this question, and further studies need to be conducted.

9. IMPLICATIONS FOR A PEDAGOGICAL MODEL

9.1 Introduction

This chapter aims to propose a pedagogical model of learning based on metacognitive abilities. It is based on the original DAC pedagogical model (with the *resistance, confusion* and *abduction* phases), the metacognition models described in the Introduction chapter, and the results and observations of the studies presented in this thesis. It proposes a similar idea to Piaget: When learners face information that fit their schemes, they assimilate it. Nevertheless, when they meet information that crashes with their schemes, they enter a state of disequilibrium and need to change their model to accommodate this new information. DAC's pedagogical model begins from this state of disequilibrium.

As mentioned before (see Introduction for more detailed information about it), the original DAC model comprised three phases or stages: *resistance, confusion* and *abduction.* In the first one, the learner "resists" this new information that crashes with their mental schemes (which, behaviourally, could be presented in high confidence errors). This phase would be similar to Buch's unconsciously incompetence stage. Then, the learner enters into a state of *confusion,* where they begin to realise that their model was wrong but do not understand the new information yet. In this case, this stage is similar to Buch's *consciously incompetence* one. They know they do not know but have not gained the knowledge yet. Finally, students enter in a phase of *abduction,* in similar lines to the eureka effect (similar to a sudden, seemingly unaccountable moment of inspiration, (Perkins, 2001)), where they begin to understand the new knowledge, and they are aware of it. This phase would be similar to Buch's *consciously competence* stage.

When considering the initial *resistance* phase proposed by DAC's pedagogical model, a question arises: What happens with those learners that, despite performing well, report low confidence and do not improve? Can there be a reversed *resistance* phase, one where the learner, instead of denying the new knowledge, denies their ability? In this case, students would also have difficulties to gain new understanding, as they would be “resisting” their ability to acquire it or understand it.

Thus, we propose two resistance phases: one for knowledge and one for confidence. The first one, which we will call *content-resistance*, comes from the original DAC model, where learners show low performance but high confidence in their answers. The second one, which we will call *self-resistance*, adds-up to the original DAC model. It refers to the learners that, despite showing a larger performance, report low confidence in their answers. We also must consider the existing relation between low confidence and surface learning, that is, working focusing only on reaching the minimum requirements to pass the course.

9.2 Existing theories

This section will summarise the existing theoretical models, presented in chronological order, more relevant for DAC's phases of learning. For a more comprehensive revision on models of SRL, check (Panadero, 2017).

a) Hegel's Dialectic model

Piaget's theory on cognitive disequilibrium could be compared with Hegel's dialectic model (***Figure 58***) where a new idea (called *synthesis*) comes from the conflict between a previous knowledge (called a *thesis*) and the *antithesis* (that is, new knowledge that contradicts the previous one). In similar lines, DAC's pedagogical model also sees generating new knowledge from contradictory ideas. It sees learning as the gain of new knowledge when the previous mental model (either of the world, *content-resistance* phase, or of one's self, *self-resistance* stage) crash with the current situation.

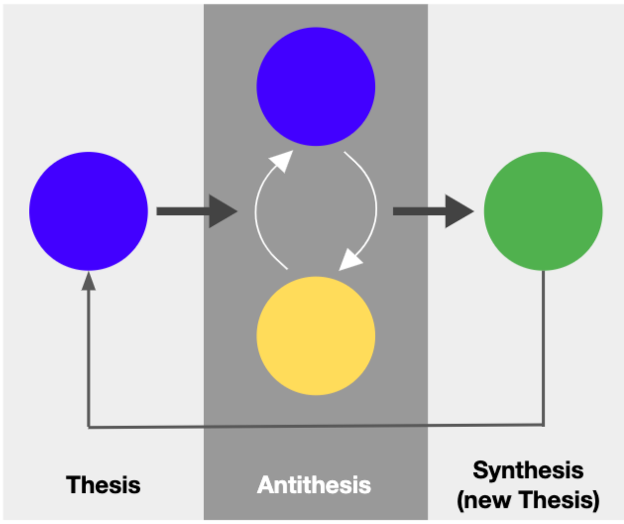


Figure 58. Hegel's Dialectic model

b) Peirce's theory of Abductive Reasoning

Considering Peirce theory (Peirce, 1997), every agent has his idea of the world (what he defines as *ego*), in contrast with the real way the world is (defined as *non-ego*). When the agent encounters a situation that does not fit his inner world model (*ego*), he would be surprised. That surprise sets off a struggle. If it were the non-ego what collapsed in the struggle, the agent would attribute the experience to Perception, what would mean he correctly understood the world. Nevertheless, if It were the ego what collapsed, the agent would attribute the experience to Imagination, what would mean he did not understand the world correctly but was only imagining it.

As stated by (Kirlik & Storkerson, 2010) review of Peirce's semiotics, abduction can be considered "a phenomenological sense of knowing or logically determining that something is or might be the case". He saw it as part of discovery, the stage of scientific inquiry where one's generates theories to assess them later.

c) Piaget's theory on Cognitive Disequilibrium

As previously mentioned in Chapter 2, Constructivism sees children as active agents in their learning, constructing their knowledge through interacting with the world surrounding them (Piaget, 1977). The term "cognitive schemes" was coined by Jean Piaget to represent the mental maps children create through recollecting information and reorganising it in their brain, creating meaning out of their own experiences (Piaget, 1952). These dynamic schemes grow and change over time through experience.

Piaget's theory presents three steps: *equilibrium, assimilation,* and *accommodation*. If children's existing schemas can explain the world surrounding them, they are in what he called a state of *equilibrium* (that is, a state of cognitive balance). When they face new information, this one can match or not their mental schemes. The second term, *assimilation*, represents dealing with a new idea by using existing schemes to fit it in one's mental models. However, if the information does not fit their schemes, children enter a state of *disequilibrium.* When this happens, learners go through a process of *accommodation* by updating their mental schemes to include the new information. ***Figure 59*** represents this process.

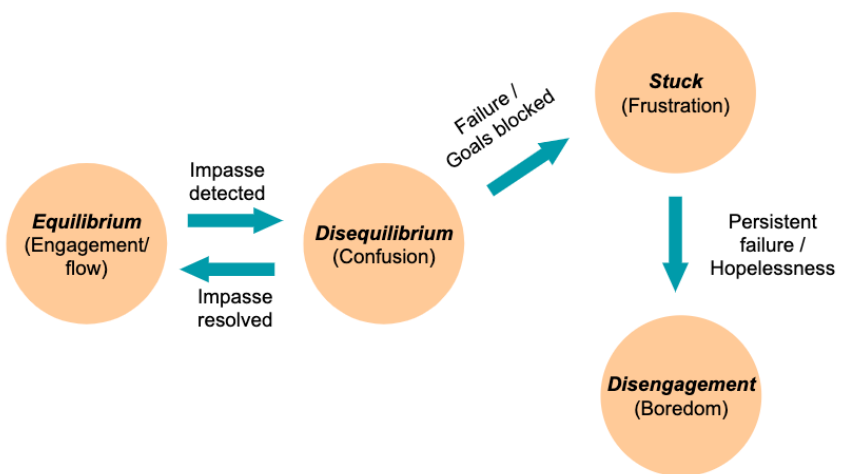


Figure 59. Piaget's Cognitive Disequilibrium model, adapted from (D'Mello & Graesser, 2010).

d) Buch's Conscious Competence Theory

Noel Burch, an employee with Gordon Training International, developed the Four stages of competence in the 1970s. This model can be known for several names: Four stages of competence / Conscious Competence Ladder / Conscious Competence Matrix, the Learning Matrix, and the Four Stages of Learning (Noel Burch, 1970). The model proposes two factors affecting how we learn a new skill: consciousness (as awareness) and skill level (competence).

1. *Unconscious incompetence* (*Ignorance*) The learner does not know how to do something, and they are not aware of their lack of knowledge. They may even be unaware or deny the usefulness of the skill. To move to the next stage, learners need to recognise their lack of knowledge and its value.

2. *Conscious incompetence* (*Awareness*) The learner still does not know how to do something but recognises they do not know and the value of the skill. To move to the next stage, the learner needs to practice that skill until arriving at the required competency level.

3. *Conscious competence* (*Learning*) The learner knows how to do something. Nevertheless, they still need to put effort to do it (reflected in taking more time, for example). To move to the next stage, the learner needs to practice that skill still, until it becomes "automatic", not requiring that much effort.

4. *Unconscious competence* (*Mastery*) The learner has had so much practice with a skill that it has become "automatic" and can be carried out easily (even, sometimes, while performing another task). Moreover, in this stage, the learner would be able to teach it to others.

e) Zimmerman's Self-Regulated Learning models

Zimmerman's work can be divided into three models: The Triadic Analysis of SRL, the Multi-level model and the Cyclical Phases of SRL. The first one, similar to Bandura's triadic model of social-cognition, represents the interactions of three forms of SRL (environment, behaviour and person level). The second model presents four stages in which learners acquire their self-regulatory competency (Zimmerman, 2000). These are observation (the vicarious induction of skill from a proficient model), emulation (the imitative performance of the general pattern of the style of a model's skill with social assistance), self-control (the independent display of the model's skill under structured conditions), and self-regulation (the adaptive use of skill across changing personal and environmental conditions).

The third model, the Cyclical Phases of SRL, presents the interrelation of metacognitive and motivational processes (Zimmerman, 2000). In the first phase, the *Forethought* phase, students analyse the task, set goals, plan how to reach them. In the second, the *Performance* phase, learners perform the task, while monitoring their progress, using self-control strategies to keep themselves cognitively engaged and motivated to finish. Finally, in the *Self-reflection* phase, learners assess how they have performed the task, making attributions about their success or failure.

f) Boekaerts's Dual Processing Model

Boekaerts' work can be divided into four models: The Structural Model, the adaptable Learning Model and its extended version, and the Dual Processing model. In this case, we will not enter in the first one, as it is too general for this issue. The second one, the Adaptable Learning model, which later evolved to the Dual Processing one, presents two parallel models: The Coping or Well-being mode and Mastery or Learning mode. These models, in similar lines to the dichotomy between performance- or mastery-oriented learners from the Achievement Goal Framework (Elliot & McGregor, 2001) (seen in Chapter 5, section 5.2.c) depending on the origin of learners' strategies. If they are in the Coping/Well-being mode, they will pursue their task goals by avoiding negative feedback, similar to the Performance-Avoidance goal in AGF. On the contrary, if their strategies intend to master the task, they would be in the Mastery/Learning mode, similar to the Mastery-oriented goal in AGF.

The last model, Boekaerts' Dual Processing Model (Cascallar, Boekaerts, & Costigan, 2006), proposes three different purposes for self-regulation. In the first one, following a "top-down" approach, students' strategies to pursue their learning goals come from their values, needs and personal goals, following a mastery/growth pathway. In the second, following a "bottom-up" approach, those strategies come from preventing the self from being damage. Here, there can be a mismatch between the learning goals of the task and the students' ones. The final one represents students trying to move from this "bottom-up" / well-being approach to the "top-down" / mastery one. Students can pursue this change through internal forces or via external help (either from teachers or peers), as in Vygotsky's ZPD theory.

Köhler's *Insight theory* presented several steps in a learning process: failure, pause for thought, perception, insight, success. According to Köhler, the first step is *failure*, where the learner does not get the right result. Consequentially, learners *pause for thought* (to reflect on why the result was not correct) and think about possible approaches to achieve the correct result through *perception*. Finally, they then have an *insight*, a flash of inspiration, that leads them to *success*. These steps are related to a proposal for learning phases, explained in the following section.

9.3 Phases of Learning

A preliminary version of this model appeared in (Vouloutsi et al., 2015). This thesis extends the model in two ways. The first one, to account for the difference between two types of resistance (content and self). The second, to add a new phase to represent the use of the obtained knowledge. ***Figure 60*** depicts the stages.

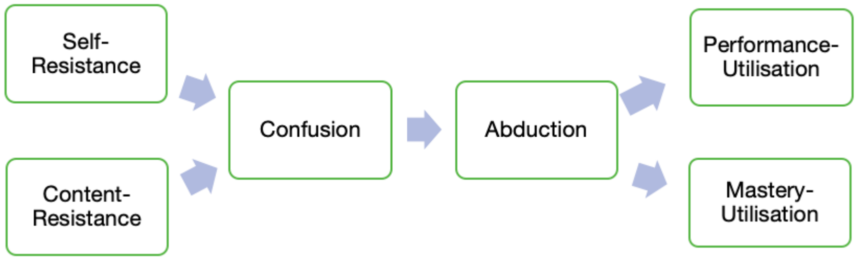


Figure 60. DAC's phases of learning

a) Resistance: Content and Self

Peirce holds that inquiry begins when experience contradicts a belief. Surprises generate doubts and demands for explanations. We propose that the learner, when faced with new information (or feedback about their performance) that crashes with their current knowledge, enters in a state of resistance. The directionality of this state depends on the learner's bias. If they are highly confident on an incorrect theory, students find themselves in the *resistance* phase (from now on, *content-resistance* face), as they are not aware that they do not know. This face would be similar to the unconsciously incompetent stage from. Contrarily, if they have low confidence in their correctness, students find themselves in the *self-resistance* phase, as they are not aware that they know.

The self-resistance / lack of agency phase accounts for a stage in which the learner, although having enough knowledge to solve the task, does not feel confident in his abilities. Despite not being as troublesome as the resistance phase in short-term, a learner that does not trust his abilities will not explore new strategies, be highly affected by feedback and has more possibilities of giving up that a student with proper metacognition. This is because low confidence correlates to a performance-oriented approach where the learner only wants to solve the task but does not care about learning from it, and to a fixed mindset, making him feel that his knowledge is something static that cannot improve.

Both *self-resistance* and *content-resistance* relate to students' abilities to monitor their errors accurately. Self-resistance could be affected by self-handicapping behaviour. Self-handicapping behaviour consists of claiming (or creating) an obstacle before performing a task (Berglas & Jones, 1978; Jones & Berglas, 1978). Like this, if one fails the task, one can present the self-handicap as the reason behind that failure, protecting the perceived ability in the specific domain and global self-esteem (Feick & Rhodewalt, 1997; Mccrea & Hirt, 2001; Rhodewalt, Morf, Hazlett, & Fairfield, 1991). This kind of behaviour has been shown to correlate with performance-avoidance goals, that is, when students focus on passing the test/course (instead of mastering it and exploring) and avoiding negative feedback (Midgley & Urdan, 2001). It should not be confused with defensive pessimism, that is, a purposely decrease of effort or procrastination to attribute the poor outcome to lack of effort instead of lack of ability (Garcia & Pintrich, 1994; Midgley, Arunkumar, & Urdan, 1996).

b) Confusion

During the confusion phase, the learner begins to realise that their mindset does not reflect reality (either by being highly confident on a wrong theory, as in the *content-resistance* phase; or by being unsure about their view, as in *self-resistance*). Consequentially, the discrepancy between actual and perceived performance will decrease (here, we consider this discrepancy as an absolute, thus non-signed, version of the *bias* measure). Inquirers respond by formulating hypotheses to account for unexpected phenomena.

As previously stated by researchers like D'Mello, confusion can be beneficial for learning (D'Mello, Lehman, Pekrun, & Graesser, 2014). Nevertheless, we always have to make sure to maintain the learner in a limited range of confusion, as we do not want to provoke frustration or to make them give up the task. Thus, confusion must be externally regulated, adapted to the learners' characteristics. One measure to do so would be considering students' intolerance for uncertainty, as that would allow us to maintain the learner in a tolerable range of confusion.

c) Abduction

After facing the new information that crashed with their knowledge and going through a confusion phase, the learner gains new knowledge through abduction. DACs pedagogical model proposes an equivalent paradigm, where this difference between the inner world of the learner derived from his previous knowledge (similar to Peirce's *ego*) and the world itself, which presents him new knowledge (Peirce's *non-ego*) causes *confusion*, what the learner has to overcome to abduct and acquire knowledge.

Abductive reasoning does not provide the same certainty (as deductive one does, for example), so a larger tolerance for uncertainty might help to go through it without giving up. In abduction, the learner creates frameworks for interpreting and analysing phenomena, allowing to go from mere intuition to systematised knowledge. Like this, the learner generates predictions that can be tested by further observations.

d) Utilisation: Performance and Mastery

After realising the strategy, the learner can take the last step of the learning phases: using that strategy. In the utilisation phase, the learner has not only proved to be able to solve the task, like in the abduction phase but can make use of this learning to solve a task. Therefore, this phase needs to be assessed in activities where learners can answer freely without having to provide a constrained answer.

The *utilisation* phase could be split between *performance-oriented utilisation* and *mastery-oriented utilisation*. In the former, the learner would be using that strategy in a performance-oriented way, that is, prioritising the score. They would take the strategy with a higher chance of scoring (either by being easier or due to previous experience or higher performance/confidence). An exploitation manner would dictate this behaviour. The learner would be exploiting previously rewarded strategies. In the latter, the learner would be focusing on mastering the task in an *exploitation* manner. To do so, they would utilise strategies that they know but not only the ones with a higher chance of scoring. It must be highlighted, though, that this behaviour would not be completely exploratory, as the learner would be using previously explored strategies.

9.4 Behavioural outcomes

This section is divided into two subsections. The first one will present what would be the expected behavioural outcomes of the previously presented phases. The second one will present what examples of those behaviours we could find in the studies presented in this thesis.

a) Expected behavioural outcomes

There are some behaviours we could expect from students, given the phases mentioned above. ***Table 11*** presents the possible behaviours/actions that would make a student remain or move in a specific learning phase. These ideas are based on the results and results of the previously presented studies and literature.

One would expect that, if learners are in the *content-resistance* phase, they will remain there either if they ignore feedback (due to not trusting the system, for example) or if they manage to get correct answers with wrong strategies. In the last option, learners would not realise they do not know because of their (incorrect) strategy worked. To move to the next phase (*confusion*), they need to realise they are committing mistakes. On the other side, learners in a *self-resistance* phase would have difficulties to pursue more complex strategies and move to the following stage due to not being aware of their abilities. In this line, they would remain in the *self-resistance* phase if they do not realise that they are performing correctly (by showing low confidence in their responses). To move to the next phase, they need to recognise the correctness of their answers and increase their confidence.

The way to detect that learners are in the *confusion* phase would depend on the nature of the task. Suppose students' actions in the task consist of providing the expected outcome and their confidence about it. In that case, confusion could be detected by a confidence decrease after high-confidence errors or a confidence increase after low-confidence corrects. If the task allows for a free answer, where they chose what process would lead to the desired outcome, it would be detected by their exploration patterns: Their decision to explore previous high-confidence errors or low-confidence corrects. Doing so, they would be testing their previously incorrect (either due to self- or content-resistance) theories.

Students would remain in the *confusion* phase if they were presented with many different setups (following different rules, for example), as the feedback needed to understand the rule would not be specific enough. This is why adaptivity is essential to scaffold students' learning process. On the contrary, learners could move to the next phase if they received specific exercises (with feedback), adapted to their learning needs, to increase both performance and confidence at the same time.

It becomes crucial, thus, to diminish the "bias" (the difference between perceived and real accuracy) to perform abduction. If the difference is too high, this can mean that the confidence is too low or too high. In any case, the learner would not consider new information as valuable: If they have overconfidence (positive bias), the learner could think their model of the world is enough, no need to go further and abduct; if they are underconfident (negative bias), they could resist to new knowledge, because it crashes with their world.

Finally, suppose learners were in the *abduction* phase. In that case, that is, answering correctly (and with medium-high confidence), the next step would be for them to move towards the last phase, *utilisation*, where they would not only answer correctly but also actively use those strategies in following tasks. Thus, the expected behaviour that would reflect they were remaining in this phase would be not using those strategies that they answered correctly when asked to respond freely. This could be due to a fear of uncertainty, what would make them not explore strategies to avoid failing and also larger error weight, as they could be more affected by the negative effect of failures than the positive impact of correctness, in line with a *Performance-avoidance* goal directness. On the contrary, if, when asked to use a strategy to answer freely, learners used the newly obtained strategies, they would be moving towards the *utilisation* phase. More specifically, if learners only (or mainly) used the strategies that previously gave them a larger score, they would be in the *performance-oriented utilisation* phase. In contrast, if they explored new strategies, despite not always obtaining larger results, they would move to the *mastery-oriented utilisation* phase.

Table 11. Learners' expected behaviour leading to remaining or moving in each learning phase

	Remain	**Move**
Content-resistance	- Getting correct answers with wrong strategies - Ignoring feedback*	- Realising you're committing mistakes (by seeing the balance fall)
Self-resistance	- Not increasing confidence	- Seeing that your answers were correct
Confusion	- Getting feedback for many different strategies (too many reasonings to correct)	- Getting feedback for different setups with the same rule/item type
Abduction	- Not using the correct strategies	- Using the correct strategies

b) Examples in this thesis

The first study presented in this thesis (Chapter 3, *Students' difficulties in recognising their competence level: A revision of the Unskilled and Unaware effect*) assessed the discrepancy between actual and perceived accuracy in a logic task. Despite not having feedback, what would have made learners move through DAC's phases, it showed a clear distinction between the two proposed *resistance* phases. Like this, participants in the bottom quartile (low performers) who reported high confidence in their answers would be in the *content-resistance* phase, as they were not aware of their lack of knowledge. On the contrary, participants in the top quartile (high performers) who reported low confidence in their answers would be in the *self-resistance* phase, as they were not aware of their actual knowledge. ***Figure 61***, coming from the results of the study, depicts the mapping between the *resistance* phases and participants' behavioural patterns in the task.

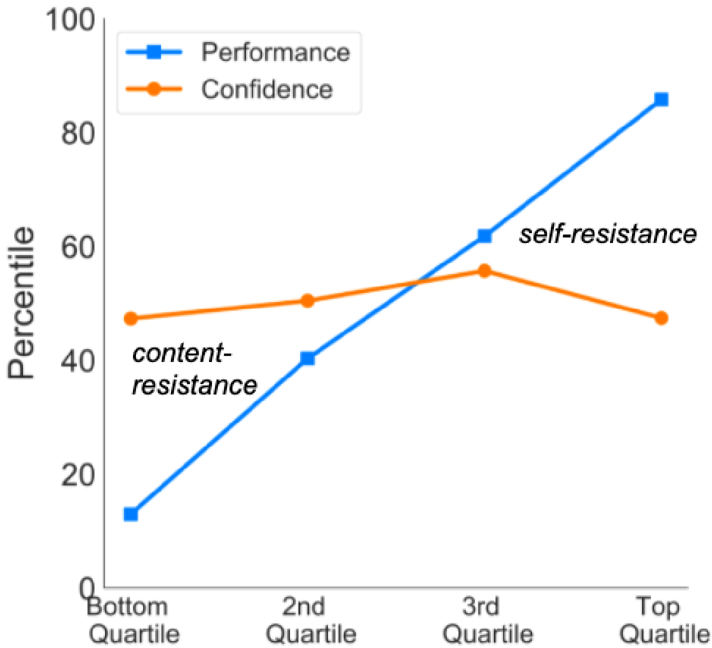


Figure 61. Perceived logical reasoning ability as a function of actual test performance. Results in blue represent actual performance, while orange results represent perceived one (confidence). Participants in the bottom-quartile showing larger confidence could be considered as being in the *content-resistance* phase; while participants in the top-quartile showing lower confidence could be considered as being in the *self-resistance* phase

The use of feedback in the second study (Chapter 4, *How confident are you? A study on gender and feedback*) allows us to analyse the participants' journey through more phases than the previous one. Like this, in similar lines to the previous study, participants can be divided between being in the *content-resistance* phase (those who showed high confidence and low performance) or in the *self-resistance* stage (showing low confidence and high performance). ***Figure 62*** depicts the evolution of these two groups during the task. During the task, participants received feedback about the correctness of their answers, what moved the bias measure (the difference between perceived and actual accuracy) towards 0. This could represent subjects moving towards the *confusion* phase, as those in the overconfident group began to report lower confidence in their wrong answers and those in the underconfident group began to report higher confidence in their correct answers. Finally, in the post-test, participants' bias moved even closer to zero (even without feedback), showing a more accurate perception of their abilities, what could be in line with the *abduction* phase.

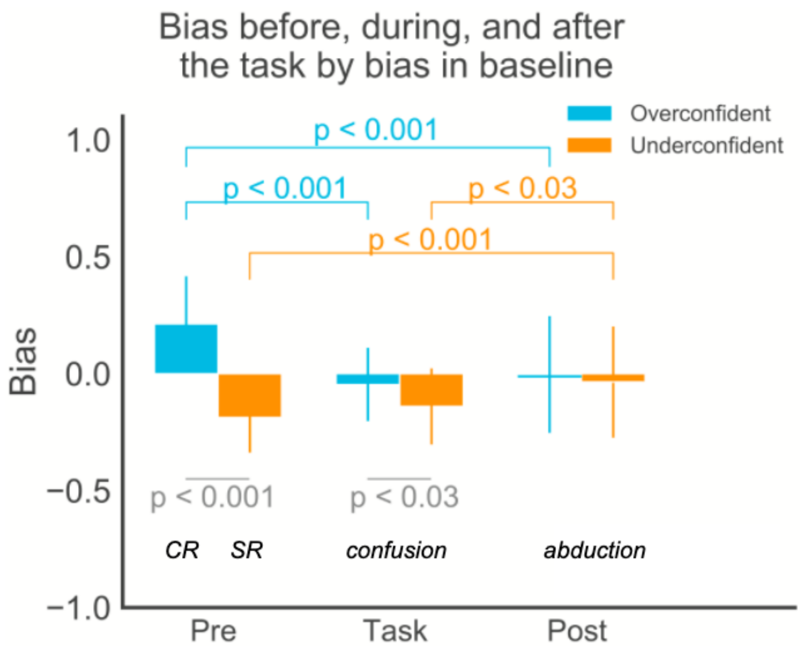


Figure 62. DAC's phases of learning mapped to the evolution of the bias measurement for the whole experiment, when participants were classified as overconfident and underconfident. The results illustrate how the differences in bias in the pre-questionnaire disappear in the post-questionnaire. These results match the DAC's phases of learning predictions regarding the confidence of the learner.CR accounts for *content-resistance* and SR, for *self-resistance.*

The third study (Chapter 5, *Metacognitive factors behind rule change in the Balance Scale task*) was designed with the proposed last phase of learning (*utilisation*) in mind. Like this, in similar lines to the previous study, the pre-test allowed us to classify participants either in the self- or content-resistance phase. Later on, receiving feedback about the correctness of their answers would move them towards the *confusion* phase, as they would have to deal with high-confidence errors receiving negative feedback and low-confidence corrects receiving positive feedback.

The exploratory blocks of this study permit a more in-depth insight into the utilisation phase. Considering this, participants who mainly used strategies (in this case, represented by the type of setup, also known as item type) that had previously received positive feedback would be classified as being in the *performance-utilisation* phase, as their strategies would follow an exploitation approach. On the other side, participants that mainly used strategies not seen previously (or testing ones that scored as wrong before) would be in the *mastery-utilisation* phase, as their strategies would follow an exploration approach.

In this case, the order of the *abduction* and *utilisation* phases would not be as linear as in the proposed model. This is because it is not until the post-test where they provide answers without receiving feedback about it (as happened in the previous study). We used the study in Chapter 5 to analyse the *utilisation* phases, by finding the patterns related to either *performance*- or *mastery-utilisation*.

Considering the study in Chapter 5, we can classify students given the testing (or not) of previous item types. Like this, ***Table 12*** presents the following types, depending if they test or not previous items (and if these ones were previously correct or not).

Table 12. Learners' classification depending on their testing of previous items.

	Not testing previous wrongs	**Testing previous wrongs**	**Other**
Not testing previous corrects	Avoidant Exploratory	Informative Exploratory	Just exploratory, no specific strategy for wrong items
Testing previous corrects	Avoidant Self-enhancement	Informative Self-enhancement	Just Self-enhancement, no specific strategy for wrong items
Other	Just Avoidant, no specific strategy for correct items	Just Informative, no specific strategy for correct items	No specific strategy, neither for correct nor wrong items

The main profile in the study in Chapter 5 is the "self-only", that is, the one where learners showed higher number of previously correct items during the exploration phases. This kind of learner did not take a specific strategy for wrong items (neither avoiding nor focusing on them). This pattern is in line with an exploitation-related strategy, in similar lines to the *Performance-utilisation* phase. A further extent of this strategy, not only testing correct items but also avoiding wrong ones, is represented in the "Avo self" profile, which decreases from the first to the second exploration phase. On the contrary, learners in the "Exp only" profile, seemed to focus on not testing correct items during the exploration phases. This approach would represent the firsts steps toward the *Mastery-utilisation* phase, as learners are aware of what they know and do not waste trials on testing it but on testing other hypotheses.

At the moment, the fourth study (Chapter 6, *Metacognitive factors behind rule change in the*) does not allow for an exact mapping between DAC's phases of learning and visitors' behaviour during the experience. To do so, we would need, for example, to assess their previous knowledge (both actual and perceived) of the future content. This would allow us to see how their model made them navigate the range of information provided by the system.

Moreover, the application does not (yet) offer a sequential path based on item complexity. Like this, one approach would be to classify the items based on their levels of comprehension. That would mean knowing which items require having visited previous ones (and which ones are they) to understand their content correctly. That could create an initial baseline level that all visitors would need to go through (to gain initial knowledge before exploring more complex items). It could also imply the possibility of having different (even parallel) levels so that visitors could "extrapolate" what they learnt in one to another (in similar lines to the *utilisation* phase).

The first one would be the metacognitive regulation of the *planning*, *monitoring* and *information processing* derived from the nature of the task. Their behaviour could also go around the metacognitive experience about their *feeling of knowing*, as they can freely explore the information, it would be essential to assess what they think they know or do not know, as it could be driving their behaviour. Finally, we could also consider their metacognitive knowledge about the understanding of false beliefs, as visitors could face moments where they need to be aware of their thoughts about the topic.

It would be important, thus, to understand how to evaluate that comprehension. One option would be fact retrieval, as assessed by (Pacheco, Sánchez-Fibla, Duff, & Verschure, 2017). Nevertheless, in this option, although easier to quantify than others, it would be challenging to divide understanding from memory. Another option, then, would be to ask participants to present several items/situations and ask them to relate cause and effect. Two options arise here: Either users are provided with a narrative (for example, about a character) containing facts and they have to report what they think that will happen given those facts, or they can receive a fact (or a decision from that character) and they would have to report which of the previously explored fact(s) could have led to that situation. The main concern of this option would be the need to assess participants' analytical skills, as if they are not able to perform this task, we should be able to distinguish between lack of understanding of the provided content or a lack of analytics skills from their part. The approaches presented here would allow us to assess if a system that guides your content presentation depending on your previous knowledge/understanding could lead to larger learning gains than free exploration or a predefined guided path.

The fifth study presented (Chapter 7, CREA) did not directly measure metacognition (by comparing perceived and actual performance). Nevertheless, the constant contact with feedback (and failure) faced through the process of coding, is thought to promote students' resilience and persistence and change their beliefs about the source of learning (hard work rather than immediate mastery).

Finally, the last study (Chapter 8, *Collaboration variability in Autism Spectrum Disorder*) assessed a different dimension of metacognition and self-regulation: perceived against actual collaboration while playing with another agent. In this case, results showed differences in participants' collaboration with the other player depending on their group (either neurotypical individuals or in the autistic spectrum), while finding no differences against reported perceived collaboration. The lack of adaptivity (or, better said, the longer adaptivity process) of participants in the spectrum could relate to some kind of *content-resistance,* as they would take more time to adapt to the behaviour of the other agent. Moreover, the larger difficulties in adapting to the other player under uncertainty (that is, in the block where each agent appeared in random order), could also relate to the *content-resistance* phase.

The explanation of these results could go in two directions: Were participants with ASD not aware of their affected collaboration or was it a sign of hypermentalizing (that is, excessive mental state attribution) as a way to countermand the difficulties of their condition? Unfortunately, this study does not allow to assess this directly. To do so, we could design a task less related to their lacks due to their neurodiversity, something they may not feel they need to counteract for.

9.5 Limitations and Further Steps

As discussed in the previous section, part of the studies present behaviour supporting the proposed learning phases. At the moment, more development and more in-depth analysis are required to see the reflection of DAC's learning phases is less controlled environments as the studies presented in Chapters 6 and 7.

We also need to consider that *confidence* may not be a perfect measure, as it may be a value computed merging accuracy with noise. This "noise" can come from other distal variables, as learners' self-efficacy or perceived competence in a field. On another note, at the moment, the model seems to require providing feedback to the learners for them to move through the stages. One possibility would be to test how much time students need to go through these phases with and without feedback.

Further steps could go in the line of analysing the outcome of the strategies proposed. This analysis will allow us to assess if feedback adapted to learners' needs (not only about the content of the task but also on metacognitive accuracy) promotes a faster learning process. Another possible study would be to analyse the process and outcome of these phases under different levels of uncertainty.

10. GENERAL DISCUSSION AND CONCLUSION

10.1 Introduction

This chapter contains the general conclusions of the studies presented in this dissertation. This thesis presents ideas around the relationship between students' metacognitive abilities and their performance during a task. In this case, two studies with the Balance Scale task are presented, one more preliminary, focusing on the notion of confidence, and another, larger study, comprising several metacognitive factors. These factors, presented as distal variables, reflect several characteristics of the students that could affect a task. More specifically, that could impact strategy decision, regardless of the actual knowledge of the student. They are intolerance for uncertainty, mathematical anxiety, goal-directness, and perceived competence.

On another note, we also present a study based on the Dunning-Kruger effect. In our case, the task was focused on science literacy and presented to master students in a cognitive science master program. In it, we could see how, in similar lines with the original study, there is a general trend for reporting confidence over the mean, regardless of the performance. More specifically, individuals with a low performance show higher confidence, while the ones with high performance report lower confidence. Moreover, it showed a relationship between differences in females' performance per year and the gender ratio on that year.

Another part of this thesis, Chapter 8, focuses on differences between perceived and actual behaviour in autistic individuals during a collaborative task. In this study, we have seen how participants in the spectrum reported a similar perception of the task and their collaboration with the other partner to their neurotypicals counterparts while their behaviour was different.

Finally, this thesis also presented two studies in real-life scenarios. The first one presents the first steps towards a methodology to teach programming and robotics/electronics to elementary school children. Through Computational Thinking and self-reflection, children undergo six steps from the observation of robots' behaviour towards the construction of small projects. During this time, the teacher takes a constructivist approach and becomes a "guide in the side", giving general instructions of each step and allowing the children to explore and self-regulate their learning.

The scenario of the second study introduced an immersive application to teach the history of the Holocaust in a museum. This research, which takes a more user-study-oriented approach, is currently serving as the validation of the system in that museum. We wanted to assess how users navigate in the virtual environment (both in spatial and information terms) and how this relates to their previous knowledge, interest and emotional state and how it affects their perception of the experience.

10.2 How this thesis answers the research questions

As previously stated in Chapter 1, this thesis was conducted with six purposes: (1) to assess how actual and perceived performance relate to each other when participants evaluate their answers, (2) to analyse how those differences differ depending on students' gender and are affected by feedback, (3) to assess which other distal variables would be related to students' metacognitive abilities and behaviour in a mathematical task, (4) to analyse how behaviour and self-assessment differ in autism during a collaborative task, (5) to assess how users navigate in digital experience to provide Cultural Heritage content, and (6) to develop a methodology to teach robotics and programming in a tangible, discovery-oriented way. This chapter discusses the findings of each study with regards to each research question.

The first three studies corresponding to chapters three, four and seven, respectively, answer the first question ("How does actual and perceived performance relate to each other when participants have to evaluate their answers?"). In all the three studies, participants overestimated their competence by reporting larger confidence. Despite the general ceiling effect in the second study (Chapter 4, "How confident are you? A study on feedback and gender"), which did not allow to check for overconfidence in participants that did all the task correctly, the ones outside of that range showed overconfidence in their performance. The same studies also help us answer the second question ("How do those differences differ depending on students' gender?"). In all the three studies, female participants showed lower confidence than males, despite this difference being smaller when performance is low. Moreover, in the study of Chapter 4, only underconfident female participants did not improve their calibration.

For the third question, we extended the range of the measures to cover not only confidence but also other metacognitive variables. Like this, we asked, "Which other distal variables would relate to students' metacognitive abilities and behaviour in a mathematical task?". It is important to remember that distal variables are obtained outside of the task, contrarily to proximal ones, which are obtained during it. To answer this, we carried out the study in Chapter 5, "Metacognitive factors behind rule change in the Balance Scale task". The distal variables that we tested were students' perceived competence (cognitive, self, social and mathematical), goal orientation, mathematics anxiety and intolerance for uncertainty. We studied how these traits related to their behaviour during a task and the discrepancy between perceived and actual competence.

Another purpose of this thesis was to assess the deployment of real-world learning applications. First, we run a study to determine users' behaviour while using a digital experience to present knowledge about the Holocaust. This fifth question ("How do users navigate in a virtual experience to provide Digital Heritage content?") is answered in Chapter 6 ("Virtual Reality for Historical and Cultural Learning"). In it, we could see that users mostly explored content related to personal stories and seemed to change their emotional state towards more positive and empathetic emotions. We also created a course to teach coding and robotics to elementary-school children, under the question ("How can we create a methodology to teach programming, in a tangible and discovery-oriented way?"). In Chapter 7 ("A methodology to teach programming and robotics to children"), we created a methodology, based on Computational Thinking, focused on visual representations, verbal explanations and discovering on their own. Children enjoyed more those parts of the method that involved crafting, creativity and the usage of computers, as well as those that promoted teamwork.

In our last question, we wanted to assess how prediction and collaboration abilities differ in autism, represented in the question "How do behaviour and self-assessment differ in autism during a collaborative task?". The study in Chapter ("Collaboration Finally, variability in Autism Spectrum Disorder") was created to answer that question. We found differences in predictive and collaborative behaviour between neurotypical participants and participants in the autistic spectrum, which were not reflected in the self-reports.

Further sections present more detailed conclusions of each of the parts of this thesis.

10.3 How this thesis supports the presented learning principles

Conclusions on Principle 1: Principled Conceptual Knowledge

This principle accounts for the importance of structuring new and existing knowledge around the central concepts of the discipline. This knowledge should be both accessible and usable. Experts' content knowledge is usually structured around the central concepts of the specific domain.

In this case, this principle was followed by presenting the information ordered by difficulty. Like this, participants in the studies related to the Balance Scale task received content tailored to the rule they referred to.

Conclusions on Principle 2: *Prior knowledge*

This principle postulates that learners use their previous knowledge to construct new one, by linking the new one to what they already know, what implies adding, modifying, or reorganising existing knowledge or skills. Moreover, when facing new content, students do not only already possess knowledge but also beliefs and misconceptions about that topic, which can significantly affect how they approach new learning (Wandersee et al., 1993).

In four out of the six studies presented, we followed Principle 1 by assessing participants' prior knowledge before carrying out the task. In the second study (Chapter 4), children had to fill in a pre-test for us to know their baseline knowledge. In the third one (Chapter 5) the task contained a first block where participants were presented with examples of the five item types. Their performance in each of the item types allowed us to determine the rule they were following, to later adapt the task difficulty to their level. Moreover, in the previous session, participants filled in four psychological scales and an arithmetic test, which allowed us to split them equally in the six conditions.

In the fourth study (Chapter 6), we assessed users' perceived knowledge about the applications' content. At the moment, that information was only used for informative purposes. Nevertheless, further research could explore how adapting the content to their previous knowledge affect their experience with the system. In similar lines to the second and third studies, the pre-test in the fifth study (Chapter 7) allowed us to assess children's previous knowledge in programming and electronics components.

Conclusions on Principle 3: *Metacognition*

This principle defends the use of metacognitive strategies to identify, monitor and regulate cognitive processes. Learners need to assess what they already know and what else they need to know in any given situation to be effective problem solvers. To do so, they have to consider both factual (the one about the task, their skills, and the goals) and strategic knowledge (the one about how and when to use a specific procedure to solve the problem).

The relationship between students' metacognitive abilities and their learning process is one of the cornerstones of this thesis. The next section (10.4, "Conclusions on metacognition") presents more detailed conclusions about the topic.

Conclusions on Principle 4: *Differences among learners*

This principle highlights the importance of considering students' different abilities, strategies and approaches when facing a learning task. There can be substantial differences in cognitive skills between learners, even among the ones of the same age. Moreover, these differences can also relate to emotional and motivational characteristics, as we will see in the next principle

The importance of considering learners' characteristics and differences during their learning process is one of the cornerstones of this thesis. The following section (10.5, "Conclusions on individual differences in metacognition") presents more detailed conclusions about the topic.

Conclusions on Principle 5: *Motivation*

This principle accounts for the effect of learners' motivation to learn and sense of self on their learning process. Regardless of this motivation being either extrinsic (performance-oriented) or intrinsic (learning-oriented), it can affect students' willingness to persist in a task, affecting their results. For example, learners' beliefs about their abilities in a subject area strongly relate to their success in learning about that domain. Moreover, those results can also be affected by learners' perceptions of their abilities as predetermined or as substantially affected by their effort (Dweck & Henderson, 1989).

This principle is related to the results of the study in Chapter 5 ("Metacognitive factors behind rule change in the Balance Scale task"), as the factors used in the study can also be considered motivational factors.

Conclusions on Principle 6: *Situated learning*

This principle defends that knowledge is contextually situated and shaped by the context and the activity in which it is used. Knowledge has thus to be presented in authentic contexts, that is, those settings and situations that would usually involve that knowledge. Moreover, situated learning theory also defends the idea of learning by doing, where learners can see the implications of their knowledge.

This principle is related to the studies in chapters six and seven ("Virtual Reality for Historical and Cultural learning", and "A methodology to teach programming and robotics to children", respectively). In the first case, it is due to the location of the study, not only in the memorial itself but also considering the history behind the building. In the second case, it is due to the idea of learning by doing, as it is the cornerstone of the methodology.

Conclusions on Principle 7: *Learning communities*

This principle highlights the importance of socially supported interactions to enhance students' learning process. Providing learners with the opportunity to interact and collaborate with others allows them to discuss their ideas with others and to learn from observing others. Moreover, social interaction is essential for students' metacognitive skills and the formation of their sense of self. Meaningful conversations do not consist only on discussing facts or procedures; but also, on applying ideas and raising questions.

This principle is related to the results of the studies in chapters 7 and 8 ("A methodology to teach programming and robotics to children" and "Collaboration variability in autism spectrum disorder", respectively). In the first case, it is due to the collaborative nature of the methodology, where learners work in groups towards their final outcome and share their process and results with their peers. In the second case, it is due to the implications of the collaborative deficits in individuals in the autistic spectrum.

10.4 Conclusions on metacognition

Traditional educational paradigms promote learning as a mere knowledge transfer; however, equally important is scaffolding student's learning process. In other words, the process of acquiring new information is as essential as knowledge acquisition itself. The assessment of confidence during learning allows for a better understanding of the learning process undertaken by the student. At the same time, it permits observing the implications of self-perception of one's skills on knowledge acquisition.

The central studies in this thesis related to metacognition are the first three ones (Chapters 3, 4 and 5) They show, in escalated order of complexity, the relationship between students' performance and their metacognitive abilities (mostly, focused on the confidence in their answers).

The results of Chapter 3 present the importance of assessing students' perception of their abilities. Despite the varied spectrum of performance in the sample, participants' confidence in their answers remained almost stable through them. Despite the Dunning-Kruger effect being known as unskilled being unable to perceive their performance, another explanation has also arisen in the literature. Poor performers may possess the metacognitive abilities to recognise their incompetence; nevertheless, their self-enhancement needs may lead them to ignore it when the domain of that ability is not essential for them (Kim, Chiu, & Bregant, 2015).

The results of this study go in line with three of the seven principles of learning presented in the introduction. The main principle they relate to is *Metacognition* (Principle 3), for the reasons previously stated. The presented gender differences in the discrepancy between actual and perceived perforqmance highlight the importance of considering Principle 4, that is, *Differences among learners* when assessing a learning experience. Finally, females' performance variability depending on the class' gender ratio reinforces Principle 6 (*Situated learning*), the idea of how context can affect one's learning experience. These last two parts of gender will be further explored in the next section.

In the second study presented (Chapter 4), we found a highly significant correlation between the bias during the pre-questionnaire and the improvement after the task. More specifically, overconfident students (positive bias) improved more compared to underconfident ones. This result could be explained by the *hypercorrection effect,* which postulates that high-confidence errors are more likely to be remembered later due to the inconsistencies provoked by the feedback to the learner's internal model. These results support the *resistance* phase of DAC's pedagogical model, as this discrepancy between the perceived and real performance leads to *confusion*, accompanied by a drop in one's bias.

Engaging learners in externally reporting their current confidence is a way of supporting their awareness of their learning process. It must be said, though, that our primary purpose is not to directly increase learners' confidence, as it could be counterproductive if their abilities are low. Our goal is to promote the improvement of their metacognitive skills to help them to be aware of their needs and achievements. This process would help them deal with failure more productively, as they could learn from it. Moreover, that could help them not see their knowledge as something that they can improve instead of as something static, as happens with learners with a fixed mindset (Mueller & Dweck, 1998).

Enabling learners to be aware of their current abilities and properly self-monitor their improvement may serve as a way to set the ground for a proper learning process. Providing learners with content that suits their current needs and skills can fail in increasing their knowledge if they are not aware of those needs. A learner that does not have a proper model of their abilities may not be able to obtain reasonable profit from the provided feedback. Thus, it is essential to provide students with tools that allow them to detect their capabilities. By doing so, they may foster a mastery-oriented approach in which they will look for new challenges whenever they have mastered the previous content.

Improving students' metacognitive abilities is not only a cornerstone to promote self-regulated learning but also to palliate school drop-outs. Moreover, helping learners to calibrate their metacognition could be a step towards eradicating effects as the *confidence gap* mentioned before, helping female students to pursue science-related careers. Although more exploration may be needed, this can be seen as insight on possible approaches to improve learners' metacognition and, specifically, scaffold more accurate confidence in female students.

a) Implications

An analysis of subjects' metacognition during the whole study (during the task, and both pre- and post-questionnaires) supports DAC's learning phases. More specifically, DAC proposes that initially, the student tries to defend their current knowledge and world model, or what we call resistance. *Resistance* implies a certainty of one's world model, and hence, we would expect the learner to be overconfident of their performance. Indeed, the positive bias in the pre-questionnaire shows that participants showed overconfidence in their abilities, as they did not accurately measure their performance. Additionally, in *confusion*, learners understand that their world model is not correct and needs to update it. Hence, one would expect a drop in confidence, as now, we can assume uncertainty of the updated model.

Our results demonstrate that during the task, a significant decrease in the bias occurs: as they begin to receive feedback about their performance, the model of their abilities is threatened. Finally, *abduction* signifies acquisition of stabilisation of knowledge, which in turn may suggest confidence that matches the user's performance. In our scenario, the intervention acted as a way to acquire information; and the responses of the post-questionnaire could match with abduction. Indeed, results showed that students' bias is close to zero. We argue that a bias close to zero indicates that learners can adequately evaluate their abilities without needing feedback about them, suggesting that they have reached the desired metacognitive skills, or they have arrived at abduction.

b) Strengths & Limitations

The main strength of these studies is the sample size of two of them (Chapter 3 and 5). In comparison to the original research, our Unskilled and Unaware study has about four times more participants than the original study. This allowed us to run, not only the general analysis but more specific ones, considering, for example, participants' gender or differentiating among question types. Another strength is the study of gender differences, which did not appear in the original study. The sample size of the study in Chapter 5 allowed us to run a finite mixture model, in this case

One of the main limitations of the study presented in Chapter 3, as happened in the original one, is the sample profile. The use of student sampling, despite benefits like the easy access and low cost -if any- of data collation, presents a biased sample, which may not be representative of the general population. Another limitation could be the design of the test, which, as previously mentioned, requested students to report their demographics (age and gender) at the beginning of the test. Asking participants to previously report their gender could affect females' performance and confidence reports, as it has been shown to affect how females answer questions (mostly, if these contain science-related topics). For that reason, new versions of the test ask students for their demographical data at the end.

One should also take into consideration issues like availability and production deficiencies in regulation (Veenman, 2013). This notion suggests that learners may differ in their consistency to regulate their cognition. It is crucial, thus, to identify the type of metacognitive deficiency expressed by the leaners, so teachers (or adaptive systems) can make more informed decisions for their instructional actions. It should also be considered is that children's overconfidence came from an evaluation not (only) of their performance but their perceived amount of effort.

In the case of the presented work, the study in Chapter 5 aimed to evaluate possible traits related to students' metacognition and behaviour. Other variables that could have been considered are belongingness (the degree to which students feel they belong in a course, it relates to) or perception of instrumentality (the perceived usefulness of a task or subject).

c) Further steps

The presented studies' further steps fall into two categories: One related to the psychological/educational implications and another one regarding the development of educational technologies. First and foremost, this study will lead to further research on learners' metacognitive abilities and their effect on learning. At the moment, this study focuses on confidence self-reports, but we can see future work on other variables, like self-efficacy, perceived competence, or mathematical anxiety.

Further steps in this research would come from analysing different types of metacognitive training depending on the discrepancy between actual and perceived performance and the side of the percentiles the participant is in. Considering low performers as lacking sufficient task-relevant knowledge to properly assess their performance, a solution could be on first focusing on training them to firstly perform well and then train their metacognition. Nevertheless, that could appear as a too simplistic strategy, as if the perception of their accuracy is not corrected, it is possible that some may not benefit from skill training, due to resisting to accepting new knowledge crashing with their existent one.

Using incentives as a motivation to improve metacognitive awareness has neither reached strong outcomes (Ehrlinger et al., 2008). Contrarily to what was expected, in that study, low performers became even more overconfident when there was either a monetary or a social incentive. Other steps go towards studying meta-monitoring. For example, following (Miller & Geraci, 2011)'s example, asking participants to provide second-order judgments (e.g. confidence judgments for previous confidence judgements), analysing, like that, learners' awareness of their metacognitive skills.

Further steps could also go in the line of evaluating others' judgements. The divergence mentioned above does not only account for self-perception but also the perception of others. Thus, less competent individuals will not just overestimate their skill level but also not identify skill in other subjects, even after being trained to improve their abilities.

When considering a change of strategy, there could be several options why a student would choose or not going towards a more complex strategy. There is their knowledge on the task (you would not select strategies that do not even know that exist); but what happens when a student has the ability to pursue a more complicated strategy but remains in the current one? They can be not perceiving their skills (as reflected in the self-resistance learning phase from DAC) or even, although they perceived them, not feel sure enough to pursue the next rule. This could be due to fear of uncertainty, not knowing how their strategy will come out and being afraid of changing it, it could also be that the person has a very performance-oriented goal, where they focus on doing only what they know, either to gain more positive feedback (approach-oriented) or to avoid negative one (avoidance-oriented). The question here would be how do we help the student to become a mastery-oriented individual, not afraid to take chances and that not only answers them correctly but *utilises* the newly-gained strategies?

10.5 Conclusions on individual differences in metacognition

In this thesis, we have shown how students' characteristics relate to their metacognitive abilities and their behaviour during a task. In this case, the main two factors this thesis has focused on have been gender and neurodivergence. Another possible characteristic to consider is age, as there seems to be differences between children and adults in how both age groups deal with errors. Like this, children may remain overly optimistic even after receiving feedback, as a protection against a loss of motivation (Bjorklund, 1997). Contrarily, adults tend to use more their prior error experience to regulate their metacognition, what is known as "underconfidence with practice". In any case, as our studies with both ages group are quite different, we will not compare them. The following subsection, thus, will discuss the relationship between metacognitive abilities and learners' gender or neurodivergences.

a) Gender

As we have already begun to see in the previous section, students' metacognition is highly related to their gender. Some studies have found that girls rate their ability as lower than that of boys as early as the first year of primary school – even when their actual performance does not differ from that of boys (Fredericks and Eccles, 2002; Herbert and Stipek, 2005). Nevertheless, there is no clear consensus in gender differences in metacognition, with some studies showing more accurate metacognition in girls (Bidjerano, 2005; Zimmerman & Martinez-Pons, 1990) while others show it in boys (Niemivirta, 1997). On another side, some studies discuss the possibility that female students are not less confident than male students, but they are relatively less overconfident.

In the first study (Chapter 3, “Students’ difficulties in recognizing their competence level”), we found differences between genders in the discrepancy between actual and perceived competence, significant for the 2nd and top quartiles. When dividing between correct and wrong answers, females seemed to show higher accuracy than males to distinguish between them (reflected in significant differences in females but not in males between confidence reports for these two kinds of answers). Moreover, females showed lower confidence than males in wrong answers but not in correct ones. This result is in line with previous results showing larger error-sensitivity in females. Unexpectedly, we found differences in performance between genders, with males scoring higher (which is reflected in larger confidence). It is essential to highlight, though, that females’ performance seems highly related to the gender ratio in the class.

In the second study (Chapter 4, “How confident are you? A study on feedback and gender”), we found significant differences between genders in confidence but not in performance, larger in the pre-test. Moreover, when assessing the evolution of the bias (the difference between actual and perceived performance) during the experiment, only the group of the underconfident females did not improve their calibration at the end of the task. In the third study (Chapter 5, “Metacognitive factors behind rule change in the Balance Scale task”), we assessed students’ differences between actual and perceived performance and several metacognitive and motivational factors related to them. Females showed higher mathematical anxiety and lower perceived competence (in the cognitive, self, and mathematics domains) than males. Unexpectedly, we also found initial gender differences in performance, with males scoring higher than females.

The results of this thesis highlight the importance of considering students' gender when creating learning experiences. Moreover, the relationship between gender ratio in the class and performance in females underline the importance of considering representation and students' distribution when designing classes and collaborative activities. These results are not only useful for educational psychologists but also for teachers and instructional designers.

b) Neurodivergence

The results of Chapter 8 ("Collaboration variability in Autism") showed differences in collaborative patters between neurotypicals and people in the spectrum. Nevertheless, the groups did not differ in the self-perception of their collaboration. Participants in the spectrum showed more variable behaviour and slower and less adaptation to the other agent than neurotypicals. Moreover, they focused more on the movement of the other agent when describing it, rather than in other characteristics, like the colour.

Usually, individuals appear to know whether they are right or wrong and change their actions accordingly to the desired result. Nevertheless, people in the spectrum can be described as knowing when they are correct, but not knowing when they have made a mistake (so no clash of intent) and have a related decreased ability to change actions accordingly. Some possible explanations arise Deficits in discriminating between intended and unintended outcomes; monitoring the intentional outcome and reporting on the source of an unexpected result.

Considering this, a question arises: Are individuals with ASD aware of their performance but choose to mentalise, that is, adapt their responses to counteract for their lacks or they are not aware of it and have a lack in metacognition? To assess this, we would need to design a task less related to their lacks due to their neurodiversity, something they don't feel they need to counteract for.

The relation between metacognitive regulation and executive functioning may be consistent with the executive dysfunction in

ASD (for example, perseverating with erroneous responses). Some authors discuss that one possibility is that social difficulties in autism are not caused by an inability to accurately recognise emotions but by to assess whether that emotion was correctly recognised. However, it is still not clear if this possible overlap between executive dysfunction and meta-representation of mental states is due to a deficit metacognitive monitoring and regulation.

c) Implications

These results showed the importance of considering students' differences when creating learning experiences for them.

Despite the advancements towards gender equity in terms of cognitive performance, the gender gap mentioned above is still present regarding its effect in mathematics learning. The current educational paradigms need more efforts to support females' interest in mathematics. The early appearance of negative patterns towards STEM-related subjects already in early adolescence indicates the relevance of including interventions to improve girls' attitudes for these subjects already in elementary school years. Nevertheless, not everything remains inside the curriculum: It is also crucial to consider learners' social environments (like family, peers and teachers) that could be affecting their perceived competence (Pekrun, 2000). For example, previous research showed the effect of teachers' implicit gender stereotype in science on female students' motivational beliefs and educational choices (Thomas, 2017).

In order to palliate genders' differences in self-perception in STEM-related subjects, it is first crucial to know if there is stereotype threat in the classroom. Teachers should consider if the materials (or even themselves) may be communicating stereotyped-messages at the school; they should also provide examples and opportunities for both genders to express their ideas and knowledge in class; finally, teachers should give opportunities for practice and feedback before assessments, considering how the lack of them appears to negatively impact performance in females (Nicol & Macfarlane-Dick, 2006). Moreover, in line with the first suggestion, previous research showed the overrepresentation of male protagonists in science books, together with the stereotypical portrayal of scientists (Makarova & Herzog, 2015), which have been shown to have an impact on secondary school students' mathematical anxiety and understanding of science, regardless of their gender (Good, 2014).

In the case of a student with neurodivergence, the study presented adds to the current research fields of autism and collaboration a new measure of collaboration. Nevertheless, studies showing ASD behaviour during collaboration should not only matter to researchers and practitioners in the field of autism but also to teachers and any education-related agent that wants to open up their classroom to neurodivergences.

We have previously presented the importance of promoting collaborative activities in the classroom. If we want to encourage collaboration between students, we need to take into consideration the individual differences of each of them, not only in characteristics like age, gender or previous knowledge but also their neurodivergences. Moreover, the variability inside of the ASD group highlights the importance of considering not only students' neurodivergence but also the variability inside of the same neurodivergence

While this study provides a preliminary insight, several limitations need to be discussed. First of all, further studies with larger sample sizes are required in order to better control for individual differences. Furthermore, it is essential to note that our study lacks female participants, as the primary general users of the ASD centre we collaborated with were males. This is in line with the larger occurrence of autism in male individuals compared to female ones. Despite these limitations, this study proposes a simple (and non-invasive) method to evaluate the predictive abilities of individuals in the autism spectrum. More data would be needed to achieve this, as the main limitation of this study is the weakness of its statistical power.

d) Further steps

Further research in the gender domain could go in the line of assessing differences between subjects (comparing STEM-related and non-related ones) in females' metacognition. An examination of the gender-by-age interaction in metacognition and motivation would also be useful to assess how they evolve during primary and secondary school years. Moreover, in the studies presented, none of the participants reported a different gender as their sex. Nevertheless, we find it important that further research takes into consideration this differentiation.

Regarding the neurodivergence domain, possible uses of this application would go in the line of an environment where the user could train their social abilities in a controlled and adaptive way. The system could be used to improve the skills of non-neurotypical people by training their predictive skills. Like this, individuals in the spectrum could not only train their tracking of moving objects and predict their trajectories but also train their reading and understanding of non-verbal cues. The task offers the possibility of merging these two types of prediction (related to objects and social interaction), in a game-like manner.

To the moment, the PSSE has not been contrasted with any kind of diagnostic tool for ASD. In the future, a validation of the PSSE measurement in comparison with a screening tool could allow for a stronger claim on distinguishing between these two groups. However, at this point, we do not claim that it can be either a diagnostic tool or a tool to be able to distinguish between the two groups, but we highlight the possibility.

10.6 Conclusions on educational applications

This thesis developed on three parts regarding educational applications. In Chapter 7, it presented a methodology to teach programming and robotics to elementary-school children. It also introduced a VR application to provide content about the Holocaust in a museum (Chapter 6). Finally, it proposed a pedagogical model of learning that could be used when developing a tutoring system.

a) Methodologies to teach programming and robotics

The purpose of the work presented in Chapter 7 ("A methodology to teach programming and robotics to children") was to provide the necessary information for educators to develop a robotics curriculum. More specifically, we presented the current state of the art in the field of Educational Robotics (ER), and how Computational Thinking (CT) can be used to teach programming and robotics. We also discussed the relevant learning theories that support ER and teaching, and how they are related to the developments of metacognition, learning, and the role of gender in STEM. We further presented a methodology developed to teach robotics and coding in elementary schools, namely "Coding Robots through Exploring their Affordances" (CREA). We discussed how CREA was designed and organised, and what were the main outcomes of this methodology when applied in an extra scholar activity of children between 9 and 11 years old.

Results suggested that the use of worksheets (where students present their thoughts in a way where their classmates can use them) can promote thinking in a way that is understandable by computers. The use of an open-sandbox software, worksheets and code-based puzzles can help students overcome their resistance to syntaxes and programming environments. Working together with their classmates (both in the same team and others) and presenting their work to the class promotes collaboration and communication. The main challenge found during the development of this methodology was the differences between children in their interest in the topic. One limitation is the lack of females (only 1 out of 10), which would allow us to look for gender differences, as suggested by the literature.

Furthermore, the CREA method allowed students to practice the 4 C's of the 21st century: collaboration, communication, critical thinking and creativity. We propose that robotics is not only about acquiring knowledge on technical issues, but also develop another set of skills that can be used in a variety of domains. As Bers (Bers, 2017) advocates in her book "Coding as a playground", the study does not aim to see coding and robotics as a way to fulfil the demand of the future workforces but to present them as a new way of thinking. ER is not the destination, but the journey.

ER provides children with opportunities for creation and action. It offers them the chance to increase their autonomy by influencing their problem-solving skills given by learning how to program. Building robots seems to increase students' confidence in their capabilities to master technology (Huang, Varnado, & Gillan, 2013). Similarly, ER also allows learners to improve their metacognitive skills (MS) -that is, thinking about thinking (Flavell, 1979)-, which have been proven to be one of the cornerstones of learning. Some studies showed ER's positive outcomes in MS (Gaudiello & Zibetti, 2013; Huang, Varnado, & Gillan, 2014; Lin & Liu, 2011; Paglia, Barbera, Caci, & Cardaci, 2010). When failure is seen as an opportunity to learn rather than an outcome to avoid, children are more prone to explore their limits.

Introducing children to the world of robotics does not only educate them with high tech digital skills but also provides them with a new way of observing and understanding the world surrounding them. It gives children the means to understand how the devices and algorithms they use every day empowers them to be literate citizens that can think critically on the meaning of the technology they use and take decisions in a more informed way. For example, the children of today will be (if they are not already now) the users of tomorrow's social networks. Thus, computational literacy will move them from a passive role of digital consumers to a more active one, where they can create those devices or, at least, understand how they work. They are no more mere recipients of information but become active agents in it. In this kind of approach, teachers step aside from their traditional role to become, as Constructivism already presented, "a guide on the side". In this line of thinking, the technology can serve as a communicational mediator between the student and their peers and teachers.

b) Implications

The authors' recommendations for other researchers in the field follow similar lines to Carnegie Mellon University's Computational Thinking Practices Framework (Flot, McKenna, & Shoop, 2016). It is recommended to support students' analysis of their work and the work of others, promote their design and implement creative solutions, allow children to communicate their thoughts and their results, and promote collaboration with their peers. Moreover, it is vital to provide students with tools to think critically about the technology surrounding them.

The authors would also like to highlight the importance of encouraging metacognition in class by asking questions, facilitating classroom discussions, and giving students choice and ownership. Open-ended tasks are essential, as there is no one clear answer, and all students can contribute to the class. Other recommendations, also considering the previous conclusions on gender, would be to reduce gender differences in STEM, promote students' access to learning activities on ER and increase female representation in the field. In Computer Science related topics, female students seem to prefer activities related to cooperation and brainstorming more than males (Akinola, 2015). Additionally, taking into account gender differences in competitiveness, it might be preferable to avoid competitive activities, to ensure the participation of girls in STEM education. Moreover, gender stereotypes in school textbooks could also be considered (Papadakis, 2018).

It is also important to understand that teaching children about (or with) technology does not mean providing them only with technological literacy. Fields like roboethics may play a role in the educational robotics paradigm. According to Veruggio (Veruggio & Operto, 2008), "roboethics is applied ethics whose objective is to develop scientific/cultural/technical tools that can be shared by different social groups and beliefs". These tools aim to promote and encourage the development of Robotics for the advancement of human society and individuals and help to prevent its misuse against humankind. Thus, roboethics is not the ethics of robots, nor any artificial ethics, but it is the human ethics of robots' designers, manufacturers, and users (Breazeal, 2004; Veruggio, Operto, & Bekey, 2016). Including roboethics into ER is crucial, because, as Parker (Parker, Swope, & Baker, 1990) pointed out almost thirty years ago, today's young people who are getting their hands-on robotics kits will be the robotics professionals and consumers of tomorrow (Veruggio & Operto, 2008). Instead of viewing technology/robotics as the digital panacea to improve the 21st Century education, we need more curricula focused on the process as well as the outcome of the project.

c) Recommender Systems for Museums

The study in Chapter 6 ("Virtual Reality for Historical and Cultural Learning") presents the first steps in the validation of an immersive experience to present information about the Holocaust. The application seemed to induce some changes in users' emotional states. It helped them to understand how the camp worked and increased their interest in knowing more about the topic and visiting similar sites. It was perceived as innovative and interesting. Moreover, we found a strong focus on content about personal stories.

This study shows the possibilities of immersive experiences for presenting information about the Holocaust. The main limitation of this study is the sample size, which we plan to increase in further iterations. Due to time and space constraints, we did not conduct interviews or focus groups, which would give us more detailed insights on users' perception of the experience.

Possible next steps would be to implement an introductory video/text, as some users suggested, to introduce visitors to the history of the camp. Further steps could also go in the line of conducting experiments related to the content. One possibility would be to assess users' narratives when navigating through the information.

This research can be useful not only to historians and museum curators but also to researchers working on the fields of VR and AR, as the project studies different ways of presenting information to the user. Moreover, it can also be interesting for researchers on memory and learning, as visitors' decisions on the explored content can be informative about their learning processes.

d) Considering learners' metacognitive abilities when designing Intelligent Tutoring Systems

These studies are also part of a goal to develop a learning tool able to adapt to each individual and use the appropriate strategies to present the various knowledge elements, according to the needs, preferences, and performance of each learner (for more information about the project, check: Reidsma et al., 2016; Vouloutsi et al., 2016). Here, we are mainly interested in assessing the required inputs that would potentially allow for this adaptation to the user's metacognitive state.

The components of metacognition that a pedagogical intervention should try to address can also be divided between monitoring-related interventions and control-related interventions, like the classification mentioned above of metacognitive skills (Efklides, 2006b). When promoting metacognitive monitoring, one can ask the student about confidence in their answer, perceived difficulty, and self-efficacy (before the task). This would allow the learner to be more aware of their learning process. To promote metacognitive control, the system can, for example, prompt the student with strategies to use. As (Stamper & Koedinger, 2011) suggested, a proper student model can promote more effective learning by using resequencing, knowledge tracing, creating new tasks, and changing instructional messages, feedback, or hints.

Regarding educational technologies, the results we acquired helped to define a set of guidelines important to develop a system that adapts to the learner: previous knowledge and bias, confidence, and gender. As a future step, we aim to implement these factors to an Intelligent Tutoring System to adjust the educational content to learners according to their metacognitive abilities, and at the same time, provide them with tools to recognise and improve their metacognition. Additionally, it is essential to explore the adaptability of a system with younger age groups that may have less accuracy in their metacognition, as we can hypothesise their bias would be less precise. We argue that assessing the evolution of learners' metacognitive abilities before, during, and after the task can help us to understand the role of confidence in this task, and to further use this information for the development of a model and an adaptive learning system to scaffold students' learning process. Moreover, further steps can also go in the line of analysing more in detail the effect of the platform in the learning process.

e) Further steps

This work shows an exploratory study for the design and implementation of CREA, an effective and sustainable ER methodology to teach programming and robotics to elementary-school children. The proposed approach contains six main steps of increasing complexity to scaffold children's learning process. From the qualitative data obtained in the interviews, the authors understand that all the children liked the parts that involved crafting, using the computer and activities that promote teamwork. Some students enjoyed writing the instructions for the robot (but it becomes an issue for students with low (actual or self-perceived) writing skills. There is no clear consensus on whether students preferred to have many options to explore or to have constrained possibilities, so the task becomes easier to perform and conclude. Interestingly, all students reported not liking to deal with the Arduino syntax.

Another point to take into consideration is how to make the CREA methodology available for children with different needs. For example, for colour-blind students, the paper blocks could not only be different by colours but also by patterns. To avoid generating difficulties for learners with challenges in writing, the parts where children have to write down the instructions could be adapted to write in a computer or using already prewritten instructions. This last option would also allow the use of Parsons problems, where children are provided with lines of code that they have to order to make it work (Morrison, Margulieux, Ericson, & Guzdial, 2016).

10.7 Contribution of this thesis

The first contribution of this thesis is the advancement of the existing literature on metacognition, specifically, in the study of the mismatch between actual and perceived performance. This is presented in Chapters 3, 4, and 5. The second contribution enters more in-depth into the metacognition field by assessing gender differences in the effect of feedback during a task. We found that both students' gender and the feedback received played a role in metacognition and learning, as we found differences in confidence between genders but not in performance, and that those differences were still present during the study in females but not in males, even after receiving feedback. This is presented in Chapters 4 and 5.

The third contribution continues in the metacognitive spectrum by proposing several traits related to students' behaviour during a learning task and their discrepancy between actual and perceived performance. This is presented in Chapter 5. The fourth contribution presents the first steps towards the assessment of users' preferences for the development of tutoring systems in the field of History education. This is illustrated in Chapter 6.

The fifth contribution of this thesis proposes a collaborative methodology, based on Computational Thinking, focused on visual representations, verbal explanations and discovering on their own. The sixth contribution advances the study of ASD by presenting a task to measure prediction and collaboration, and the differences in individuals in the spectrum between actual and perceived behaviour during the task. This is presented in Chapter 8.

The contribution made by this thesis to STEAM education, in general, is to highlight the effect of student confidence and other characteristics (such as their gender or traits like perceived competence in mathematics or mathematical anxiety) on student learning and experiences. This is achieved by presenting a substantial range of literature on the various theoretical standpoints and several studies.

Self-perception in Mathematics and mathematical anxiety have been shown in this thesis to be problematic, especially for students with less mathematical knowledge, and especially females. Not only is improving student self-perception of themselves and their mathematical abilities a worthwhile aim, but the clear links found between metacognitive and motivational factors and achievement demonstrate that improving those factors would also promote improved achievement.

The development of metacognitive skills is critical in the preparation of learners for positive engagement within lifelong learning, the knowledge economy and, indeed, the knowledge society. The goal of this thesis was to understand metacognitive processing that occurs during learning.

These outcomes contributed directly to the goal of this thesis, by providing a better understanding of metacognition, and will help in the development of educational interventions that foster the metacognitive skills to enable learners to progress within learning contexts.

10.8 General future research directions

Altogether, the findings and limitations of this thesis offer ideas for future research directions. One of the possible improvements has to do with the length of the studies. Due to logistic constraints, all the studies presented in this thesis presented one-session experiments (in exception to the research presented in Chapter 5, which had two). Long-term studies would allow one to assess, for example, the evolution of learners' metacognitive abilities during a school year and how it related to their perceived competence or mathematical anxiety. It could also allow testing long-term retention and the effect of feedback after finishing the task.

Other improvements could go towards assessing participants' ability to extrapolate the acquired knowledge. That would permit to study a different dimension of metacognition, as it would require learners to identify the fundamental concepts of the current information they are facing, relate them to the previous knowledge and see how their previous strategies adapt to the new task.

10.9 General Conclusions

This thesis shows how students' characteristics affect their learning process and its outcomes. Thus, these characteristics, either being related to the gender or neurodiversity of the learners or their metacognitive abilities, should be considered by teachers and researchers when designing educational activities.

Metacognition is vital in education, mental health, and public life. It has been shown to improve learning in school settings (De Jager, Jansen, & Reezigt, 2005; Payne & Manning, 1992; Tauber & Rhodes, 2010; Yeager et al., 2019); it regulates compulsion, anxiety, and depression (Rouault, Seow, Gillan, & Fleming, 2018); promotes effective leadership (Edelson, Polania, Ruff, Fehr, & Hare, 2018; Fleming & Bang, 2018), and encourages moderation in political and religious discussion (Rollwage et al., 2018). Metacognition is an essential ingredient not only of our capacity to know ourselves, but to know ourselves together; to make decisions in groups that are better informed, fairer, and more reasonable than the decisions that each of us can make alone (Heyes et al., 2020).

One of the most critical reasons for improving metacognition is that it will enhance the application of knowledge, abilities and character traits outside the specific context in which they have been taught. This could lead to the transfer of skills across fields — important for learners training for real-life situations where specific boundaries in expertise are lacking, and skills must be chosen from the full spectrum in their knowledge to successfully overcome the tasks at hand. Nevertheless, this knowledge transfer may not be online across disciplines but also within the same one. For example, learners should know how to apply an idea learned with one specific example to another task.

The educational debate in the last century was if learners were or not empty vessels to be filled, for then moving towards studying how to continue filling that vessel. Now is the moment to asks us, not about the initial or later content, but about providing each of them with the necessary content and make sure, before that, that the vessel can receive it. It is not only about creating the knowledge but on helping students to welcome it or, even better, question it and create new one.

Supporting learners' self-reflection about their learning process, their challenges, and their strategies to overcome them, is crucial to scaffold their independence in the classroom. Like this, when they leave the educational system, they will be resilient adults equipped with tools to be lifelong learners.

APPENDIX

This appendix contains the tests, scales and interviews used in the studies presented.

Test from study in Chapter 3

Instructions

- Please answer the personal information questions. Only group results will be analysed
- Answer ALL the questions in order without going back to the earlier questions
- It is not expected that you perform complicated calculations to solve any problem. Just choose the answer that seems most likely to be correct to you
- For each question evaluate the rate of certainty of your answer - how confident you are you gave the correct answer (0% - 100%)
- The survey should take no more than 30 minutes to complete

Sex ……. **Age** ……..

Question 1

- You are offered the following gamble: ,100 flips of an unbiased coin. For every head you are paid 2€, for every tail you pay 1€. How probable is that you would accept the gamble?
    - Very improbable
    - Improbable
    - Neither improbable nor probable
    - Probable
    - Very probable

Certainty (0-100%):

0	10	20	30	40	50	60	70	80	90	100

Question 2

- Here is a description of a hypothetical person named Linda: *Linda is 31 years old, single, outspoken, and very bright. She had a double major in philosophy and music. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.*

Please choose the more probable case:

- ○ Linda is a bank teller
- ○ Linda is a bank teller and a feminist

Certainty (0-100%):

0	10	20	30	40	50	60	70	80	90	100

Question 3

- Here is a description of a hypothetical person named Linda: *Linda is 31 years old, single, outspoken, and very bright. She had a double major in philosophy and music. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.*

Please choose the more probable case:

- ○ Linda is a feminist
- ○ Linda is a bank teller and a feminist

Certainty (0-100%):

0	10	20	30	40	50	60	70	80	90	100

Question 4

- Read the following passage three times: *These functional fuses have been developed after years of scientific investigation of electric events, combined with the fruit of long experience on the part of the two investigators who have come forward with them for our meeting today.*

Read the passage one more time counting the number of *f* letters.
(It was placed on the other side of the paper)
Turn page.
How many times does the letter *f* appears in the passage? (DON'T LOOK BACK)

- ○ 6 or less
- ○ 7
- ○ 8
- ○ 9
- ○ 10 or more

Certainty (0-100%):

0	10	20	30	40	50	60	70	80	90	100

Question 5

- Which of the following sequences of heads (H) and tails (T) was more likely to have been generated by a random process with an unbiased coin?
  - ○ H T H T T H
  - ○ T H T T T T
  - ○ Both sequences are equally likely

Certainty (0-100%):

0	10	20	30	40	50	60	70	80	90	100

Question 6

- You are in a game show with an attractive prize behind one of a million doors. You are asked to pick the door that has the prize. You pick a door, say door 12.586. The host, who knows what is behind each door, selectively opens all of the other doors except door 698.453. You are given the option of changing your choice; in this case, you could pick door 698.453. Is it your advantage to switch your choice in this game?
    - Yes
    - No

Certainty (0-100%):

0	10	20	30	40	50	60	70	80	90	100

Question 7

- Solve the problem. Each of four cards has a number on one side and a letter on the other. Which cards must be turned over to test the hypothesis, *"If a card has a vowel on one side, then it has an odd number on the other side"*? The four cards are showing **6**, **L**, **U** and **9**, respectively.
    - L and U
    - U
    - U and 9
    - 6, L and U
    - U and 6

Certainty (0-100%):

0	10	20	30	40	50	60	70	80	90	100

Question 8

- You are in a game show with an attractive prize behind of three doors. You are asked to pick the door that has the prize. You pick up a door, say door 1. The host, who knows what is behind each door, selectively opens another door, say door 3, with no prize behind it. You are then given the option of changing your choice; in this case you could pick door 2. Is it your advantage to switch your choice in this game?
    - Yes
    - No

Certainty (0-100%):

0	10	20	30	40	50	60	70	80	90	100

Question 9

- What hero does Clark Kent become when he changes in a toll booth?
    - Spiderman
    - Superman
    - Batman
    - Can’t say

Certainty (0-100%):

0	10	20	30	40	50	60	70	80	90	100

Question 10

- A father and a son went for a drive and had a serious car accident killing the father and seriously injuring the son. The boy was rushed to the hospital, and the doctor on duty said, “I can’t operate on this boy, he’s my son.” Is this possible?
  - Yes
  - No

Certainty (0-100%):

0	10	20	30	40	50	60	70	80	90	100

Question 11

- A psychic claims to know the score of an important game before it begins. Do you believe her?
  - Yes
  - No

Certainty (0-100%):

0	10	20	30	40	50	60	70	80	90	100

Question 12

- Consider the following categorical syllogistic reasoning:

  All psychologists are scientists.
  Some scientists are mortal.
  ____
  Some psychologists are mortal

  (You can choose more than one option)
  - This argument is valid
  - This argument is invalid
  - The conclusion is true

- The conclusion is not true

Certainty (0-100%):

0	10	20	30	40	50	60	70	80	90	100

Question 13

- Consider the following categorical syllogistic reasoning:

  All xenos are oxons.
  All oxons are red.
  ——————————-
  Some xenos are red

  (You can choose more than one option)
  - This argument is valid
  - This argument is invalid
  - The conclusion is true
  - The conclusion is not true

Certainty (0-100%):

0	10	20	30	40	50	60	70	80	90	100

Question 14

- If **p**, then **q**. I observe **p**. Therefore, I conclude that **q** must be the case.

  (You can choose more than one option)
  - This argument is valid
  - This argument is invalid
  - The conclusion is true
  - The conclusion is not true

Certainty (0-100%):

0	10	20	30	40	50	60	70	80	90	100

Question 15

- If **p**, then **q**. I observe **p**. Therefore, I conclude that **p** must be the case.

  (You can choose more than one option)
  - This argument is valid
  - This argument is invalid
  - The conclusion is true
  - The conclusion is not true

Certainty (0-100%):

0	10	20	30	40	50	60	70	80	90	100

How confident are you that you answered the questionnaire correctly? (0-100%):

0	10	20	30	40	50	60	70	80	90	100

Tests from study in Chapter 4

Pre-test

- User ID *[We would fill it in for them]*

- Age
    - 8
    - 9
    - 10
    - 11

- Gender
    - Female
    - Male

- Do you have a mobile phone?
    - Yes
    - No

- Do you have a tablet?
    - Yes
    - No

Look at the card deck close to you. Each of the cards represents an activity. Do the following:

1. Choose the five activities (cards) in which you spend more part of your time. Choose only five cards.
2. Order those five cards based on how much time you spend doing each activity
3. Then, select them from this list (1 meaning the activity you spend the most time with; and 5, the activity you spend the least time with).

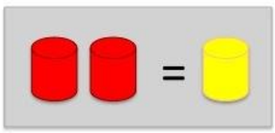
- Reading
- Listening to music
- Watching TV
- Cycling
- Doing puzzles
- Dancing
- Doing homework
- Playing an instrument
- Doing sports
- Playing board games
- Drawing / Painting
- Surfing Internet
- Playing with a pet
- Playing videogames
- Playing with dolls
- Playing with vehicles

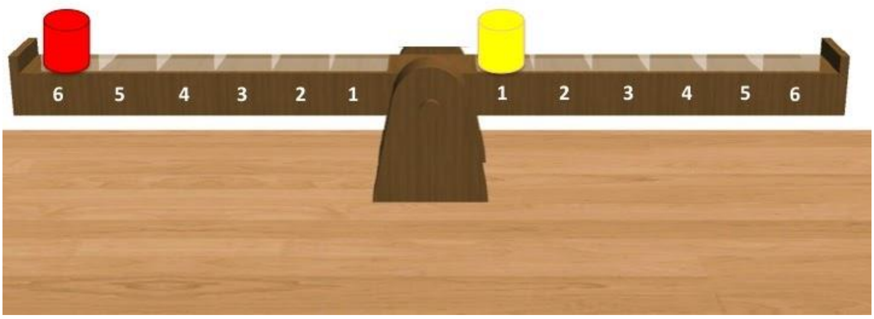
- Doing activities with friends
- Using a phone or tablet
- Playing with Lego

We will show you different setups of the balance. The weights will be already placed on both sides. You will have to answer where will the balance fall to.

You will have to:

1. Read carefully the instructions of the exercise
2. Look at the image
3. Mark down the answer (left, balance, right). Remember you can only give one answer!

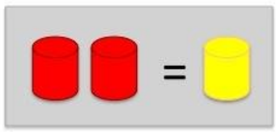


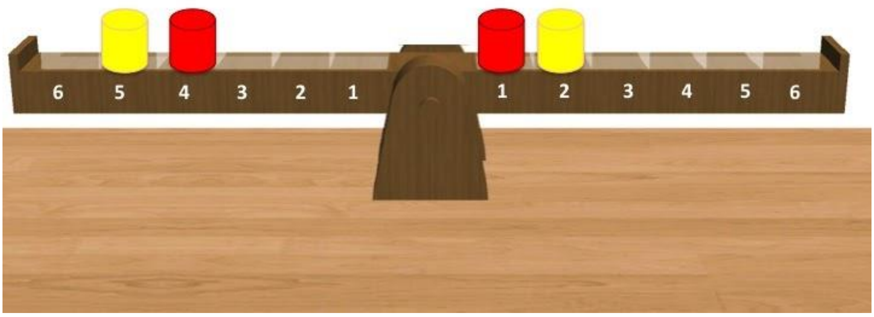


- Where will the balance fall?

- Left
- It won’t fall (it will remain in balance)
- Right

- How sure are you how having answered correctly? [Likert scale from 0 to 10, from *I am not confident at all* to *I am totally confident*]



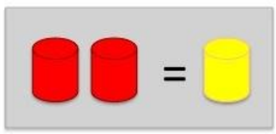


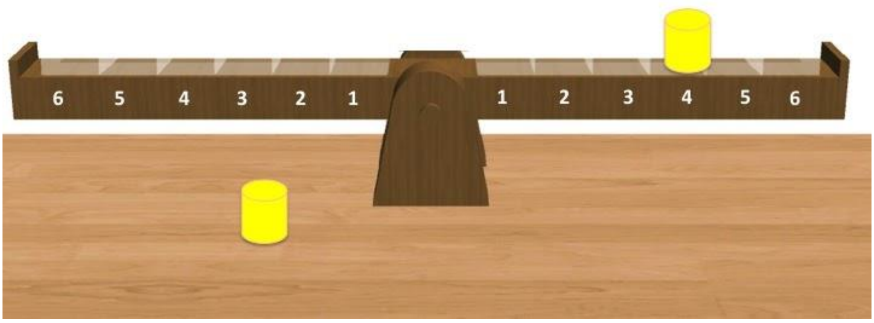
- Where will the balance fall?
    - Left
    - It won’t fall (it will remain in balance)
    - Right

- How sure are you how having answered correctly? [Likert scale from 0 to 10, from *I am not confident at all* to *I am totally confident*]

In the next images, the balance has only one weight on top. You will have to:

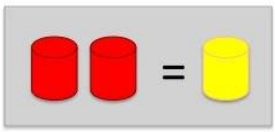
1. Read carefully the instructions of the exercise
2. Look at the image
3. Mark down the answer (1, 2, 3, 4, 5 o 6) of the position where you think the other weight should go. Remember you can only give one answer!

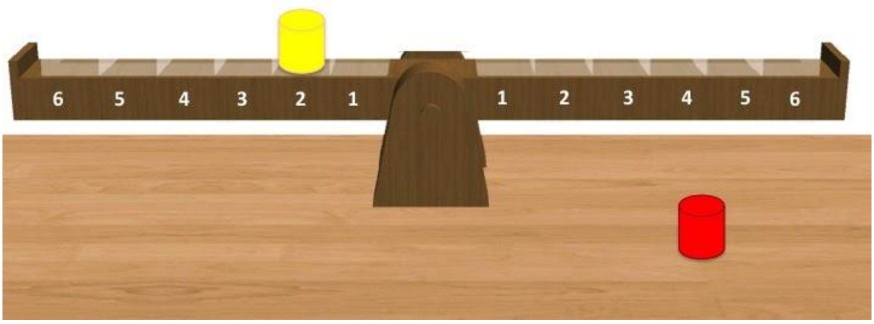




- In which position should you place the yellow weight at the left side so the balance remains in balance? [To choose from 1 to 6]

- How sure are you how having answered correctly? [Likert scale from 0 to 10, from *I am not confident at all* to *I am totally confident*]

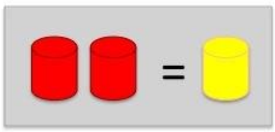


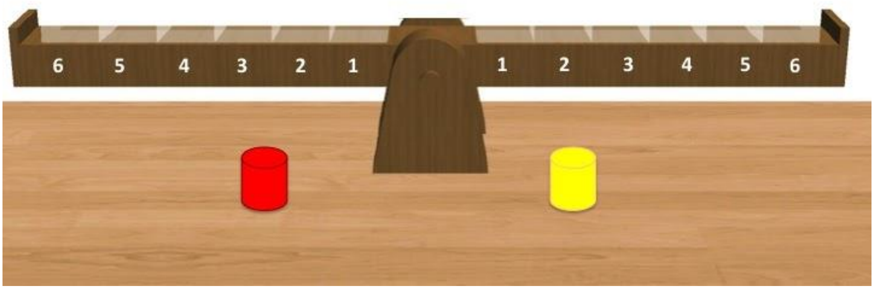


- In which position should you place the red weight at the right side so the balance falls to the right? [To choose from 1 to 6]
- How sure are you how having answered correctly? [Likert scale from 0 to 10, from *I am not confident at all* to *I am totally confident*]

In the next images, the balance does not have any weight on top. You will have to:

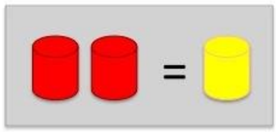
1. Read carefully the instructions of the exercise
2. Look at the image
3. Mark down the answer (1, 2, 3, 4, 5 o 6) of the positions where you think each weight should go to make the balance remain in equilibrium. Remember you can only give one answer per weight!

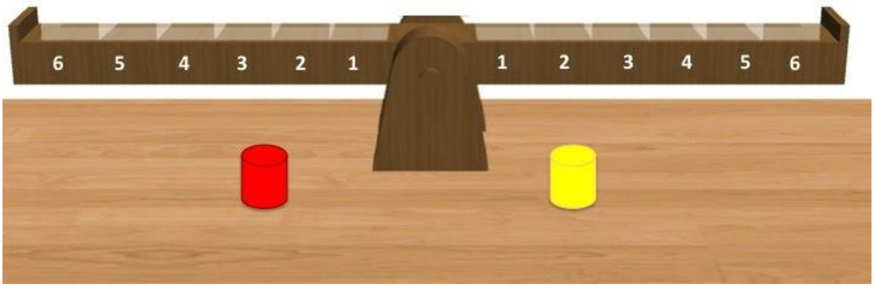




- In which position should you place the red weight at the left side so the balance remains in balance? [To choose from 1 to 6]
- In which position should you place the yellow weight at the right side so the balance remains in balance? [To choose from 1 to 6]

- How sure are you how having answered correctly? [Likert scale from 0 to 10, from *I am not confident at all* to *I am totally confident*]





- In which position should you place the red weight at the left side so the balance falls to the right? [To choose from 1 to 6]
- In which position should you place the yellow weight at the right side so the balance falls to the right? [To choose from 1 to 6]
- How sure are you how having answered correctly? [Likert scale from 0 to 10, from *I am not confident at all* to *I am totally confident*]

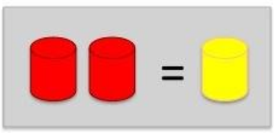
Post-test

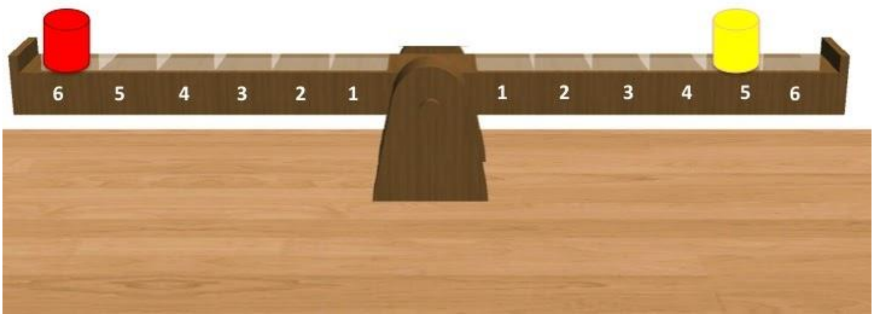
- User ID *[We would fill it in for them]*

We will show you different setups of the balance. The weights will be already placed on both sides. You will have to answer where will the balance fall to.

You will have to:

1. Read carefully the instructions of the exercise
2. Look at the image
3. Mark down the answer (left, balance, right). Remember you can only give one answer!

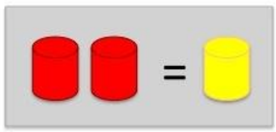


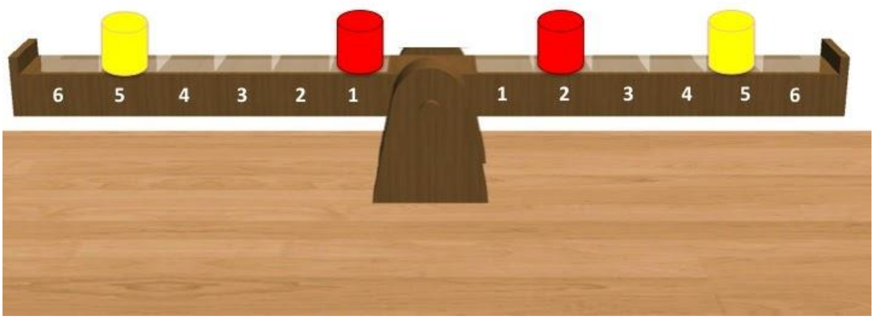


- Where will the balance fall?
  - Left

- It won’t fall (it will remain in balance)
- Right

- How sure are you how having answered correctly? [Likert scale from 0 to 10, from *I am not confident at all* to *I am totally confident*]



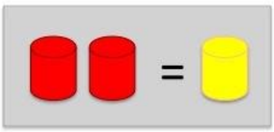


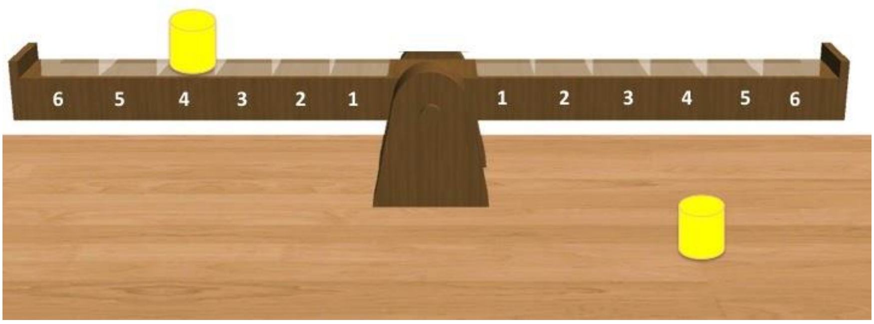
- Where will the balance fall?
  - Left
  - It won’t fall (it will remain in balance)
  - Right

- How sure are you how having answered correctly? [Likert scale from 0 to 10, from *I am not confident at all* to *I am totally confident*]

In the next images, the balance has only one weight on top. You will have to:

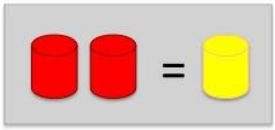
1. Read carefully the instructions of the exercise
2. Look at the image
3. Mark down the answer (1, 2, 3, 4, 5 o 6) of the position where you think the other weight should go. Remember you can only give one answer!

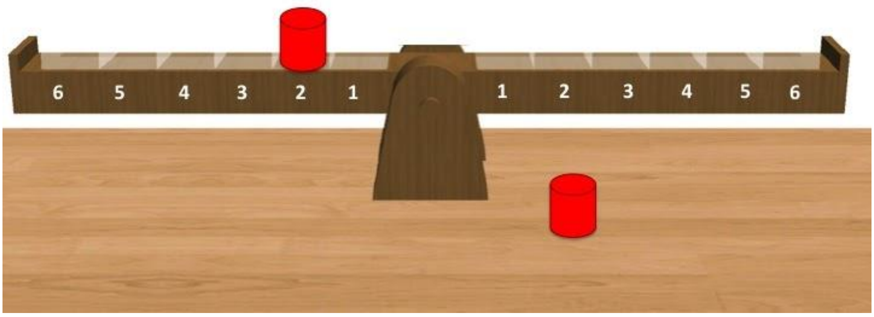




- In which position should you place the yellow weight at the right side so the balance falls to the right? [To choose from 1 to 6]

- How sure are you how having answered correctly? [Likert scale from 0 to 10, from *I am not confident at all* to *I am totally confident*]



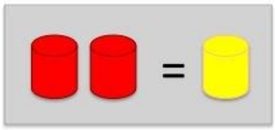


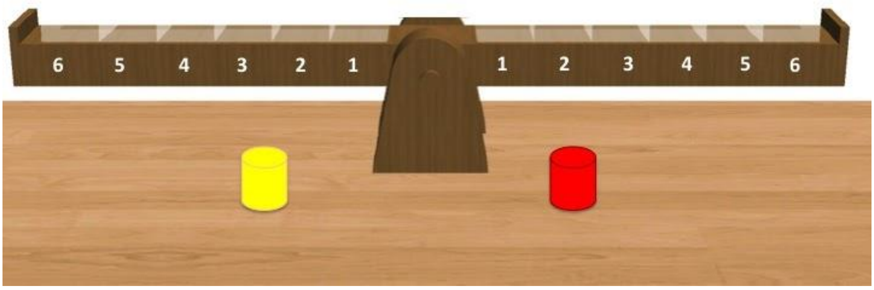
- In which position should you place the red weight at the right side so the balance falls to the left? [To choose from 1 to 6]
- How sure are you how having answered correctly? [Likert scale from 0 to 10, from *I am not confident at all* to *I am totally confident*]

In the next images, the balance does not have any weight on top. You will have to:

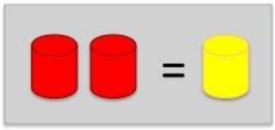
1. Read carefully the instructions of the exercise
2. Look at the image

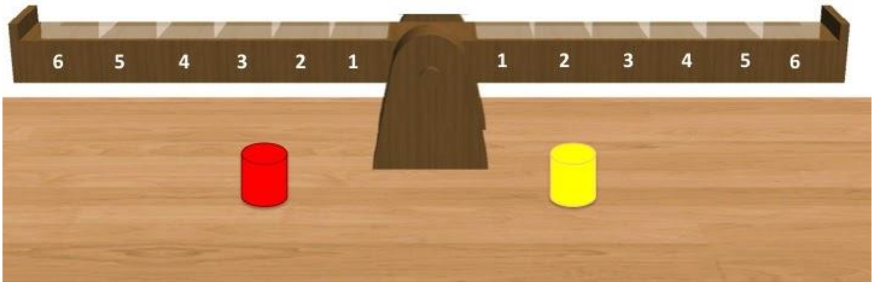
3. Mark down the answer (1, 2, 3, 4, 5 o 6) of the positions where you think each weight should go to make the balance remain in equilibrium. Remember you can only give one answer per weight!





- In which position should you place the yellow weight at the left side so the balance falls to the right? [To choose from 1 to 6]
- In which position should you place the red weight at the right side so the balance falls to the right? [To choose from 1 to 6]
- How sure are you how having answered correctly? [Likert scale from 0 to 10, from *I am not confident at all* to *I am totally confident*]





- In which position should you place the red weight at the left side so the balance remains in balance? [To choose from 1 to 6]
- In which position should you place the yellow weight at the right side so the balance remains in balance? [To choose from 1 to 6]
- How sure are you how having answered correctly? [Likert scale from 0 to 10, from *I am not confident at all* to *I am totally confident*]

- Would you do this task again? [5-point Likert scale, from *Not at all!* to *Yes, totally!*]

- Would you recommend this task to your friends? [5-point Likert scale, from *Not at all!* to *Yes, totally!*]

- Do you think the task was difficult? [5-point Likert scale, from *Very easy* to *Very difficult*]

Scales and tests used in Chapter 5

You will answer some questions about yourself.

Only us will know the answers. Neither your parents, classmates or teachers will see them.

If you had any doubt about any word during while answering the questions, please raise your hand in silence and the experimenter will come to help you.

The intolerance of uncertainty scale for children [Corner et al 2009]

Answer the following questions about yourself. Read the sentences at the left and think if you are like this. *Totally disagree* means that the sentence does not resemble at all how you are. *Totally agree* means that the sentence resembles a lot how you are.

Example:
In this example, the sentence is *I like ice-cream.*
- If I do not like ice-cream at all, I will click on the circle at the left (*Totally disagree*), which means something like *Not at all*
- If I like ice-cream a lot, I will click on the circle at the right (*Totally agree*), which means something like *A lot*.
- If neither I do not like it or I like it, I click on the circle in the middle.

Children were asked to rate all the following sentences in a 5-point Likert scale, from *Totally disagree* to *Totally agree.*

1. Doubts stop me from having strong opinions
2. Being unsure means that a person is mixed-up
3. Not knowing what will happen in the future makes life hard
4. It's not fair that we can't predict the future
5. I can't relax if I don't know what will happen tomorrow
6. Not knowing what will happen in the future makes me uneasy, anxious, or stressed
7. Surprise events upset me greatly
8. It frustrates me to not have all of the information I need
9. Not knowing what could happen keeps me from enjoying life
10. One should always think ahead to avoid surprises
11. Plans can be ruined by things you didn't think would happen
12. When it is time to do things, not knowing what could happen keeps me from acting
13. Being unsure of things means that I am not great
14. When I am not sure of something, I can't go forward

15. When I am not sure of something, I can’t work very well
16. Other kids have less doubt than I do
17. Not knowing what will happen makes me unhappy or sad
18. I always want to know what will happen to me in the future
19. I don’t like being taken by surprise
20. The smallest doubt can stop me from doing things
21. I should be able to prepare for everything in advance
22. Being unclear about things means that I am not confident
23. It’s not fair that other kids are surer of things
24. Not knowing what can happen keeps me from sleeping well
25. I must get away from all situations where I don’t know what will happen
26. Things that are unclear stress me
27. I don’t like being undecided about the future

Item	Parent-report form	Child-report form
1	Uncertainty stops my child from having strong opinions	Doubts stop me from having strong opinions
2	My child believes that being uncertain means one is mixed-up	Being unsure means that a person is mixed-up
3	Uncertainty makes my child's life intolerable	Not knowing what will happen in the future makes life hard
4	My child thinks it's unfair that we can't predict the future	It's not fair that we can't predict the future
5	My child's mind can't be relaxed if he/she doesn't know what will happen tomorrow	I can't relax if I don't know what will happen tomorrow

Item	Parent-report form	Child-report form
6	Uncertainty makes my child uneasy, anxious, or stressed	Not knowing what will happen in the future makes me uneasy, anxious, or stressed
7	Unforeseen events upset my child greatly	Surprise events upset me greatly
8	It frustrates my child to not have all the information he/she needs in a situation	It frustrates me to not have all of the information I need
9	Uncertainty keeps my child from living a full life	Not knowing what could happen keeps me from enjoying life
10	My child believes that one should always look ahead so as to avoid surprises	One should always think ahead to avoid surprises

Item	Parent-report form	Child-report form
11	My child believes that a small unforeseen event can spoil everything, even with the best planning	Plans can be ruined by things you didn't think would happen.
12	When it's time to act, uncertainty paralyzes my child	When it is time to do things, not knowing what could happen keeps me from acting
13	My child believes that being uncertain means that he/she is not first rate	Being unsure of things means that I am not great
14	When my child is uncertain he/she can't go forward	When I am not sure of something I can't go forward
15	When my child is uncertain he/she can't function very well	When I am not sure of something I can't work very well

Item	Parent-report form	Child-report form
16	Other children seem to be more certain than my child	Other kids have less doubts than I do
17	Uncertainty makes my child unhappy or sad	Not knowing what will happen makes me unhappy or sad
18	My child always wants to know what the future has in store for him/her	I always want to know what will happen to me in the future
19	My child can’t stand being taken by surprise	I don’t like being taken by surprise
20	The smallest doubt can stop my child from acting	The smallest doubt can stop me from doing things

Item	Parent-report form	Child-report form
21	My child feels as though he/she should be able to organize everything in advance	I should be able to prepare for everything in advance
22	My child feels as though being uncertain means that he/she lacks confidence	Being unclear about things means that I am not confident
23	My child feels as though it's unfair that other people seem to be sure about their future	It's not fair that other kids are more sure of things
24	Uncertainty keeps my child from sleeping soundly	Not knowing what can happen keeps me from sleeping well
25	My child tries to get away from all uncertain situations	I must get away from all situations where I don't know what will happen

Item	Parent-report form	Child-report form
26	The ambiguities of life stress my child	Things that are unclear stress me
27	My child can’t stand being undecided about the future.	I don’t like being undecided about the future.

Modified Abbreviated Math Anxiety Scale

(Carey, 2017)

Answer the following questions about yourself. Read the sentences at the left and think if you get nervous when those things happen to you. *Not nervous at all* means that you do not get nervous if what the sentence says happens. *Very nervous* means that you get nervous if what the sentence says happens.

Example:
In this example, the sentence is *Asking for an ice-cream.*
- If I do not get nervous at all when I ask for an ice-cream, I will click on the circle at the left (*Not nervous at all*)
- If I get very nervous when I ask for an ice-cream, I will click on the circle at the left (*Very nervous*)
- If neither I do not get it or I get nervous, I click on the circle in the middle.

Children were asked to rate all the following sentences in a 5-point Likert scale, from *Not nervous at all* to *Very nervous.*

1. Having to complete a maths worksheet by yourself
2. Thinking about a maths test the day before you use it
3. Watching the teacher work out a maths problem on the board
4. Taking a maths test
5. Being given maths homework with lots of difficult questions that you have to hand in the next day
6. Listening to the teacher talk for a long time in maths
7. Listening to another child in your class explain a maths problem
8. Finding out that you are going to have a surprise maths quiz when you start your maths lesson
9. Starting a new topic in maths

2 x 2 Achievement Goal Framework

(Elliot & McGregor, 2001)

Answer the following questions about yourself. Read the sentences at the left and think if you are like this. *Totally disagree* means that the sentence does not resemble at all how you are. *Totally agree* means that the sentence resembles a lot how you are.

Example:
In this example, the sentence is *I like ice-cream.*
- If I do not like ice-cream at all, I will click on the circle at the left (*Totally disagree*), which means something like *Not at all*
- If I like ice-cream a lot, I will click on the circle at the right (*Totally agree*), which means something like *A lot*.
- If neither I do not like it or I like it, I click on the circle in the middle.

Children were asked to rate all the following sentences in a 7-point Likert scale, from *Totally disagree* to *Totally agree*

1. It is important for me to do better than other students
2. It is important for me to do well compared to others in this class
3. My goal in this class is to get a better grade than most of the other students
4. I worry that I am not learn all that I possibly could in this class
5. Sometimes I am afraid that I may not understand the content of this class as thoroughly as I'd like
6. I am often concerned that I may not learn all that there is to learn in this class
7. I want to learn as much as possible from this class
8. It is important for me to understand the content of this course as thoroughly as possible
9. I desire to completely master the material presented in this class
10. I just want to avoid doing poorly in this class
11. My goal in this class is to avoid performing poorly
12. My fear of performing poorly in this class is often what motivates me

Perceived Competence Scale for Children

(Harter, 1982)

In these questions, you will see two types of children, one at the left and the other at the right. Think which one you resemble de most, the one at the left or the one at the right. Then, think if you resemble him/her a lot or little. If you resemble them a lot, click on the circle under *Very true for me*. If you resemble them but only a little, click on the circle under *A bit true for me.*

In this example, you will see two types of children:
- Ones like more chocolate ice-cream (written at the left part)
- The others like more vanilla ice-cream (written at the right part)

Knowing that, think Am I more like the children that like chocolate ice-cream more or the ones that like vanilla ice-cream more? If you are more like the children that prefer chocolate ice-cream, you will only care about the circles at the left (for this question). If you are more like the children who prefer vanilla ice-crem, you will only care about the circles at the right (for this question).

Please, click only one circle per line.
A question cannot have clicked answers at the left and at the right at the same time, it must be only one side.

REALLY TRUE FOR ME	SORT OF TRUE FOR ME	X	Q#	Y	SORT OF TRUE FOR ME	REALLY TRUE FOR ME
4	3	Some kids feel they are very good at their school work.	1. BUT	Other kids worry about whether they can do the school work assigned to them.	2	1
1	2	Some kids are often unhappy with themselves.	2. BUT	Other kids are pretty pleased with themselves.	3	4
4	3	Some kids feel like they are just as smart as other kids their ages.	3. BUT	Other kids aren't so sure and wonder if they are as smart.	2	1
1	2	Some kids don't like the way they are leading their life.	4. BUT	Other kids do like the way they are leading their life.	3	4
1	2	Some kids are pretty slow in finishing their school work.	5. BUT	Other kids can do their school work quickly.	3	4
4	3	Some kids are happy with themselves as a person.	6. BUT	Other kids are often not happy with themselves as a person.	2	1
REALLY TRUE FOR ME	SORT OF TRUE FOR ME	X	Q#	Y	SORT OF TRUE FOR ME	REALLY TRUE FOR ME

REALLY TRUE FOR ME	SORT OF TRUE FOR ME	X	Q#	Y	SORT OF TRUE FOR ME	REALLY TRUE FOR ME
1	2	Some kids often forget what they learn.	7. BUT	Other kids can remember things easily.	3	4
4	3	Some kids like the kind of person they are.	8. BUT	Other kids often wish they were someone else.	2	1
4	3	Some kids do very well at their classwork.	9. BUT	Other kids don't do very well at their classwork.	2	1
4	3	Some kids are very happy being the way they are.	10. BUT	Other kids wish they were different.	2	1
1	2	Some kids have trouble figuring out the answers in school.	11. BUT	Other kids almost always can figure out the answers.	3	4
1	2	Some kids are not very happy with the way they do a lot of things.	12. BUT	Other kids think the way they do things is fine.	3	4
REALLY TRUE FOR ME	SORT OF TRUE FOR ME	X	Q#	Y	SORT OF TRUE FOR ME	REALLY TRUE FOR ME

Tempo Test Automatizeren (TTA)

6	+	0	=	4	+	5	=	13	+	5	=
3	+	2	=	3	+	6	=	11	+	7	=
7	+	2	=	2	+	8	=	13	+	7	=
6	+	4	=	10	+	7	=	3	+	16	=
0	+	9	=	2	+	10	=	6	+	8	=
9	+	4	=	80	+	20	=	57	+	3	=
4	+	7	=	40	+	4	=	92	+	8	=
5	+	6	=	8	+	70	=	27	+	5	=
8	+	5	=	73	+	5	=	48	+	6	=
30	+	60	=	3	+	43	=	47	+	4	=
6	+	55	=	42	+	27	=	17	+	26	=
9	+	63	=	47	+	43	=	25	+	27	=
17	+	30	=	44	+	56	=	29	+	63	=
40	+	35	=	18	+	17	=	46	+	55	=
11	+	42	=	13	+	18	=	65	+	28	=
23	+	18	=								
32	+	59	=								
44	+	27	=								
54	+	38	=	Para, no segueixis							
29	+	28	=	Espera el senyal							

4 - 2 =	9 - 2 =	18 - 12 =
3 - 0 =	10 - 7 =	16 - 11 =
5 - 2 =	14 - 2 =	14 - 13 =
8 - 8 =	20 - 8 =	19 - 16 =
6 - 4 =	18 - 4 =	20 - 15 =

11 - 4 =	90 - 20 =	48 - 6 =
13 - 7 =	60 - 10 =	59 - 3 =
14 - 8 =	21 - 1 =	86 - 4 =
16 - 9 =	57 - 7 =	60 - 7 =
15 - 7 =	89 - 9 =	80 - 9 =

99 - 10 =	70 - 33 =	48 - 24 =
46 - 20 =	90 - 57 =	69 - 35 =
31 - 21 =	22 - 6 =	65 - 41 =
78 - 58 =	73 - 4 =	76 - 33 =
60 - 18 =	85 - 7 =	89 - 45 =

32 - 18 =

47 - 29 =

53 - 38 =

66 - 27 =

84 - 38 =

Para, no segueixis

Espera el senyal

4 x 1 =	3 x 4 =	6 x 3 =
6 x 2 =	9 x 2 =	9 x 4 =
5 x 10 =	3 x 3 =	6 x 5 =
7 x 0 =	9 x 10 =	7 x 3 =
8 x 2 =	4 x 4 =	8 x 4 =
0 x 5 =	9 x 5 =	3 x 7 =
8 x 3 =	6 x 10 =	6 x 6 =
4 x 5 =	7 x 5 =	5 x 7 =
6 x 4 =	4 x 0 =	3 x 6 =
9 x 3 =	8 x 6 =	8 x 7 =
3 x 8 =	6 x 8 =	3 x 9 =
9 x 6 =	4 x 9 =	9 x 8 =
2 x 8 =	8 x 6 =	4 x 7 =
6 x 7 =	2 x 9 =	7 x 9 =
4 x 6 =	2 x 6 =	4 x 8 =
9 x 9 =		
8 x 8 =		
9 x 7 =		
5 x 9 =	Para, no segueixis	
10 x 1 =	Espera el senyal	

Questionnaires from study in Chapter 6

Pre-test

- Which parts of this museum have you seen before today? (Mark all that apply)
    - Memorial, Interviews
    - Study on controversial images
    - This is the first part I see

- Have you visited any other location from the Jewish Cultural Quarter? (Mark all that apply)
    - Jewish Historical Museum
    - National Holocaust Museum
    - Children's Museum, Portuguese Synagogue
    - This is the first location I visit

- How many other museums, memorial or cultural institute related to the Holocaust have you visited?
    - *0,*
    - *1-3,*
    - *4 or more*

- How much do you know about the Holocaust?
    - 5-point Likert scale: *I know very little - I know a lot (expert)*

- How much do you know of the second world war?
    - 5-point Likert scale: *I know very little - I know a lot (expert)*

- How do you feel at the moment? Choose all the emotions that represent how you are feeling and click on them in the first column. Then, from those ones, click (in the second column) the ones you are feeling the most.
    - Interest
    - Compassion
    - Awe

- Annoyance
- Sadness
- Disgust
- Boredom
- Contempt
- Anger
- Shame
- Amazement
- Shock
- Fascination
- Relaxation
- Irritation
- Pride
- Love
- Confidence
- Negative surprise
- Amusement
- Delight
- Calm
- Positive surprise
- Inspiration
- Excitement
- Gratitude
- Affection
- Tense
- Despair
- Joy
- Hope
- Embarrassment
- Afraid
- Fear
- Guilt
- Sensual

- What are you looking for in this application?
    - *Only exploring,*
    - *Learning about Bergen-Belsen*
    - *Learning about the Holocaust.*

Post-test

- How do you feel at the moment? Choose all the emotions that represent how you are feeling and click on them in the first column. Then, from those ones, click (in the second column) the ones you are feeling the most.
  - Interest
  - Compassion
  - Awe
  - Annoyance
  - Sadness
  - Disgust
  - Boredom
  - Contempt
  - Anger
  - Shame
  - Amazement
  - Shock
  - Fascination
  - Relaxation
  - Irritation
  - Pride
  - Love
  - Confidence
  - Negative surprise
  - Amusement
  - Delight
  - Calm
  - Positive surprise
  - Inspiration
  - Excitement
  - Gratitude
  - Affection
  - Tense
  - Despair
  - Joy
  - Hope
  - Embarrassment
  - Afraid

- Fear
- Guilt
- Sensual

Rate the following statements from "Not at all" to "Yes, very much". Please consider some scales are inverted (going from "Yes, very much" to "Not at all").

- The presentation made me understand how Bergen-Belsen looked like
  Not at all A bit Pretty well A lot Yes, very much

- The presentation gave me an impression of the daily life in the camp
  Not at all A bit Pretty well A lot Yes, very much

- I have learnt more about concentration camps because of this presentation
  Yes, very much A lot Pretty well A bit Not at all

- The presentation gave me an impression of the how the camp operated
  Not at all A bit Pretty well A lot Yes, very much

- Because of the presentation, I can better imagine the Bergen-Belsen concentration camp
  Yes, very much A lot Pretty well A bit Not at all

- It was a boring experience
  Yes, very much A lot Pretty well A bit Not at all

- It was an exciting experience
  Not at all A bit Pretty well A lot Yes, very much

- It was an interesting experience
  Yes, very much A lot Pretty well A bit Not at all

- It was a fascinating experience

Not at all   A bit   Pretty well   A lot   Yes, very much

- It fulfilled my expectations
  Yes, very much   A lot   Pretty well   A bit   Not at all

- It was not sufficiently challenged
  Not at all   A bit   Pretty well   A lot   Yes, very much

- I thought it was innovative
  Not at all   A bit   Pretty well   A lot   Yes, very much

- I would surely recommend this presentation to my friends
  Yes, very much   A lot   Pretty well   A bit   Not at all

- Because of the presentation, I plan to find more information about Bergen-Belsen
  Yes, very much   A lot   Pretty well   A bit   Not at all

- Because of the presentation, I plan to visit more memorials and museums on this subject
  Not at all   A bit   Pretty well   A lot   Yes, very much

- It was easy for me to learn how to operate the installation
  Not at all   A bit   Pretty well   A lot   Yes, very much

- I can imagine others finding it easy to learn how to operate the installation
  Yes, very much   A lot   Pretty well   A bit   Not at all

- It was easy for me to move and navigate in the virtual environment
  Yes, very much   A lot   Pretty well   A bit   Not at all

- It was easy for me to understand the text's content
  Not at all   A bit   Pretty well   A lot   Yes, very much

- It was easy for me to understand the text in the way it was written
  Yes, very much   A lot   Pretty well   A bit   Not at all

- I got dizzy or nauseous while using the installation
  Not at all   A bit   Pretty well   A lot   Yes, very much

- How much time do you think you spent with this installation? *[Open ended text answer]*

- What type of information did you find the most interesting?
    - Children
    - Perpetrators
    - Daily Life
    - Diamond Workers Group
    - Religious Life
    - Appelplatz

- What type of information surprised you the most?
    - Children
    - Perpetrators
    - Daily Life
    - Diamond Workers Group
    - Religious Life
    - Appelplatz

- What information do you think is missing? *[Open ended text answer]*

- What functionality do you think is missing? *[Open ended text answer]*

- Did you find the added stories an important addition?
    - *Yes*
    - *No*

- How do you feel about the use of new technology in this subject? *[Open ended text answer]*

- Age *[Open ended text answer]*

- Gender

- Female
- Male
- Other

- Nationality *[Please write your nationality in the box]*

- Highest educational level achieved *[Choose from list]*

Tests and Interview from study in Chapter 7

Pre-test

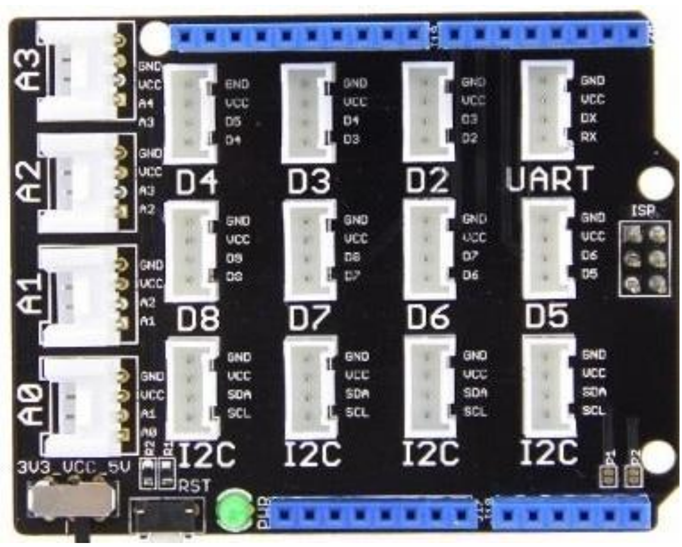
A button is…

a) Digital input
b) Analogic input
c) Digital output
d) Analogic output

How many values could an analogic output have?

a) 256
b) 255
c) 2 (high or low)
d) Infinite

Which pins could be used for digital output?



a) All white pins

b) All pins
c) From D4 to D8
d) From A0 to A3

A LED is…

a) Digital output
b) Analog or digital output
c) Analog input
d) Analog or digital input

How many values could a digital output have?

a) 256
b) 255
c) 2 (high or low)
d) Infinite

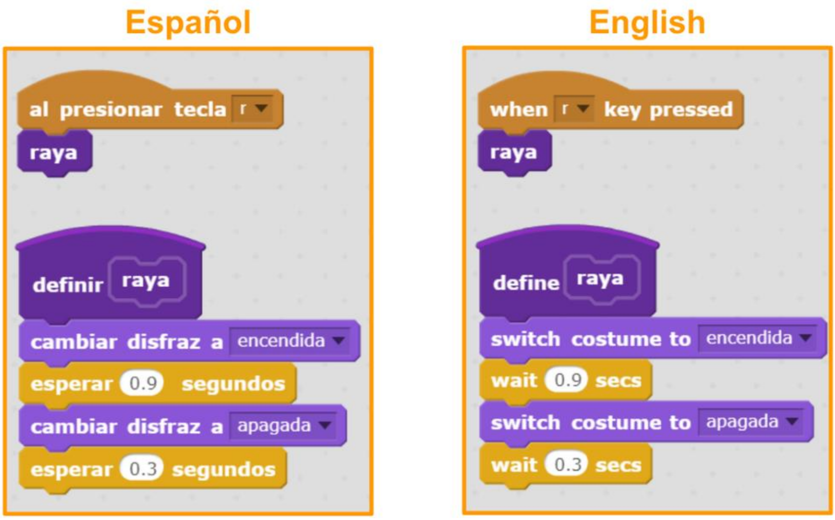
When we press the “up arrow”, the bulb…

a) It will turn on JUST for a moment
b) It will remain on
c) Nothing will happen
d) It will turn off

A potentiometer is…

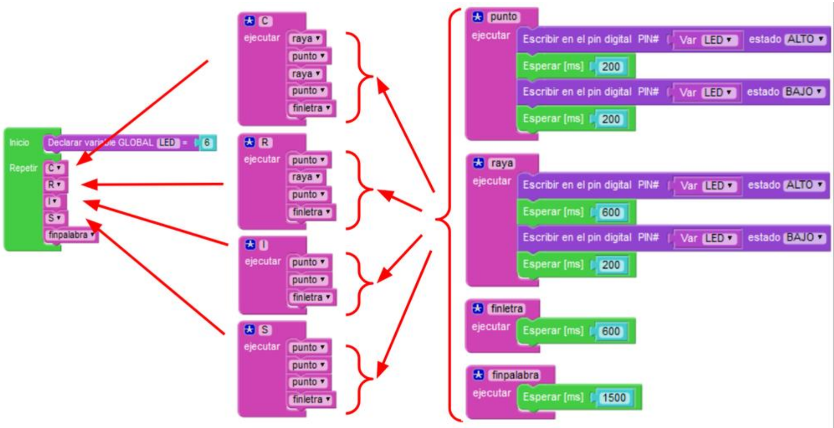
a) A digital output
b) An analogical input
c) A digital input
d) An analogical output

What is a function?



a) Only “r”
b) “Raya” and “r” are functions
c) Just “raya”
d) “Raya” and “r” are not functions

How many different functions are there?

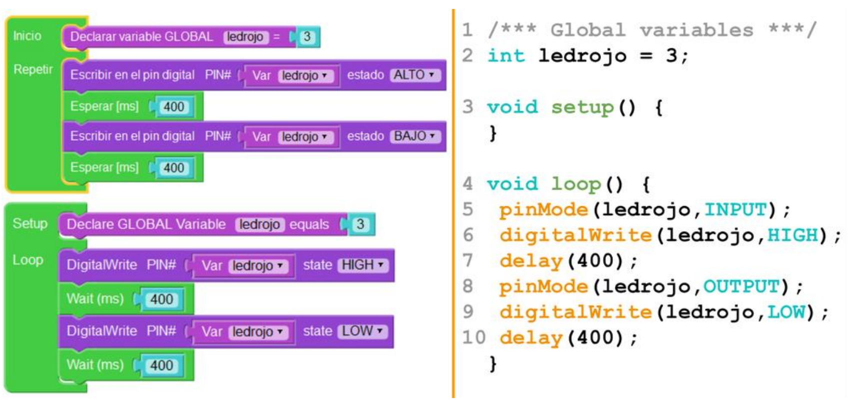


a) “C” “R” “I” “S” “finpalabra”
b) "C" "R" "I" "S" "punto" "raya" "finletra" "finpalabra"
c) "punto" "raya" "finletra"
d) "punto" "raya" "finletra" "finpalabra"

A light sensor is…

a) A digital output
b) An analogical input
c) A digital input
d) An analogical output

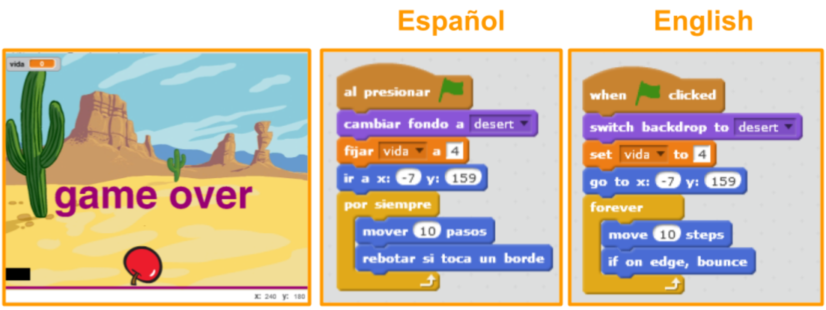
In which line of the core there is an error?



a) Line 8

b) Line 5
c) Line 6
d) Line 9

What is a variable?



a) The apple
b) Gave over
c) Vida (life)
d) The manzana and the palo negro

Post-test

Same questions of pre-test, with different question and answer order

Interview

Children were given cards (printed photos) representing each session of CREA. After making sure they understood the mapping between each session and the photo, the interview began. In it, they were asked to order them from less to more depending on the question. Then, they were asked to explain their answer. The instructions were:

- Please, order the sessions from the one you liked the most to the one you liked the less.

- Please, order the sessions from the one you find the easiest to the one you find the most difficult.

- Please, order the sessions from the one you would not like to do again to the one you would for sure do again.

- Which session would you recommend to a friend?

- Please, order the sessions, from the one you did more individually to the one you did it more collaboratively

- Is there anything that you would have liked to do but we did not?

Questionnaires and Interview from study in Chapter 8

Questionnaire before the task

- Gender
    - Female
    - Male
    - Other

- Age
- Do you usually play videogames?
    - Yes, everyday
    - Yes, every week
    - I do not play much
    - I do not play at all

Questionnaire during the task

- How engaged are you? (In all blocks)
- How predictable is the target's direction? (In all blocks)
- How predictable are the other agent's actions? (Only in the blocks where the participant plays with another player)
- How good is the collaboration with the other agent? (Only in the blocks where the participant plays with another player)

Questionnaire after the task

- Have you used any strategy during the game? If so, can you please explain it?
- Do you think there were differences between each part? If so, can you please explain them?
- Do you think you the other player was a human player or a computer?
    - Human player
    - Computer
    - I don’t know
- How predictable were all the other agents’ actions?
- How good is the collaboration with all the other agents?

Interview after the task

I will interview you now about the task you have just performed. Please, answer sincerely. There is no correct or wrong answer.

- *What do you think about the graphical aspects of the game?*
- *Has the game been difficult for you?*
- *What can you tell me about the other player?*
    - *Describe the green player*
    - *Describe the blue player*
    - *Describe the white player*
- *What do you think about the last part?*
- *How well do you think you did it?*
- *What would you improve from the game?*

When the interviewed finished, the experimenter would say to the participant: *I have finished with my questions. Do you have any question?*

BIBLIOGRAPHY

Abramson, L. Y., Seligman, M. E. P., & Teasdale, J. D. (1978). Learned Helplessness in Humans - Critique and Reformulation. *Journal of Abnormal Psychology*, *87*(1), 49–74. Retrieved from https://pdfs.semanticscholar.org/ef52/775276f83a46162a9b364335d9ee5ee73b99.pdf

Ackermann, E. (2001). Piaget’s Constructivism , Papert’s Constructionism : What’s the difference ? *Constructivism: Uses and Perspectives in Education*, 1–11. https://doi.org/10.1.1.132.4253

Aghababyan, A., Lewkow, N., & Baker, R. (2017). Exploring the asymmetry of metacognition, (November), 115–119. https://doi.org/10.1145/3027385.3027388

Aghababyan, A., Lewkow, N., & Baker, R. S. (2018). Enhancing the clustering of student performance using the variation in confidence. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *10858 LNCS*(May), 274–279. https://doi.org/10.1007/978-3-319-91464-0_27

Akinola, S. O. (2015). Computer Programming Skill and Gender Difference: An Empirical Study. *American Journal of Scientific and Industrial Research*, *7*(1), 1–6.

Aleven, V., & Koedinger, K. R. (2000). Limitations of student control: Do students know when they need help? In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 1839, pp. 292–303). https://doi.org/10.1007/3-540-45108-0_33

Aleven, V., & Koedinger, K. R. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based cognitive tutor. *New York*, *26*(2), 9–16. https://doi.org/10.1207/s15516709cog2602_1

Alhajri, R., Alhunaiyyan, A. A., & AlMousa, E. (2017). Understanding the impact of individual differences on learner performance using hypermedia systems. *International Journal of Web-Based Learning and Teaching Technologies*, *12*(1), 1–18. https://doi.org/10.4018/IJWLTT.2017010101

Alimisis, D., & Kynigos, C. (2009). *Teacher Education on Robotics-Enhanced Constructivist Pedagogical Methods*. *Teacher Education on RoboticsEnhanced Constructivist Pedagogical Methods*. Retrieved from http://www.terecop.eu/

Amaral, D. G., Schumann, C. M., & Nordahl, C. W. (2008). Neuroanatomy of autism. *Trends in Neurosciences*, *31*(3), 137–145. https://doi.org/10.1016/j.tins.2007.12.005

Ames, C. (1992). Classrooms: Goals, Structures, and Student Motivation. *Journal of Educational Psychology*, *84*(3), 261–271. https://doi.org/10.1037/0022-0663.84.3.261

Anderson, J. R., Boyle, C. F., & Reiser, B. J. (1985). Intelligent Tutoring Systems. *Science*, *228*(May 2015), 456. https://doi.org/10.1017/CBO9781107415324.004

Ansuini, C., Cavallo, A., Bertone, C., & Becchio, C. (2015). Intentions in the brain: The unveiling of Mister Hyde. *Neuroscientist*, *21*(2), 126–135. https://doi.org/10.1177/1073858414533827

Arís, N., & Orcos, L. (2019). Educational Robotics in the Stage of Secondary Education: Empirical Study on Motivation and STEM Skills. *Education Sciences*, *9*(2), 73. https://doi.org/10.3390/educsci9020073

Ashcraft, M. H. (2002). Math anxiety: Personal, educational, and cognitive consequences. *Current Directions in Psychological Science*, *11*(5), 181–185. https://doi.org/10.1111/1467-8721.00196

Atmatzidou, S., & Demetriadis, S. (2016). Advancing students' computational thinking skills through educational robotics: A study on age and gender relevant differences. *Robotics and Autonomous Systems*, *75*, 661–670. https://doi.org/10.1016/j.robot.2015.10.008

Ausubel, D. (1978). *Educational psychology : a cognitive view* (2d ed.). New York: Holt Rinehart and Winston.

Avlund, S. H., Thomsen, P. H., Schendel, D., Jørgensen, M., & Clausen, L. (2020). Time Trends in Diagnostics and Clinical Features of Young Children Referred on Suspicion of Autism: A Population-Based Clinical Cohort Study, 2000–2010. *Journal of Autism and Developmental Disorders*, 1–15. https://doi.org/10.1007/s10803-020-04555-8

Balafoutas, L., & Sutter, M. (2012). Affirmative action policies promote women and do not harm efficiency in the laboratory.

Science, *335*(6068), 579–582. https://doi.org/10.1126/science.1211180
Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.
Bandura, Albert. (1986). *Social Foundations of Thought and Action: A social Cognitive Theory*. *Contemporary Sociology* (Vol. 16). https://doi.org/10.2307/2071177
Bang, D., & Fleming, S. M. (2018). Distinct encoding of decision confidence in human medial prefrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(23), 6082–6087. https://doi.org/10.1073/pnas.1800795115
Barker, B. S., & Ansorge, J. (2017). Robotics as Means to Increase Achievement Scores in an Informal Learning Environment Robotics as Means to Increase Achievement Scores in an Informal Learning Environment. *Taylor & Francis*, *1523*(June), 229–243. https://doi.org/10.1080/15391523.2007.10782481
Baron-Cohen, S. (1997). *Mindblindness: An Essay on Autism and Theory of Mind*. Bradford Books. Retrieved from http://books.google.es/books?id=MDbcNu9zYZAC
Baron-Cohen, S., Jolliffe, T., Mortimore, C., & Robertson, M. (1997). Another advanced test of theory of mind: evidence from very high functioning adults with autism or asperger syndrome. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *38*(7), 813–822. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9363580
Barth, C. M., & Funke, J. (2010). Negative affective environments improve complex solving performance. *Cognition and Emotion*, *24*(7), 1259–1268. https://doi.org/10.1080/02699930903223766
Bauman, M. L., & Kemper, T. L. (2005). *The neurobiology of autism*. Johns Hopkins University Press. Retrieved from https://books.google.es/books?hl=en&lr=&id=s7nWIlnimY8C&oi=fnd&pg=PA121&dq=Structural+brain+anatomy+in+autism:+what+is+the+evidence%3F&ots=85tHUzFBTa&sig=HSkJJdDryrKVg_tSBY3MUrlXvKk&redir_esc=y#v=onepage&q=Structural brain anatomy in autism%3A what is the evidence%3F&f=false
Bell, M. A., & Cuevas, K. (2013). NIH Public Access, *13*(3), 1–11.

https://doi.org/10.1080/15248372.2012.691143.Using
Ben-Sasson, A., Hen, L., Fluss, R., Cermak, S. A., Engel-Yeger, B., & Gal, E. (2009). A meta-analysis of sensory modulation symptoms in individuals with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *39*(1), 1–11. https://doi.org/10.1007/s10803-008-0593-3
Berglas, S., & Jones, E. E. (1978). Drug choice as a self-handicapping strategy in response to noncontingent success. *Journal of Personality and Social Psychology*, *36*(4), 405–417. https://doi.org/10.1037/0022-3514.36.4.405
Bers, M. U. (2008). *Blocks to Robots: Learning with Technology in the Early Childhood Classroom*. Teachers College Press.
Bers, M. U. (2017). *Coding as a playground: Programming and computational thinking in the early childhood classroom*. Routledge. https://doi.org/10.4324/9781315398945
Bers, Marina U., Ponte, I., Juelich, K., & Schenker, J. (2002). Teachers as designers: Integrating robotics in early childhood education. *Information Technology in Childhood Education Annual*, *2002*, 123–145. https://doi.org/10.1992/pssb.200301795
Betz, N. E. (1978). Prevalence, distribution, and correlates of math anxiety in college students. *Journal of Counseling Psychology*, *25*(5), 441–448. https://doi.org/10.1037/0022-0167.25.5.441
Beyer, S. (2015). Gender Differences in the Recall of Positive and Negative Performance Feedback, (May 2003).
Biemmi, I., & Satta, C. (2017). Childhood, Education and Gender. The construction of gender culture in classroom, after-school and family settings Editorial 1, *6*. https://doi.org/10.15167/2279-5057/AG2017.6.12.491
Bjorklund, D. F. (1997). The role of immaturity in human development. *Psychological Bulletin*, *122*(2), 153–169. https://doi.org/10.1037/0033-2909.122.2.153
Bobadilla, J., Ortega, F., Hernando, A., & Gutiérrez, A. (2013). Recommender systems survey. *Knowledge-Based Systems*, *46*, 109–132. https://doi.org/10.1016/j.knosys.2013.03.012
Boekaerts, M., & Rozendaal, J. S. (2010). Using multiple calibration indices in order to capture the complex picture of what affects students' accuracy of feeling of confidence. *Learning and Instruction*, *20*(5), 372–382.

https://doi.org/10.1016/j.learninstruc.2009.03.002
Bol, L., & Hacker, D. J. (2001). A comparison of the effects of practice tests and traditional review on performance and calibration. *Journal of Experimental Education*, *69*(2), 133–151. https://doi.org/10.1080/00220970109600653
Boldt, A., Blundell, C., & De Martino, B. (2019). Confidence modulates exploration and exploitation in value-based learning. *Neuroscience of Consciousness*, *2019*(1). https://doi.org/10.1093/nc/niz004
Bonawitz, E., Shafto, P., Gweon, H., Goodman, N. D., Spelke, E., & Schulz, L. (2011). The double-edged sword of pedagogy: Instruction limits spontaneous exploration and discovery. *Cognition*, *120*(3), 322–330. https://doi.org/10.1016/j.cognition.2010.10.001
Borràs, J., Moreno, A., & Valls, A. (2014). Intelligent tourism recommender systems: A survey. *Expert Systems with Applications*, *41*(16), 7370–7389. https://doi.org/10.1016/j.eswa.2014.06.007
Botelho, S. S. C., Braz, L. G., & Rodrigues, R. N. (2012). Exploring Creativity and Sociability with an Accessible Educational Robotic Kit. In *Proceedings 3rd International Conference on Robotics in Education (RiE 2012)* (pp. 55–60).
Boxtel, C. Van, & Drie, J. Van. (2004). Novices Contextualise Historical Sources, (1999), 89–97.
Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How People Learn: Brain, Mind, Experience, and School*. *Committee on learning research and educational practice* (Vol. Expanded E). https://doi.org/10.1016/0885-2014(91)90049-J
Breazeal, C. (2004). Designing Sociable Robots. In *Designing Sociable Robots* (p. 263). MIT Press. https://doi.org/10.7551/mitpress/2376.003.0007
Brickman, P., Gormally, C., Armstrong, N., & Hallar, B. (2009). Effects of inquiry-based learning on students’ science literacy skills and confidence. *International Journal for the Scholarship of Teaching and Learning*, *3*(2), 1–22. https://doi.org/10.1016/j.iheduc.2009.11.003
Brooks, J. J. G., & Brooks, M. G. (1999). In Search of Understanding: The Case for Constructivist Classrooms. *Association for Supervision and Curriculum Development*, 1–

136. Retrieved from http://www.ascd.org
Brown, A. L., Bransford, J. D., Ferrar, R. A., & Campione, J. C. (1983). Learning, Remembering, and Understanding. In P. H. Mussen (Ed.), *Handbook of Child Psychology* (pp. 77–166). New York: John Wiley.
Bruner, J. (1961). The Act of Discovery. *Harvard Educational Review*. Retrieved from https://psycnet.apa.org/record/1962-00777-001
Bruner, J. (1978). The role of dialogue in language acquisition. In A. Sinclair, J. Jarvelle, R., & W. J. M. Levelt (Eds.), *The Child's Concept of Language*. New York: Springer-Verlag.
Bryman, A. (2006). Integrating quantitative and qualitative research: How is it done? *Qualitative Research*, *6*(1), 97–113. https://doi.org/10.1177/1468794106058877
Butler, A. C., Karpicke, J. D., & Roediger, H. L. (2008). Correcting a Metacognitive Error: Feedback Increases Retention of Low-Confidence Correct Responses. *Journal of Experimental Psychology: Learning Memory and Cognition*, *34*(4), 918–928. https://doi.org/10.1037/0278-7393.34.4.918
Butler, A. C., Karpicke, J. D., & Roediger III, H. L. (2008). Correcting a Metacognitive Error: Feedback Increases Retention of Low-Confidence Correct Responses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(4), 918–928. https://doi.org/10.1037/0278-7393.34.4.918
Butterfield, B., & Metcalfe, J. (2001). Errors committed with high confidence are hypercorrected. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(6), 1491–1494. https://doi.org/10.1037/0278-7393.27.6.1491
Carey, E., Hill, F., Devine, A., & Szucs, D. (2017). The modified abbreviated math anxiety scale: A valid and reliable instrument for use with children. *Frontiers in Psychology*, *8*(JAN), 1–13. https://doi.org/10.3389/fpsyg.2017.00011
Carey, E., Hill, F., Devine, A., & Szücs, D. (2016). The chicken or the egg? The direction of the relationship between mathematics anxiety and mathematics performance. *Frontiers in Psychology*, *6*(JAN), 1–6. https://doi.org/10.3389/fpsyg.2015.01987
Carruthers, P. (2009). How we know our own minds: The relationship between mindreading and metacognition. *Behavioral and Brain Sciences*, *32*(2), 121–182.

https://doi.org/10.1017/S0140525X09000545
Cascallar, E., Boekaerts, M., & Costigan, T. (2006). Assessment in the Evaluation of Self-Regulation as a Process, 297–306. https://doi.org/10.1007/s10648-006-9023-2
Cates, G., & Rhymer, K. (2003). Examining the Relationship Between Mathematics Anxiety and Mathematics Performance: An Instructional Hierarchy Perspective. *Journal of Behavioral Education*, *12*(1), 23–34. https://doi.org/10.1023/A:1022318321416
Chambers, J., Carbonaro, M., & Murray, H. (2008). Developing conceptual understanding of mechanical advantage through …. *Australasian Journal of …*, *24*(4), 387–401. https://doi.org/10.14742/ajet.1199
Chambon, V., Farrer, C., Pacherie, E., Jacquet, P. O., Leboyer, M., & Zalla, T. (2017). Reduced sensitivity to social priors during action prediction in adults with autism spectrum disorders. *Cognition*, *160*, 17–26. https://doi.org/10.1016/j.cognition.2016.12.005
Chan, J. C. Y., & Lam, S. fong. (2010). Effects of different evaluative feedback on students' self-efficacy in learning. *Instructional Science*, *38*(1), 37–58. https://doi.org/10.1007/s11251-008-9077-2
Chatard, A., Guimond, S., & Selimbegovic, L. (2007). "How good are you in math?" The eVect of gender stereotypes on students' recollection of their school marks. *Journal of Experimental Social Psychology*, *43*, 1017–1024. https://doi.org/10.1016/j.jesp.2006.10.024
Chaudhary, V., Agrawal, V., Sureka, P., & Sureka, A. (2017). An Experience Report on Teaching Programming and Computational Thinking to Elementary Level Children Using Lego Robotics Education Kit. In *Proceedings - IEEE 8th International Conference on Technology for Education, T4E 2016* (pp. 38–41). IEEE. https://doi.org/10.1109/T4E.2016.016
Chi, M. T. H. (2000). Self-explaining: The dual processes of generating inferences and repairing mental models. *Advances in Instructional Psychology*.
Choi, S., & Clark, R. E. (2006). Cognitive and affective benefits of an animated pedagogical agent for learning english as a second language. *Journal of Educational Computing*

Research. https://doi.org/10.2190/A064-U776-4208-N145
Chrastil, E. R., & Warren, W. H. (2012). Active and passive contributions to spatial learning. *Psychonomic Bulletin & Review*, *19*(1), 1–23.
Chukoskie, L., Townsend, J., & Westerfield, M. (2013). *Motor skill in autism spectrum disorders: A subcortical view. International Review of Neurobiology* (1st ed., Vol. 113). Elsevier Inc. https://doi.org/10.1016/B978-0-12-418700-9.00007-1
Church, M. A., Elliot, A. J., & Gable, S. L. (2001). Perceptions of classroom environment, achievement goals, and achievement outcomes. *Journal of Educational Psychology*, *93*(1), 43–54. https://doi.org/10.1037/0022-0663.93.1.43
Cockburn, A., & Bryant, A. (1997). Leogo: An equal opportunity user interface for programming. *Journal of Visual Languages and Computing*, *8*(5–6), 601–619. https://doi.org/10.1006/jvlc.1997.0152
Comer, J. S., Roy, A. K., Furr, J. M., Gotimer, K., Beidas, R. S., Dugas, M. J., & Kendall, P. C. (2009). The Intolerance of Uncertainty Scale for Children: A Psychometric Evaluation. *Psychological Assessment*, *21*(3), 402–411. https://doi.org/10.1037/a0016719
Conati, C., & Zhou, X. (2002). Modeling Students' Emotions from Cognitive Appraisal in Educational Games (pp. 944–954). https://doi.org/10.1007/3-540-47987-2_94
Cook, J. L., Blakemore, S. J., & Press, C. (2013). Atypical basic movement kinematics in autism spectrum conditions. *Brain*, *136*(9), 2816–2824. https://doi.org/10.1093/brain/awt208
Cooper, S., Dann, W., & Pausch, R. (2000). Alice: a 3-D tool for introductory programming concepts. In *5th Annual CCSC Northeastern Conference on the Journal of Computing in Small Colleges* (pp. 107–116). https://doi.org/10.1145/1953163.1953243
Copeland, T. (2006). Constructing Pasts: Interpreting the Historic Environment. In A. Hems & M. R. Blockley (Eds.), *Heritage interpretation* (pp. 83–95). London, New York: Routledge. Retrieved from http://eprints.glos.ac.uk/55/
Corbett, A. T., Anderson, J. R., Carver, V. H., & Brancolini, S. A. (1994). Individual differences and predictive validity in student modeling. Retrieved from

https://philpapers.org/rec/CORIDA-4
Council, N. R. (2002). *Learning and Understanding*. *Learning and Understanding*. National Academies Press. https://doi.org/10.17226/10129
Courchesne, E., Redcay, E., Morgan, J. T., & Kennedy, D. P. (2005). Autism at the beginning: Microstructural and growth abnormalities underlying the cognitive and behavioral phenotype of autism. *Development and Psychopathology*, *17*(3), 577–597. https://doi.org/10.1017/S0954579405050285
Cross, C. P., Copping, L. T., & Campbell, A. (2014). Sex differences in impulsivity: a meta-analysis. *Igarss 2014*, (1), 1–5. https://doi.org/10.1007/s13398-014-0173-7.2
Cvencek, D., Meltzoff, A. N., & Greenwald, A. G. (n.d.). Math-Gender Stereotypes in Elementary School Children. https://doi.org/10.1111/j.1467-8624.2010.01529.x
D’Cruz, A.-M., Ragozzino, M. E., Mosconi, M. W., Shrestha, S., Cook, E. H., & Sweeney, J. A. (2013). Reduced behavioral flexibility in autism spectrum disorders. *Neuropsychology*, *27*(2), 152–160. https://doi.org/10.1037/a0031721
D’Mello, S., Lehman, B., Pekrun, R., & Graesser, A. (2014). Confusion can be beneficial for learning. *Learning and Instruction*, *29*, 153–170. https://doi.org/10.1016/j.learninstruc.2012.05.003
De Jager, B., Jansen, M., & Reezigt, G. (2005). The development of metacognition in primary school learning environments. *School Effectiveness and School Improvement*, *16*(2), 179–196. https://doi.org/10.1080/09243450500114181
De Vos, T. (2010). *Manual tempotoets automatiseren*. Amsterdam: Boom Testuitgevers.
Deaux, K., & LaFrance, M. (1998). Gender. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The Handbook of Social Psychology* (pp. 788–827). New York: Mc Graw-Hill. Retrieved from http://web.mit.edu/curhan/www/docs/Articles/15341_Readings/Motivation/Pittman_1998_Motivation.pdf
Deshmukh, A., Castellano, G., Kappas, A., Barendregt, W., Nabais, F., Paiva, A., … Aylett, R. (2013). Towards empathic artificial tutors. *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 113–114. https://doi.org/10.1109/HRI.2013.6483527

Desoete, A., & De Craene, B. (2019). Metacognition and mathematics education: an overview. *ZDM - Mathematics Education*, *51*(4), 565–575. https://doi.org/10.1007/s11858-019-01060-w
Destan, N., & Roebers, C. M. (2015). What are the metacognitive costs of young children's overconfidence? *Metacognition and Learning*, *10*(3), 347–374. https://doi.org/10.1007/s11409-014-9133-z
Detsikas, N., & Alimisis, D. (2011). Status and trends in educational robotics worldwide with special consideration of educational experiences from Greek schools. In D. Bezakova & I. Kalas (Eds.), *Proceedings of the International Conference on Informatics in Schools: Situation, Evolution and Perspectives* (pp. 1–12).
Dewey, J. (1902). *The child and the curriculum*. Chicago: University of Chicago.
Dewey, J. (1938). Experience & Education.
Di Martino, A., Kelly, C., Grzadzinski, R., Zuo, X. N., Mennes, M., Mairena, M. A., … Milham, M. P. (2011). Aberrant striatal functional connectivity in children with autism. *Biological Psychiatry*, *69*(9), 847–856. https://doi.org/10.1016/j.biopsych.2010.10.029
Dickhäuser, O., & Meyer, W.-U. (2006). Gender differences in young children's math ability attributions. *Psychology Science*, *48*(1), 3–16. Retrieved from https://www.researchgate.net/profile/Wulf_Uwe_Meyer/publication/26514479_Gender_differences_in_young_children%27s_math_ability_attributions/links/575d6a6808aec91374aecaeb/Gender-differences-in-young-childrens-math-ability-attributions.pdf
Ding, Y., Wang, E., Zou, Y., Song, Y., Xiao, X., Huang, W., & Li, Y. (2017). Gender differences in reward and punishment for monetary and social feedback in children: An ERP study. *PLoS ONE*, *12*(3), e0174100. https://doi.org/10.1371/journal.pone.0174100
Dugas, M. J., Buhr, K., & Ladouceur, R. (2004). The Role of Intolerance of Uncertainty in Etiology and Maintenance. In *Generalized anxiety disorder: Advances in research and practice.* (pp. 143–163).
Dugas, M. J., & Koerner, N. (2005). Cognitive-behavioral

treatment for generalized anxiety disorder: Current status and future directions. *Journal of Cognitive Psychotherapy*. https://doi.org/10.1891/jcop.19.1.61.66326
Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self evaluations undermine students' learning and retention. *Learning and Instruction*, *22*(4), 271–280. https://doi.org/10.1016/j.learninstruc.2011.08.003
Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed Self-Assessment. *Psychological Science in the Public Interest*, *5*(3), 69–106. https://doi.org/10.1111/j.1529-1006.2004.00018.x
Dunstone, J., & Caldwell, C. A. (2018). Cumulative culture and explicit metacognition: a review of theories, evidence and key predictions. *Palgrave Communications*, *4*(1), 1–11. https://doi.org/10.1057/s41599-018-0200-y
Dweck, C. (2008). *Mindsets and math/science achievement*. Retrieved from www.opportunityequation.org
Dweck, C. S., & Henderson, V. L. (1989). *Theories of Intelligence: Background and Measures.*
Edelson, M. G., Polania, R., Ruff, C. C., Fehr, E., & Hare, T. A. (2018). Computational and neurobiological foundations of leadership decisions. *Science*, *361*(6401). https://doi.org/10.1126/science.aat0036
Efklides, A. (2006). Metacognition and affect: What can metacognitive experiences tell us about the learning process? *Educational Research Review*, *1*(1), 3–14. https://doi.org/10.1016/j.edurev.2005.11.001
Efklides, A. (2011). Interactions of metacognition with motivation and affect in self-regulated learning: The MASRL model. *Educational Psychologist*, *46*(1), 6–25. https://doi.org/10.1080/00461520.2011.538645
Eguchi, A. (2010). What is Educational Robotics? Theories behind it and practical implementation. In *Proceedings of SITE 2010--Society for Information Technology & Teacher Education International Conference* (pp. 4006–4014).
Ehrlinger, J., Johnson, K., Banner, M., Dunning, D., & Kruger, J. (2008). Why the Unskilled Are Unaware: Further Explorations of (Absent) Self-Insight Among the Incompetent. *Organ Behav Hum Decis Process*, *105*(1), 98–121.

https://doi.org/10.1016/j.obhdp.2007.05.002.Why
El-Hakim, S. F., Beraldin, J.-A., Picard, M., & Godin, G. (2009). Detailed 3D Reconstruction of Large-Scale Heritage Sites with Integrated Techniques. *IEEE Computer Graphics and Applications*, 21–38. https://doi.org/10.1016/b978-1-84334-562-6.50002-5
Elliot, A. J., & McGregor, H. A. (2001). A 2 × 2 achievement goal framework. *Journal of Personality and Social Psychology*, *80*(3), 501–519. https://doi.org/10.1037/0022-3514.80.3.501
Else-quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-National Patterns of Gender Differences in Mathematics : A Meta-Analysis, *136*(1), 103–127. https://doi.org/10.1037/a0018053
Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-National Patterns of Gender Differences in Mathematics: A Meta-Analysis. *Psychological Bulletin*, *136*(1), 103–127. https://doi.org/10.1037/a0018053
Estes, A., Shaw, D. W. W., Sparks, B. F., Friedman, S., Giedd, J. N., Dawson, G., … Dager, S. R. (2011). Basal ganglia morphometry and repetitive behavior in young children with autism spectrum disorder. *Autism Research*, *4*(3), 212–220. https://doi.org/10.1002/aur.193
Fast, N. J., Sivanathan, N., Mayer, N. D., & Galinsky, A. D. (2012). Power and overconfident decision-making. *Organizational Behavior and Human Decision Processes*, *117*(2), 249–260. https://doi.org/10.1016/j.obhdp.2011.11.009
Feick, D. L., & Rhodewalt, F. (1997). The double-edged sword of self-handicapping: Discounting, augmentation, and the protection and enhancement of self-esteem. *Motivation and Emotion*, *21*(2), 147–163. https://doi.org/10.1023/A:1024434600296
Felder, R., & Brent, R. (2005). Understanding student differences. *Journal of Engineering Education*, *94*(1), 57–72. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/j.2168-9830.2005.tb00829.x/abstract
Fetsco, Thomas, J. M. (2005). *EDUCATIONAL PSYCHOLOGY : An Integrated Approach To Classroom Decisions*. Allyn and Bacon. Retrieved from https://books.google.es/books/about/Educational_Psychology.html?id=2QNKAAAAYAAJ&redir_esc=y

Fimbel, E. J., Michaud, R., & Martin, M. (2009). Certainty in Categorical Judgment of Size. *PLoS ONE*, *4*(7), e6198. https://doi.org/10.1371/journal.pone.0006198

Fischhoff, B. (2013). *Judgment and decision making*. *Judgment and Decision Making*. Prentice Hall. https://doi.org/10.4324/9780203141939

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, *34*(10), 906–911. Retrieved from http://content.apa.org/journals/amp/34/10/906

Flavell, J. H., Miller, P. H., & Miller, S. A. (2002). *Flavell, Miller & Miller, Cognitive Development, 4th Edition* . Pearson. Retrieved from https://www.pearson.com/us/higher-education/program/Flavell-Cognitive-Development-4th-Edition/PGM180330.html

Flavell, J. H., & Wellman, H. M. (1977). Metamemory. In R. V. Kail & J. W. Hagen (Eds.), *Perspectives on the development of memory and cognition* (pp. 3–33). Hillsdale Erlbaum.

Fleming, S. M., & Bang, D. (2018). Shouldering responsibility. *Science*, *361*(6401), 449–450. https://doi.org/10.1016/j.biopsych.2017.12.017

Fleming, S. M., Weil, R. S., Nagy, Z., Dolan, R. J., & Rees, G. (2010). Relating introspective accuracy to individual differences in brain structure. *Science (New York, N.Y.)*, *329*(5998), 1541–1543. https://doi.org/10.1126/science.1191883

Flot, J., McKenna, J., & Shoop, R. (2016). Helping Students Build Conceptual Models – the Lost Manual. Carnegie Mellon Robotics Academy, Pittsburgh, PA.

Fosnot, C. T., & Perry, R. S. (2005). Constructivism: A Psychological Theory of Learning. In *Constructivism: Theory, perspectives, and practice*.

Foster, C. (2016). Confidence and competence with mathematical procedures. *Educational Studies in Mathematics*, *91*(2), 271–288. https://doi.org/10.1007/s10649-015-9660-9

Frangou, S., & Papanikolaou, K. (2008). Representative examples of implementing educational robotics in school based on the constructivist approach. *International Conference on Simulation, Modeling and Programming for Autonomous Robots*, 54–65.

Frenzel, A. C., Pekrun, R., & Goetz, T. (2007). Girls and mathematics - A “hopeless” issue? A control-value approach to gender differences in emotions towards mathematics. *European Journal of Psychology of Education*, *22*(4), 497–514. https://doi.org/10.1007/BF03173468

Furnham, A., Chamorro-Premuzic, T., & McDougall, F. (2002). Personality, cognitive ability, and beliefs about intelligence as predictors of academic performance. *Learning and Individual Differences*, *14*(1), 47–64. https://doi.org/10.1016/j.lindif.2003.08.002

Garcia, T., & Pintrich, P. R. (1994). Regulating motivation and cognition in the classroom: The role of self-schemas and self-regulatory strategies. *Self-Regulation of Learning and Performance: Issues and Educational Applications*, (January 1994), 127–153. https://doi.org/10.1017/CBO9781107415324.004

Gaudiello, I., & Zibetti, E. (2013). *Using control heuristics as a means to explore the educational potential of robotics kits*. *Themes in Science & Technology Education* (Vol. 6).

Geary, D. N., McLachlan, G. J., & Basford, K. E. (1989). Mixture Models: Inference and Applications to Clustering. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, *152*(1), 126. https://doi.org/10.2307/2982840

Geitz, G., Brinke, D. J. Ten, & Kirschner, P. a. (2016). Changing learning behaviour: Self-efficacy and goal orientation in PBL groups in higher education. *International Journal of Educational Research*, *75*, 146–158. https://doi.org/10.1016/j.ijer.2015.11.001

Georghiades, P. (2004). From the general to the situated: Three decades of metacognition. *International Journal of Science Education*. https://doi.org/10.1080/0950069032000119401

Gigerenzer, G., Hoffrage, U., & Kleinbolting, H. (1991). Probabilistic mental models: A brunswikian theory of confidence. *Psychological Review*, *98*(4), 506–528. https://doi.org/10.1037/0033-295X.98.4.506

Gignac, G. E., & Zajenkowski, M. (2020). The Dunning-Kruger effect is (mostly) a statistical artefact: Valid approaches to testing the hypothesis with individual differences data. https://doi.org/10.1016/j.intell.2020.101449

Glickstein, M. (1998). *Cerebellum and the sensory guidance of*

movement. (G. R. Bock & J. A. Goode, Eds.). Chichester, UK: John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470515563
Gneezy, U., Niederle, M., & Rustichini, A. (2003). Performance in competitive environments: Gender differences. *Quarterly Journal of Economics*. https://doi.org/10.1162/00335530360698496
Goetz, T., Bieg, M., Lüdtke, O., Pekrun, R., & Hall, N. C. (2013). Do Girls Really Experience More Anxiety in Mathematics? *Psychological Science*, *24*(10), 2079–2087. https://doi.org/10.1177/0956797613486989
Gomoll, A., Hmelo-Silver, C. E., Šabanović, S., & Francisco, M. (2016). Dragons, Ladybugs, and Softballs: Girls' STEM Engagement with Human-Centered Robotics. *Journal of Science Education and Technology*, *25*(6), 899–914. https://doi.org/10.1007/s10956-016-9647-z
Good, C. (2014). Building Bridges to Belonging: Mindsets that Reduce Stereotype Threat and Increase Participation, Achievement, and Learning in STEM. *Lletin of the American Physical Society*, *59*.
Goupil, L., Romand-Monnier, M., & Kouider, S. (2016). Infants ask for help when they know they don't know. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(13), 3492–3496. https://doi.org/10.1073/pnas.1515129113
Gowen, E., & Hamilton, A. (2013). Motor abilities in autism: A review using a computational context. *Journal of Autism and Developmental Disorders*, *43*(2), 323–344. https://doi.org/10.1007/s10803-012-1574-0
Grainger, C., Williams, D. M., & Lind, S. E. (2016). Metacognitive monitoring and control processes in children with autism spectrum disorder: Diminished judgement of confidence accuracy. *Consciousness and Cognition*, *42*, 65–74. https://doi.org/10.1016/j.concog.2016.03.003
Grant, P., & Basye, D. (2014). *Personalized learning. A Guide for Engaging Students with Technology*. Retrieved from https://books.google.es/books?hl=es&lr=&id=96apCgAAQBAJ&oi=fnd&pg=PA2&dq=Grant+and+Basye,+2014&ots=wjRSjdSvhM&sig=wImXZ6zvMhFPRlfz1NRjLTrMNnI&redir_esc=y#v=onepage&q=Grant and Basye%2C 2014&f=false

Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey Methodology, 2nd Edition | Wiley*. Wiley. Retrieved from https://www.wiley.com/en-us/Survey+Methodology%2C+2nd+Edition-p-9780470465462

Guay, F., Ratelle, C. F., Roy, A., & Litalien, D. (2010). Academic self-concept, autonomous academic motivation, and academic achievement: Mediating and additive effects. *Learning and Individual Differences*, *20*(6), 644–653. https://doi.org/10.1016/j.lindif.2010.08.001

Gura, M., & King, K. P. (2007). *Classroom robotics : case stories of 21st century instruction for millennial students*. IAP.

Guzdial, M. (2008). Paving the way for computational thinking. *Communications of the ACM*, *51*(8), 25. https://doi.org/10.1145/1378704.1378713

Hacker, D. J., Bol, L., Horgan, D. D., & Rakow, E. A. (2000). Test prediction and performance in a classroom context. *Journal of Educational Psychology*, *92*(1), 160–170. https://doi.org/10.1037/0022-0663.92.1.160

Hacker, D. J., Bol, L., & Keener, M. C. (2008). Metacognition in Education: A focus in callibration. In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of Metamemory and Memory* (pp. 429–455). Psychology Press.

Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S., & Gernsbacher, M. A. (2007). The science of sex differences in science and mathematics. *Psychological Science in the Public Interest, Supplement*, *8*(1), 1–51. https://doi.org/10.1111/j.1529-1006.2007.00032.x

Hamilton, A. F. d. C., Brindley, R. M., & Frith, U. (2007). Imitation and action understanding in autistic spectrum disorders: How valid is the hypothesis of a deficit in the mirror neuron system? *Neuropsychologia*, *45*(8), 1859–1868. https://doi.org/10.1016/j.neuropsychologia.2006.11.022

Händel, M., & Fritzsche, E. S. (2016). Unskilled but subjectively aware: Metacognitive monitoring ability and respective awareness in low-performing students. *Memory and Cognition*. https://doi.org/10.3758/s13421-015-0552-0

Harackiewicz, J. M., Barron, K. E., Pintrich, P. R., Elliot, A. J., & Thrash, T. M. (2002). Revision of achievement goal theory:

Necessary and illuminating. *Journal of Educational Psychology*, *94*(3), 638–645. https://doi.org/10.1037/0022-0663.94.3.638
Hardiman, P. T., Pollatsek, A., & Well, A. D. (1986). Learning to Understand the Balance Beam. *Cognition and Instruction*, *3*(1), 63–86. https://doi.org/10.1207/s1532690xci0301_3
Harland, T. (2012). *University Teaching*. *University Teaching*. Routledge. https://doi.org/10.4324/9780203120606
Harris, P. L., & Lane, J. D. (2014). Infants Understand How Testimony Works. *Topoi*, *33*(2), 443–458. https://doi.org/10.1007/s11245-013-9180-0
Harter, S. (1978). Effectance Motivation Reconsidered Toward a Developmental Model. *Human Development*, *21*(1), 34–64. https://doi.org/10.1159/000271574
Harter, S. (1982). The Perceived Competence Scale for Children. *Child Development*, *53*(1), 87. https://doi.org/10.2307/1129640
Hartley, B. L., & Sutton, R. M. (n.d.). A Stereotype Threat Account of Boys' Academic Underachievement. *Child Development*, *85*(5), 1716–1733. https://doi.org/10.1111/cdev.12079
Hartwig, M. K., & Dunlosky, J. (2014). The contribution of judgment scale to the unskilled-and-unaware phenomenon: How evaluating others can exaggerate over- (and under-) confidence. *Memory and Cognition*, *42*(1), 164–173. https://doi.org/10.3758/s13421-013-0351-4
Hashimoto, T., Tayama, M., Murakawa, K., Yoshimoto, T., Miyazaki, M., Harada, M., & Kuroda, Y. (1995). Development of the brainstem and cerebellum in autistic patients. *Journal of Autism and Developmental Disorders*, *25*(1), 1–18. https://doi.org/10.1007/BF02178163
Hassenzahl, M., & Ullrich, D. (2007). To do or not to do: Differences in user experience and retrospective judgments depending on the presence or absence of instrumental goals. *Interacting with Computers*, *19*(4), 429–437. https://doi.org/10.1016/j.intcom.2007.05.001
Haswell, C. C., Izawa, J., R Dowell, L., H Mostofsky, S., & Shadmehr, R. (2009). Representation of internal models of action in the autistic brain. *Nature Neuroscience*, *12*(8), 970–972. https://doi.org/10.1038/nn.2356
Hattie, B. J. (n.d.). Visible Learning: A Synthesis of over 800

Meta-Analyses Relating to Achievement.
Haun, D. E., Zeringue, A., Leach, A., & Foley, A. (2000). Assessing the Competence of Specimen-Processing Personnel. *Laboratory Medicine*, *31*(11), 633–637. Retrieved from https://academic.oup.com/labmed/article/31/11/633/2657053
Henmon, V. A. C. (1911). The relation of the time of a judgement to its accuracy. *Psychological Review*, *18*(3), 186–201. https://doi.org/10.1037/h0074579
Heyes, C. (2015). Who Knows? Metacognitive Social Learning Strategies. https://doi.org/10.1016/j.tics.2015.12.007
Heyes, C., Bang, D., Shea, N., Frith, C. D., & Fleming, S. M. (2020, May 1). Knowing Ourselves Together: The Cultural Origins of Metacognition. *Trends in Cognitive Sciences*. Elsevier Ltd. https://doi.org/10.1016/j.tics.2020.02.007
Hilton, C., Graver, K., & LaVesser, P. (2007). Relationship between social competence and sensory processing in children with high functioning autism spectrum disorders. *Research in Autism Spectrum Disorders*, *1*(2), 164–173. https://doi.org/10.1016/j.rasd.2006.10.002
Hirt, E. R., McCrea, S. M., & Boris, H. I. (2003). "I know you self-handicapped last exam": gender differences in reactions to self-handicapping. *Journal of Personality and Social Psychology*, *84*(1), 177–193. https://doi.org/10.1037/0022-3514.84.1.177
Ho, H. Z., Senturk, D., Lam, A. G., Zimmer, J. M., Hong, S., Okamoto, Y., … Wang, C. P. (2000). The affective and cognitive dimensions of math anxiety: A cross-national study. *Journal for Research in Mathematics Education*, *31*(3), 362–379. https://doi.org/10.2307/749811
Huang, L., Varnado, T., & Gillan, D. (2013). An exploration of robot builders' attachment to their LEGO robots. In *Proceedings of the Human Factors and Ergonomics Society* (Vol. 57, pp. 1825–1829). SAGE PublicationsSage CA: Los Angeles, CA. https://doi.org/10.1177/1541931213571408
Huang, L., Varnado, T., & Gillan, D. (2014). Exploring Reflection Journals and Self-efficacy in Robotics Education. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *58*(1), 1939–1943. https://doi.org/10.1177/1541931214581405
Hughes, C. (1996). Brief report: Planning problems in autism at the

level of motor control. *Journal of Autism and Developmental Disorders*, *26*(1), 99–107. https://doi.org/10.1007/BF02276237

Huijgen, T., & Holthuis, P. (2015). “Why am I accused of being a heretic?” A pedagogical framework for stimulating historical contextualisation. *Teaching History*, (January).

Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., & Williams, C. C. (2008). Diversity: Gender similarities characterize math performance. *Science*. https://doi.org/10.1126/science.1160364

Igbo, J. N., Onu, V. C., & Obiyo, N. O. (2015). Impact of Gender Stereotype on Secondary School Students’ Self-Concept and Academic Achievement. *SAGE Open*, *5*(1), 215824401557393. https://doi.org/10.1177/2158244015573934

Im, T. (2013). *The effects of emotional support and cognitive motivational messages on math anxiety, self-efficacy, and math problem solving. Dissertation Abstracts International*. Retrieved from http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=psyc10&AN=2013-99151-230%5Cnhttp://oxfordsfx.hosted.exlibrisgroup.com/oxford?sid=OVID:psycdb&id=pmid:&id=doi:&issn=0419-4209&isbn=9781267642752&volume=74&issue=2-A(E)&spage=No&pages=No+P

Insabato, A., Pannunzi, M., Rolls, E. T., & Deco, G. (2010). Confidence-Related Decision Making. *Journal of Neurophysiology*, *104*(1), 539–547. https://doi.org/10.1152/jn.01068.2009

Isen, A. M., & Labroo, A. A. (2012). Some Ways in Which Positive Affect Facilitates Decision Making and Judgment. In E. Michael Lewis, J. M. Haviland-jones, & L. Feldman Barrett (Eds.), *Emerging Perspectives on Judgment and Decision Research* (pp. 365–393). https://doi.org/10.1017/cbo9780511609978.013

Jansen, B. R. J., Louwerse, J., Straatemeier, M., Van der Ven, S. H. G., Klinkenberg, S., & Van der Maas, H. L. J. (2013). The influence of experiencing success in math on math anxiety, perceived math competence, and math performance. *Learning and Individual Differences*, *24*, 190–197.

https://doi.org/10.1016/j.lindif.2012.12.014
Jansen, B. R. J., Raijmakers, M. E. J., & Visser, I. (2007). Rule transition on the balance scale task: A case study in belief change. *Synthese*, *155*(2), 211–236. https://doi.org/10.1007/s11229-006-9142-9
Jansen, B. R. J., & van der Maas, H. L. J. (2002). The development of children's rule use on the balance scale task. *Journal of Experimental Child Psychology*, *81*(4), 383–416. https://doi.org/10.1006/jecp.2002.2664
Jia, X., Li, W., & Cao, L. (2019). The role of metacognitive components in creative thinking. *Frontiers in Psychology*, *10*(OCT). https://doi.org/10.3389/fpsyg.2019.02404
Johnson, J. (2003). Children, robotics, and education. *Artificial Life and Robotics*, *7*(1–2), 16–21. https://doi.org/10.1007/bf02480880
Johnston-Wilder, S., Brindley, J., & Dent, P. (2014). *A survey of Mathematics Anxiety and Mathematical Resilience among existing apprentices*. *The Gatsby Foundation*.
Jones, E. E., & Berglas, S. (1978). Control of Attributions about the Self Through Self-handicapping Strategies: The Appeal of Alcohol and the Role of Underachievement. *Personality and Social Psychology Bulletin*, *4*(2), 200–206. https://doi.org/10.1177/014616727800400205
Jonsson, A. C., & Allwood, C. M. (2003). Stability and variability in the realism of confidence judgements over time, content domain, and gender. *Personality and Individual Differences*, *34*(4), 559–574. https://doi.org/10.1016/S0191-8869(02)00028-4
Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, *3*(3), 430–454. https://doi.org/10.1016/0010-0285(72)90016-3
Kalelioğlu, F. (2015). A new way of teaching programming skills to K-12 students: Code.org. *Computers in Human Behavior*, *52*, 200–210. https://doi.org/10.1016/J.CHB.2015.05.047
Karabenick, S. A., & Knapp, J. R. (1988). Help seeking and the need for academic assistance. *Journal of Educational Psychology*, *80*(3), 406–408. https://doi.org/10.1037/0022-0663.80.3.406
Karahoca, D., Karahoca, A., & Uzunboylu, H. (2011). Robotics teaching in primary school education by project based learning

for supporting science and technology courses. *Procedia Computer Science*, *3*, 1425–1431. https://doi.org/10.1016/j.procs.2011.01.025
Karp, T., Gale, R., Lowe, L. A., Medina, V., & Beutlich, E. (2010). Generation NXT: Building young engineers with LEGOs. *IEEE Transactions on Education*, *53*(1), 80–87. https://doi.org/10.1109/TE.2009.2024410
Kazimoglu, C., Kiernan, M., Bacon, L., & MacKinnon, L. (2011). Understanding Computational Thinking before Programming. *International Journal of Game-Based Learning*, *1*(3), 30–52. https://doi.org/10.4018/ijgbl.2011070103
Kelleher, C., & Pausch, R. (2003). *Lowering the Barriers to Programming: a survey of programming environments and languages for novice programmers*. Retrieved from https://apps.dtic.mil/dtic/tr/fulltext/u2/a457911.pdf
Keller, Johannes. (2007). Stereotype threat in classroom settings: The interactive effect of domain identification, task difficulty and stereotype threat on female students' maths performance. *British Journal of Educational Psychology*, *77*(2), 323–338. https://doi.org/10.1348/000709906X113662
Keller, John. (1983). Motivational design of instruction. *Instructional -Design Theories and Models*.
Kenney-Benson, G. A., Pomerantz, E. M., Ryan, A. M., & Patrick, H. (2006). Sex differences in math performance: The role of children's approach to schoolwork. *Developmental Psychology*, *42*(1), 11–26. https://doi.org/10.1037/0012-1649.42.1.11
Kern, J. K., Trivedi, M. H., Garver, C. R., Grannemann, B. D., Andrews, A. A., Savla, J. S., … Schroeder, J. L. (2006). The pattern of sensory processing abnormalities in autism. *Autism*, *10*(5), 480–494. https://doi.org/10.1177/1362361306066564
Kidd, D. G., & Monk, C. A. (2009). Are unskilled drivers aware of their deficiencies? How driving skills influence the accuracy of driving performance estimates. *Proceedings of the Human Factors and Ergonomics Society*, *3*, 1781–1785. https://doi.org/10.1518/107118109x12524444082114
Kim, B., & Reeves, T. C. (2007). Reframing research on learning with technology: in search of the meaning of cognitive tools. *Instructional Science*, *35*(3), 207–256. https://doi.org/10.1007/s11251-006-9005-2

Kim, Y. H., Chiu, C. Y., & Bregant, J. (2015). Unskilled and don't want to be aware of it: The effect of self-relevance on the unskilled and unaware phenomenon. *PLoS ONE*, *10*(6), 4–12. https://doi.org/10.1371/journal.pone.0130309

Kirlik, A., & Storkerson, P. (2010). Naturalizing Peirce's semiotics: Ecological psychology's solution to the problem of creative abduction. *Studies in Computational Intelligence*, *314*, 31–50. https://doi.org/10.1007/978-3-642-15223-8_2

Kisac, I., & Budak, Y. (2014). Metacognitive Strategies of the University Students with Respect to their Perceived Self-confidence Levels about Learning. *Procedia - Social and Behavioral Sciences*, *116*, 3336–3339. https://doi.org/10.1016/j.sbspro.2014.01.759

Kleitman, S., & Stankov, L. (2007). Self-confidence and metacognitive processes. *Learning and Individual Differences*, *17*(2), 161–173. https://doi.org/10.1016/j.lindif.2007.03.004

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Neurosciences*, *27*(12), 712–719. https://doi.org/10.1016/j.tins.2004.10.007

Knowles, M. S. (1975). Self-Directed Learning: A Guide for Learners and Teachers. Association Press, 291 Broadway, New York, New York 10007 ($4.95).

Koedinger, K., McLaughlin, E., & Stamper, J. (2012). Automated Student Model Improvement. *Educational Data Mining, Proceedings of the 5th International Conference On*, 17–24. https://doi.org/10.978.17421/02764

Koedinger, K. R., & Anderson, J. R. (1997). Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education*, *8*, 30–43. Retrieved from http://telearn.archives-ouvertes.fr/hal-00197383/

Kolb, A. Y., & Kolb, D. A. (2005). Learning styles and learning spaces: Enhancing experiential learning in higher education. *Academy of Management Learning and Education*, *4*(2), 193–212. https://doi.org/10.5465/AMLE.2005.17268566

Kolb, D. A. (1984). Experiential Learning: Experience as The Source of Learning and Development. *Prentice Hall, Inc.*, (1984), 20–38. https://doi.org/10.1016/B978-0-7506-7223-8.50017-4

Koriat, A., Sheffer, L., & Ma'ayan, H. (2002). Comparing objective and subjective learning curves: Judgments of learning exhibit increased underconfidence with practice. *Journal of Experimental Psychology: General*, *131*(2), 147–162. https://doi.org/10.1037/0096-3445.131.2.147

Koziol, L. F., Budding, D. E., & Chidekel, D. (2012). From movement to thought: Executive function, embodied cognition, and the cerebellum. *Cerebellum*, *11*(2), 505–525. https://doi.org/10.1007/s12311-011-0321-y

Krueger, J., & Mueller, R. A. (2002). Unskilled, unaware, or both? The better-than-average heuristic and statistical regression predict errors in estimates of own performance. *Journal of Personality and Social Psychology*, *82*(2), 180–188. https://doi.org/10.1037/0022-3514.82.2.180

Kruger, J., & Dunning, D. (1999). Unskilled and Unaware of it: How difficulties in recognizing one's incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, *77*(6), 1121–1134. https://doi.org/10.1037/0022-3514.77.6.1121

Lave, J., & Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge, UK: Cambridge University Press.

Law, E. L.-C., & Sun, X. (2012). Evaluating user experience of adaptive digital educational games with Activity Theory. *International Journal of Human-Computer Studies*, *70*(7), 478–497.

Leary, M. R., & Hill, D. A. (1996). Moving on: Autism and movement disturbance. *Mental Retardation*, *34*(1), 39–53.

Lebreton, M., Abitbol, R., Daunizeau, J., & Pessiglione, M. (2015). Automatic integration of confidence in the brain valuation signal. *Nature Neuroscience*, *18*(8), 1159–1167. https://doi.org/10.1038/nn.4064

Legare, C. H. (2012). Exploring exploration: explaining inconsistent information guides hypothesis-testing behavior in young children. *Child Dev*, *83*, 173–185.

Legare, Cristine H. (2014). The contributions of explanation and exploration to children's scientific reasoning. *Child Development Perspectives*, *8*(2), 101–106. https://doi.org/10.1111/cdep.12070

Li, H., Yuan, J., & Lin, C. (2008). The neural mechanism underlying the female advantage in identifying negative emotions: An event-related potential study. *NeuroImage*, *40*(4), 1921–1929. https://doi.org/10.1016/j.neuroimage.2008.01.033

Liao, S. H. (2005). Expert system methodologies and applications-a decade review from 1995 to 2004. *Expert Systems with Applications*, *28*(1), 93–103. https://doi.org/10.1016/j.eswa.2004.08.003

Liberman, V. (2004, May). Local and global judgments of confidence. *Journal of Experimental Psychology: Learning Memory and Cognition*. https://doi.org/10.1037/0278-7393.30.3.729

Lichtenstein, S., & Fischhoff, B. (1977). Do Those Who Know More Also Know More about How Much They Know?, *20*, 159–183. Retrieved from https://s3.amazonaws.com/academia.edu.documents/46684681/Do_Those_Who_Know_More_Also_Know_More_ab20160621-32318-igw1a6.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1524502177&Signature=e4sJcBTV5vKOq5h%2FYEsjRb6DFRE%3D&response-content-disposition=inline%3B filename%3DDo_those_who_know_more_also_know_more_ab.pdf

Lieder, I., Adam, V., Frenkel, O., Jaffe-Dax, S., Sahani, M., & Ahissar, M. (2019). Perceptual bias reveals slow-updating in autism and fast-forgetting in dyslexia. *Nature Neuroscience*, *22*(2), 256–264. https://doi.org/10.1038/s41593-018-0308-9

Lin, C. H., & Liu, E. Z. F. (2011). A pilot study of Taiwan elementary school students learning motivation and strategies in robotics learning. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 6872 LNCS, pp. 445–449). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-23456-9_80

Lindberg, S. M., Hyde, J. S., Petersen, J. L., & Linn, M. C. (2010). New Trends in Gender and Mathematics Performance: A Meta-Analysis. *Psychological Bulletin*, *136*(6), 1123–1135. https://doi.org/10.1037/a0021276

Loyens, S. M. M., Magda, J., & Rikers, R. M. J. P. (2008). Self-directed learning in problem-based learning and its

relationships with self-regulated learning. *Educational Psychology Review*, *20*(4), 411–427. https://doi.org/10.1007/s10648-008-9082-7
Lu, J. J., & Fletcher, G. H. L. (2009). Thinking about Computational Thinking. In *SIGCSE'09 - Proceedings of the 40th ACM Technical Symposium on Computer Science Education* (Vol. 41, pp. 260–264). New York, New York, USA: ACM Press. https://doi.org/10.1145/1508865.1508959
M. Grant, D., D. Malloy, A., & C. Murphy, M. (2009). A Comparison of Student Perceptions of their Computer Skills to their Actual Abilities. *Journal of Information Technology Education: Research*, *8*, 141–160. https://doi.org/10.28945/164
Ma, X., & Xu, J. (2004). The causal ordering of mathematics anxiety and mathematics achievement: A longitudinal panel analysis. *Journal of Adolescence*, *27*(2), 165–179. https://doi.org/10.1016/j.adolescence.2003.11.003
Mabe, P. A., & West, S. G. (1982). Validity of self-evaluation of ability: A review and meta-analysis. *Journal of Applied Psychology*, *67*(3), 280–296. https://doi.org/10.1037/0021-9010.67.3.280
Machado, S., Cunha, M., Velasques, B., Minc, D., Teixeira, S., Domingues, C. A., … Ribeiro, P. (2010). Sensorimotor integration: basic concepts, abnormalities related to movement disorders and sensorimotor training-induced cortical reorganization. *Revista de Neurologia*, *51*(7), 427–436. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/20859923
Makarova, E., Aeschlimann, B., & Herzog, W. (2019). The Gender Gap in STEM Fields: The Impact of the Gender Stereotype of Math and Science on Secondary Students' Career Aspirations. *Frontiers in Education*, *4*, 60. https://doi.org/10.3389/feduc.2019.00060
Makarova, E., & Herzog, W. (2015). Trapped in the gender stereotype? the image of science among secondary school students and teachers. *Equality, Diversity and Inclusion*, *34*(2), 106–123. https://doi.org/10.1108/EDI-11-2013-0097
Maki, R. H. (1998). Test Predictions Over Text Materials. In *Metacognition in Educational Theory and Practice*. Retrieved from

https://books.google.es/books?hl=en&lr=&id=EzWRAgAAQBAJ&oi=fnd&pg=PA117&dq=Test+predictions+over+text+material.&ots=lxzaYyABbA&sig=K-lGjxJNo5lG9FIPBnkpNKCQl3w&redir_esc=y#v=onepage&q=Test predictions over text material.&f=false

Malinverni, L., Schaper, M. M., & Valero, C. (2020). Relating to materials in digital fabrication: Transform materials to transform yourself. *International Journal of Child-Computer Interaction*, 100166. https://doi.org/10.1016/j.ijcci.2020.100166

Marsh, H. W., & Martin, A. J. (2011, March 1). Academic self-concept and academic achievement: Relations and causal ordering. *British Journal of Educational Psychology*. John Wiley & Sons, Ltd. https://doi.org/10.1348/000709910X503501

Marsh, H. W., & Yeung, A. S. (1998). Longitudinal structural equation models of academic self-concept and achievement: Gender differences in the development of math and English constructs. *American Educational Research Journal*, *35*(4), 705–738. https://doi.org/10.3102/00028312035004705

Martinez, S., & Stager, G. (2019). *Invent To Learn: Making, Tinkering, And Engineering In The Classroom*. Kbh, Nota. Retrieved from https://books.google.es/books?id=4C5evgAACAAJ&dq=Invent+to+Learn:+Making,+Tinkering+and+Engineering+in+the+Classroom&hl=es&sa=X&ved=2ahUKEwic_JvAn67qAhWvzIUKHer0C1gQ6AEwAHoECAAQAg

Martino, B. De, Fleming, S. M., & Garrett, N. (2013). Confidence in value-based choice, *16*(1), 105–110. https://doi.org/10.1038/nn.3279.Confidence

Massaro, D. W. (n.d.). Survey of Psychological Literacy. Retrieved May 13, 2020, from https://mambo.ucsc.edu/survey-of-psychological-literacy/

Master, A., Cheryan, S., & Meltzoff, A. N. (2016). Computing whether she belongs: Stereotypes undermine girls' interest and sense of belonging in computer science. *Journal of Educational Psychology*, *108*(3), 424–437. https://doi.org/10.1037/edu0000061

Matsushima, K., & Kato, T. (2013). Social interaction and atypical sensory processing in children with autism spectrum disorders.

Hong Kong Journal of Occupational Therapy, *23*(2), 89–96. https://doi.org/10.1016/j.hkjot.2013.11.003

Mazzoni, G., & Nelson, T. O. (1995). Judgments of Learning Are Affected by the Kind of Encoding in Ways That Cannot Be Attributed to the Level of Recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(5), 1263–1274. https://doi.org/10.1037/0278-7393.21.5.1263

Mccrea, S. M., & Hirt, E. R. (2001). *The Role of Ability Judgments in Self-Handicapping*. Retrieved from http://kops.ub.uni-konstanz.de/volltexte/2009/7629/

McCullough, M. (2004). *Digital Ground* . MIT Press. Retrieved from https://mitpress.mit.edu/books/digital-ground

McMahon, C. M., Henderson, H. A., Newell, L., Jaime, M., & Mundy, P. (2016). Metacognitive Awareness of Facial Affect in Higher-Functioning Children and Adolescents with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *46*(3), 882–898. https://doi.org/10.1007/s10803-015-2630-3

Meelissen, M., & Luyten, H. (2008). The Dutch gender gap in mathematics: Small for achievement, substantial for beliefs and attitudes. *Studies in Educational Evaluation*, *34*(2), 82–93. https://doi.org/10.1016/j.stueduc.2008.04.004

Mehdizadeh, L., Sturrock, A., Myers, G., Khatib, Y., & Dacre, J. (2014). How well do doctors think they perform on the General Medical Council's Tests of Competence pilot examinations? A cross-sectional study. *BMJ Open*, *4*(2), e004131. https://doi.org/10.1136/bmjopen-2013-004131

Mello, S. K. D., Craig, S. D., Witherspoon, A., & Mcdaniel, B. (2008). Automatic detection of learner ' s affect from conversational cues. *User Modeling and User-Adapted Interaction*, *1*, 45–80. https://doi.org/10.1007/s11257-007-9037-6

Mengelkamp, C., & Bannert, M. (2010). Accuracy of confidence judgments: Stability and generality in the learning process and predictive validity for learning outcome. *Memory & Cognition*, *38*(4), 441–451. https://doi.org/10.3758/MC.38.4.441

Mercer, N. (2000). Development through dialogue. In N. Mercer (Ed.), *Words and Minds* (pp. 131–166). London: Routledge.

Mercer, N. (2004). Sociocultural discourse analysis: analysing

classroom talk as a social mode of thinking. *Journal of Applied Linguistics*, *1*(2), 137–168. https://doi.org/10.1558/japl.2004.1.2.137
Metcalfe, J. (2009). Metacognitive Judgments and Control of Study, *18*(3), 159–163. https://doi.org/10.1111/j.1467-8721.2009.01628.x.Metacognitive
Metcalfe, J. (2017). Learning from Errors. *Annual Review of Psychology*, *68*, 465–489. https://doi.org/10.1146/annurev-psych-010416-044022
Meyniel, F., Schlunegger, D., & Dehaene, S. (2015). The Sense of Confidence during Probabilistic Learning: A Normative Account. *PLoS Computational Biology*. https://doi.org/10.1371/journal.pcbi.1004305
Midgley, C., Arunkumar, R., & Urdan, T. C. (1996). "If I Don't Do Well Tomorrow, There's a Reason": Predictors of Adolescents' Use of Academic Self-Handicapping Strategies. *Journal of Educational Psychology*, *88*(3), 423–434. https://doi.org/10.1037/0022-0663.88.3.423
Midgley, C., & Urdan, T. (2001). Academic self-handicapping and achievement goals: A further examination. *Contemporary Educational Psychology*, *26*(1), 61–75. https://doi.org/10.1006/ceps.2000.1041
Miller, T. M., & Geraci, L. (2011). Training metacognition in the classroom: The influence of incentives and feedback on exam predictions. *Metacognition and Learning*, *6*(3), 303–314. https://doi.org/10.1007/s11409-011-9083-7
Mintzes, J. J., Wandersee, J. H., & Novak, J. D. (2005). *Teaching Science for Understanding: A Human Constructivist View*. Academic Press. Retrieved from https://books.google.es/books?hl=en&lr=&id=2zXcrHWhnDwC&oi=fnd&pg=PP1&dq=Mintzes+JJ,+Wandersee+JH,+and+Novak+JD+(editors).+Teaching+Science+for+Understanding:+a+Human+Constructivist+View.+San+Diego,+CA:+Academic,+1998&ots=mV-9-L7HU0&sig=mu2nwM3Jl4EPxkRX2h
Mitchell, P., & Ropar, D. (2004). Visuo-spatial abilities in autism: A review. *Infant and Child Development*. https://doi.org/10.1002/icd.348
Moeller, S. K., & Robinson, M. D. (2010). Sex differences in implicit punishment sensitivity: Evidence from two cognitive paradigms. *Personality and Individual Differences*, *48*(3),

283–287. https://doi.org/10.1016/j.paid.2009.10.017
Morrison, B. B., Margulieux, L. E., Ericson, B., & Guzdial, M. (2016). Subgoals help students solve Parsons Problems. *SIGCSE 2016 - Proceedings of the 47th ACM Technical Symposium on Computing Science Education*, 42–47. https://doi.org/10.1145/2839509.2844617
Mueller, C. M., & Dweck, C. S. (1998). *Praise for Intelligence Can Undermine Children's Motivation and Performance*. *Journal of Personality and Social Psychology* (Vol. 75). Retrieved from http://radio.shabanali.com/intelligence_praise_mueller_dweck.pdf
Multon, K. D., Brown, S. D., & Lent, R. W. (1991). Relation of Self-Efficacy Beliefs to Academic Outcomes: A Meta-Analytic Investigation. *Journal of Counseling Psychology*, *38*(1), 30–38. https://doi.org/10.1037/0022-0167.38.1.30
Mynttinen, S., Sundström, A., Vissers, J., Koivukoski, M., Hakuli, K., & Keskinen, E. (2009). Self-assessed driver competence among novice drivers - a comparison of driving test candidate assessments and examiner assessments in a Dutch and Finnish sample. *Journal of Safety Research*, *40*(4), 301–309. https://doi.org/10.1016/j.jsr.2009.04.006
Nagy, A., Eördegh, G., Paróczy, Z., Márkus, Z., & Benedek, G. (2006). Multisensory integration in the basal ganglia. *European Journal of Neuroscience*, *24*(3), 917–924. https://doi.org/10.1111/j.1460-9568.2006.04942.x
Narumi, T., Hayashi, O., Kasada, K., Yamazaki, M., Tanikawa, T., & Hirose, M. (2011). Digital diorama: AR exhibition system to convey background information for museums. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 6773 LNCS, pp. 76–86). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-22021-0_10
Nawijn, J., Isaac, R. K., Gridnevskiy, K., & van Liempt, A. (2018). Holocaust concentration camp memorial sites: an exploratory study into expected emotional response. *Current Issues in Tourism*, *21*(2), 175–190. https://doi.org/10.1080/13683500.2015.1058343
Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: a model and seven principles of

good feedback practice. *Studies in Higher Education*, *31*(2), 199–218. https://doi.org/10.1080/03075070600572090
Niederle, M., Segal, C., & Vesterlund, L. (2013). How costly is diversity? Affirmative action in light of gender differences in competitiveness. *Management Science*, *59*(1), 1–16. https://doi.org/10.1287/mnsc.1120.1602
Niederle, M., & Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much? *Quarterly Journal of Economics*, *112*(3), 1067–1101. Retrieved from http://www.stanford.edu/~niederle,
Nietfeld, J. L., Cao, L., & Osborne, J. W. (2005). Metacognitive Monitoring Accuracy. *The Journal of Experimental Education*, *74*(1), 7–28.
Nourbakhsh, I. R., Crowley, K., Bhave, A., Hamner, E., Hsiu, T., Perez-Bergquist, A., … Wilkinson, K. (2005). The robotic autonomy mobile robotics course: Robot design, curriculum design and educational assessment. *Autonomous Robots*, *18*(1), 103–127. https://doi.org/10.1023/B:AURO.0000047303.20624.02
Oecd. (2014). *PISA 2012 Results: Creative Problem Solving (Volume V)* (Vol. V). https://doi.org/10.1787/9789264208070-en
Oldehinkel, M., Mennes, M., Marquand, A., Charman, T., Tillmann, J., Ecker, C., … Zwiers, M. P. (2019). Altered Connectivity Between Cerebellum, Visual, and Sensory-Motor Networks in Autism Spectrum Disorder: Results from the EU-AIMS Longitudinal European Autism Project. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *4*(3), 260–270. https://doi.org/10.1016/j.bpsc.2018.11.010
Pacheco, D., Sánchez-Fibla, M., Duff, A., & Verschure, P. F. M. J. (2017). A spatial-context effect in recognition memory. *Frontiers in Behavioral Neuroscience*, *11*(August), 1–9. https://doi.org/10.3389/fnbeh.2017.00143
Paglia, F. La, Barbera, D. La, Caci, B., & Cardaci, M. (2010). Using robotics construction kits as metacognitive tools: a research in an italian primary school.
Pajares, F. (1996a). Self-efficacy beliefs in academic settings. *Review of Educational Research*, *66*(4), 543–578. https://doi.org/10.3102/00346543066004543

Pajares, F. (1996b). *Self-Efficacy Beliefs in Academic Settings* (Vol. 66). Retrieved from http://rer.aera.net

Pajares, F. (2002). Gender and perceived self-efficacy in self-regulated learning. *Theory into Practice*, *41*(2), 116–125. https://doi.org/10.1207/s15430421tip4102_8

Pajares, F. (2003). Self-Efficacy Beliefs,Motivation, and Achievement in Writing: A Review of the Literature. *Reading and Writing Quarterly*, *19*, 139–159. https://doi.org/10.1080/10573560390143085

Pallasmaa, J. (2010). The thinking hand: existential and embodied wisdom in architecture. *Choice Reviews Online*, *47*(06), 47-2983-47–2983. https://doi.org/10.5860/choice.47-2983

Palmen, S. J. M. C., Van Engeland, H., Hof, P. R., & Schmitz, C. (2004). Neuropathological findings in autism. *Brain*, *127*(12), 2572–2583. https://doi.org/10.1093/brain/awh287

Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology*, *8*(APR), 1–28. https://doi.org/10.3389/fpsyg.2017.00422

Papadakis, S. (2018). Gender Stereotypes in School Textbooks. *Revista Romaneasca Pentru Educatie Multidimensionala*, *10*(4), 65–81. https://doi.org/10.18662/rrem/73

Papastergiou, M. (2009). Digital Game-Based Learning in high school Computer Science education: Impact on educational effectiveness and student motivation. *Computers & Education*, *52*(1), 1–12. https://doi.org/10.1016/j.compedu.2008.06.004

Papert, S. (1993). *The children's machine: rethinking school in the age of the computer*. *Basic Books Inc.* (Vol. 31).

Papert, Seymour. (1980). *Mindstorms: Children, computers, and powerful ideas*. New York, USA: Basic Books Inc. Retrieved from http://dl.acm.org/citation.cfm?id=1095592

Park, Y. J., & Santos-Pinto, L. (2010). Overconfidence in tournaments: evidence from the field. *Theory Dec*, *69*, 143–166. https://doi.org/10.1007/s11238-010-9200-0

Parker, D. B., Swope, S., & Baker, B. N. (1990). *Ethical conflicts in information and computer science, technology, and business*. *QED Information Sciences*. Retrieved from http://ezproxy.lib.indiana.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=cat00016a&AN=inun.21492&site=eds-live&scope=site&scope=cite

Paulin, M. G. (1993). The role of the cerebellum in motor control and perception. *Brain, Behavior and Evolution*, *41*(1), 39–50. https://doi.org/10.1159/000113822

Pavel, S., Robertson, M., & Harrison, B. (2012). The Dunning-Kruger Effect and SIUC University's Aviation Students. *Journal of Aviation Technology and Engineering*, *2*(1), 125–129. https://doi.org/10.5703/1288284314864

Payne, B. D., & Manning, B. H. (1992). Basal reader instruction: Effects of comprehension monitoring training on reading comprehension, strategy use and attitude. *Reading Research and Instruction*, *32*(1), 29–38. https://doi.org/10.1080/19388079209558103

Peirce, C.S., & Jastrow, J. (1884). On small differences of sensation. *Memoirs of the National Academy of Sciences*, *3*(1884), 73–83. https://doi.org/52, 281-302

Peirce, Charles Sanders. (1997). Pragmatism as a Principle and Method of Right Thinking: The 1903 Harvard Lectures on Pragmatism, 305. Retrieved from https://books.google.es/books?hl=en&lr=&id=_TUqldjTO80C&oi=fnd&pg=PR7&dq=Pragmatism+as+a+principle+and+method+of+right+thinking&ots=UA4SjL2ZaH&sig=i8h5w_BLFPOO3b0sP_gI8Rr7XNg&redir_esc=y#v=onepage&q=Pragmatism as a principle and method of right thinking&f

Pekrun, R. (2000). A social-cognitive, control-value theory of achievement emotions. In *Advances in Psychology* (Vol. 131, pp. 143–163). Elsevier. https://doi.org/10.1016/S0166-4115(00)80010-2

Perez-felkner, L., Nix, S., Thomas, K., & Kirkham, N. (2017). Gendered Pathways : How Mathematics Ability Beliefs Shape Secondary and Postsecondary Course and Degree Field Choices, *8*(April), 1–11. https://doi.org/10.3389/fpsyg.2017.00386

Perkins, D. N. (2001). *The Eureka Effect: The Art and Logic of Breakthrough Thinking*. WW Norton & Company.

Perković, L., Settle, A., Hwang, S., & Jones, J. (2010). A framework for computational thinking across the curriculum. *Proceedings of the Fifteenth Annual Conference on Innovation and Technology in Computer Science Education - ITiCSE '10*, 123. https://doi.org/10.1145/1822090.1822126

Petre, M., & Price, B. (2004). *Using Robotics to Motivate "Back*

Door" Learning. *Education and Information Technologies* (Vol. 9). Kluwer Academic Publishers. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.119.4887&rep=rep1&type=pdf

Philips, S., & Tolmie, A. (2007). Children's performance on and understanding of the Balance Scale problem: the effects of parental support. *Infant and Child Development: An International Journal of Research and Practice*, *16*(1), 95–117.

Piaget, J. (1952). *When thinking begins*. *The origins of intelligence in children.*

Piaget, J. (1977). The Role of Action in the Development of Thinking. In *Knowledge and Development* (pp. 17–42). Springer US. https://doi.org/10.1007/978-1-4684-2547-5_2

Piaget, J., & Inhelder, B. (1958). *The growth of logical thinking from childhood to adolescence: An essay on the construction of formal operational structures. Routledge*. https://doi.org/10.1016/0364-5916(91)90030-N

Pierce, R., & Stacey, K. (2004). A framework for monitoring progress and planning teaching towards the effective use of computer algebra systems. *International Journal of Computers for Mathematical Learning*, *9*(1), 59–93. https://doi.org/10.1023/B:IJCO.0000038246.98119.14

Pietsch, J., Walker, R., & Chapman, E. (2003). The Relationship Among Self-Concept, Self-Efficacy, and Performance in Mathematics During Secondary School. *Journal of Educational Psychology*, *95*(3), 589–603. https://doi.org/10.1037/0022-0663.95.3.589

Pintrich, P. R. (2000). Multiple goals, multiple pathways: The role of goal orientation in learning and achievement. *Journal of Educational Psychology*, *92*(3), 544–555. https://doi.org/10.1037/0022-0663.92.3.544

Pintrich, P. R. (2003). A Motivational Science Perspective on the Role of Student Motivation in Learning and Teaching Contexts. *Journal of Educational Psychology*, *95*(4), 667–686. https://doi.org/10.1037/0022-0663.95.4.667

Pintrich, P. R., & de Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, *82*(1), 33–40. https://doi.org/10.1037/0022-0663.82.1.33

Pintrich, P. R., Wolters, C. A., & Baxter, G. P. (2000). 2 . Assessing Metacognition and Self-Regulated Learning. *Issues in the Measurement of Metacognition*, 43–97. Retrieved from https://digitalcommons.unl.edu/burosmetacognition/3

Porayska-Pomsta, K., Mavrikis, M., D'Mello, S., Conati, C., & Baker, R. S. J. D. (2013). Knowledge elicitation methods for affect modelling in education. *International Journal of Artificial Intelligence in Education*, *22*(3), 107–140. https://doi.org/10.3233/JAI-130032

Prince, M. J., & R. M. Felder. (2006). Inductive Teaching and Learning Methods: Definitions, Comparisons, and Research Bases. *J. of Engineering Education*, *95*(2), 123–138.

Pujol, L., & Champion, E. (2012). Evaluating presence in cultural heritage projects. *International Journal of Heritage Studies*, *18*(1), 83–102. https://doi.org/10.1080/13527258.2011.577796

Puustinen, M. (1998). *Help-seeking behavior in a problem-solving situation: Development of self-regulation*. *European Journal ofPsychology ofEducation*. Retrieved from https://link.springer.com/content/pdf/10.1007/BF03173093.pdf

Rahaman, H. (2018). Digital heritage interpretation: a conceptual framework. *Digital Creativity*, *29*(2–3), 208–234. https://doi.org/10.1080/14626268.2018.1511602

Rahaman, H., & Tan, B.-K. (2011). Interpreting Digital Heritage: A Conceptual Model with End-Users' Perspective. *International Journal of Architectural Computing*, *9*(1), 99–113. https://doi.org/10.1260/1478-0771.9.1.99

Ramirez, G., & Beilock, S. L. (2011). Writing about testing worries boosts exam performance in the classroom. *Science*, *331*(6014), 211–213. https://doi.org/10.1126/science.1199427

Rattan, A., Good, C., & Dweck, C. S. (2012). "It's ok - Not everyone can be good at math": Instructors with an entity theory comfort (and demotivate) students. *Journal of Experimental Social Psychology*, *48*(3), 731–737. https://doi.org/10.1016/j.jesp.2011.12.012

Rawson, K. a., & Dunlosky, J. (2007). Improving students' self-evaluation of learning for key concepts in textbook materials. *European Journal of Cognitive Psychology*, *19*(4–5), 559–579. https://doi.org/10.1080/09541440701326022

Reber, A. S. (1993). *Implicit Learning and Tacit Knowledge: An*

Essay on the Cognitive Unconscious. *Implicit Learning and Tacit Knowledge: An Essay on the Cognitive Unconscious*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195106589.001.0001

Reidsma, D., Charisi, V., Davison, D., Wijnen, F., van der Meij, J., Evers, V., … Verschure, P. (2016). *The EASEL project: Towards educational human-robot symbiotic interaction*. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 9793). https://doi.org/10.1007/978-3-319-42417-0_27

Renfrow, D. G., & Howard, J. A. (2013). Social Psychology of Gender and Race (pp. 491–531). Springer, Dordrecht. https://doi.org/10.1007/978-94-007-6772-0_17

Resnick, M., Maloney, Andrés Monroy-Hernández, J., Rusk, N., Eastmond, E., Brennan, K., Rosenbaum, E., … Kafai, Y. (2009). Scratch: Programming for all, *52*(11). https://doi.org/10.1145/1592761.1592779

Reynolds, D., Sammons, P., De Fraine, B., Townsend, T., & Van Damme, J. (2011). Educational Effectiveness Research (EER): A State of the Art Review, 43.

Rhodewalt, F., Morf, C., Hazlett, S., & Fairfield, M. (1991). Self-Handicapping: The Role of Discounting and Augmentation in the Preservation of Self-Esteem. *Journal of Personality and Social Psychology*, *61*(1), 122–131. https://doi.org/10.1037/0022-3514.61.1.122

Rinehart, N., & Mcginley, J. (2010, March 19). Is motor dysfunction core to autism spectrum disorder? *Developmental Medicine and Child Neurology*. https://doi.org/10.1111/j.1469-8749.2010.03631.x

Rinne, L. F., & Mazzocco, M. M. M. (2014). Knowing Right From Wrong In Mental Arithmetic Judgments: Calibration Of Confidence Predicts The Development Of Accuracy. *PLoS ONE*, *9*(7), e98663. https://doi.org/10.1371/journal.pone.0098663

Robic, S., Sonié, S., Fonlupt, P., Henaff, M. A., Touil, N., Coricelli, G., … Schmitz, C. (2015). Decision-Making in a Changing World: A Study in Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *45*(6), 1603–1613. https://doi.org/10.1007/s10803-014-2311-7

Robinson, M. (2005). Robotics-Driven Activities: Can They Improve Middle School Science Learning? *Bulletin of Science, Technology & Society*, *25*(1), 73–84. https://doi.org/10.1177/0270467604271244

Robinson, O. J., Standing, H. R., DeVito, E. E., Cools, R., & Sahakian, B. J. (2010). Dopamine precursor depletion improves punishment prediction during reversal learning in healthy females but not males. *Psychopharmacology*, *211*(2), 187–195. https://doi.org/10.1007/s00213-010-1880-1

Rogers, C., & Portsmore, M. (2004). Bringing Engineering to Elementary School. *Journal of STEM Education*, *5*(3), 17–29.

Rogers, Y. (2008). 57 Varieties of Activity Theory. *Interacting with Computers*, *20*(2), 247–250.

Roll, I., Aleven, V., McLaren, B. M., & Koedinger, K. R. (2007). Designing for metacognition-applying cognitive tutor principles to the tutoring of help seeking. *Metacognition and Learning*, *2*(2–3), 125–140. https://doi.org/10.1007/s11409-007-9010-0

Rollwage, M., Dolan, R. J., & Fleming, S. M. (2018). Metacognitive Failure as a Feature of Those Holding Radical Beliefs. *Current Biology*, *28*(24), 4014-4021.e8. https://doi.org/10.1016/j.cub.2018.10.053

Rosenbaum, D. A., Chapman, K. M., Coelho, C. J., Gong, L., & Studenka, B. E. (2013). Choosing actions. *Frontiers in Psychology*, *4*(FEB), 1–14. https://doi.org/10.3389/fpsyg.2013.00273

Ross, M. E., Shannon, D. M., Salisbury-Glennon, J. D., & Guarino, A. (2002). the Patterns of Adaptive Learning Survey: a Comparison Across Grade Levels. *Educational and Psychological Measurement*, *62*(3), 483–497. https://doi.org/10.1177/00164402062003006

Rouault, M., Seow, T., Gillan, C. M., & Fleming, S. M. (2018). Psychiatric Symptom Dimensions Are Associated With Dissociable Shifts in Metacognition but Not Task Performance. *Biological Psychiatry*. https://doi.org/10.1016/j.biopsych.2017.12.017

Roussou, M. (2002). Virtual Heritage: From the Research Lab to the Broad Public. In *Virtual Archaeology, Proceedings of the VAST2000 Euroconference, BAR International Series 1075* (pp. 93–100). Arezzo, Italy. Retrieved from

https://www.researchgate.net/publication/243767413
Rubinsten, O., Bialik, N., & Solar, Y. (2012). Exploring the relationship between math anxiety and gender through implicit measurement. *Frontiers in Human Neuroscience, 6*(SEPTEMBER), 279. https://doi.org/10.3389/fnhum.2012.00279
Ryan, A. M., Gheen, M. H., & Midgley, C. (1998). Why Do Some Students Avoid Asking for Help? An Examination of the Interplay among Students' Academic Efficacy, Teachers' Social-Emotional Role, and the Classroom Goal Structure. *Journal of Educational Psychology*, *90*(3), 528–535. https://doi.org/10.1037/0022-0663.90.3.528
Sadker, M., & Sadker, D. M. (1995). *Failing at fairness : how our schools cheat girls*. Touchstone.
Saks, K., & Leijen, Ä. (2014). Distinguishing Self-directed and Self-regulated Learning and Measuring them in the E-learning Context. *Procedia - Social and Behavioral Sciences*, *112*, 190–198. https://doi.org/10.1016/j.sbspro.2014.01.1155
Scharoun, S. M., & Bryden, P. J. (2016). Anticipatory planning in children with autism spectrum disorder: An assessment of independent and joint action tasks. *Frontiers in Integrative Neuroscience*, *10*(AUGUST2016), 29. https://doi.org/10.3389/fnint.2016.00029
Schirmer, A., Zysset, S., Kotz, S. A., & Von Cramon, D. Y. (2004). Gender differences in the activation of inferior frontal cortex during emotional speech perception. *NeuroImage*, *21*(3), 1114–1123. https://doi.org/10.1016/j.neuroimage.2003.10.048
Schmitz, C., Martineau, J., Barthélémy, C., & Assaiante, C. (2003). Motor control and children with autism: Deficit of anticipatory function? *Neuroscience Letters*, *348*(1), 17–20. https://doi.org/10.1016/S0304-3940(03)00644-X
Schneider, W. (1998). Performance prediction in young children: Effects of skill, metacognition and wishful thinking. *Developmental Science*, *1*(2), 291–297. https://doi.org/10.1111/1467-7687.00044
Schneider, W. (2008, September 1). The development of metacognitive knowledge in children and adolescents: Major trends and implications for education. *Mind, Brain, and Education*. Blackwell Publishing. https://doi.org/10.1111/j.1751-228X.2008.00041.x

Schraw, G. (2009). A conceptual analysis of five measures of metacognitive monitoring. *Metacognition and Learning*, *4*(1), 33–45. https://doi.org/10.1007/s11409-008-9031-3
Schraw, G., Crippen, K., & Hartley, K. (2006). Promoting Self-Regulation in Science Education: Metacognition as Part of a Broader Perspective on Learning. *Res Sci Educ*, *36*(1), 111–139. https://doi.org/10.1007/s11165-005-3917-8
Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology*. https://doi.org/10.1006/ceps.1994.1033
Schraw, G., & Moshman, D. (1995). *Metacognitive Theories*. Retrieved from http://www.springerlink.com/content/1040-726X
Schulz, L. (2012). The origins of inquiry: Inductive inference and exploration in early childhood. *Trends in Cognitive Sciences*, *16*(7), 382–389. https://doi.org/10.1016/j.tics.2012.06.004
Schulz, L. (2015). Infants explore the unexpected. *Science*, *348*(6230), 42–43. https://doi.org/10.1126/science.aab0582
Schulz, L. E., & Bonawitz, E. B. (2007). Serious Fun: Preschoolers Engage in More Exploratory Play When Evidence Is Confounded. *Developmental Psychology*, *43*(4), 1045–1050. https://doi.org/10.1037/0012-1649.43.4.1045
Schunk, D. H. (1991). Self efficacy and academic motivation. *Educational Psy*, *26*, 207–231. https://doi.org/DOI 10.1207/s15326985ep2603&4_2
Schwab, J. J. (1960). Inquiry, the Science Teacher, and the Educator. *The School Review*, *68*(2), 176–195. https://doi.org/10.1086/442536
Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2005.12.009
Seligman, M. E. P. (2007). Learned Helplessness. In *Encyclopedia of Stress* (pp. 567–570). https://doi.org/10.1016/B978-012373947-6.00233-6
Shea, N., Boldt, A., Bang, D., Yeung, N., Heyes, C., & Frith, C. D. (2014, April 1). Supra-personal cognitive control and metacognition. *Trends in Cognitive Sciences*. Elsevier Ltd. https://doi.org/10.1016/j.tics.2014.01.006
Shernoff, D. (2003). Student engagement in high school classrooms from the perspective of flow theory. *School Psychology …*, *18*,

158–176. Retrieved from http://psycnet.apa.org/journals/spq/18/2/158/

Siegler, R. S., & Chen, Z. (1998). Developmental Differences in Rule Learning: A Microgenetic Analysis. *Cognitive Psychology*, *36*(3), 273–310. https://doi.org/10.1006/COGP.1998.0686

Siegler, R. S., Strauss, S., & Levin, I. (1981). Developmental Sequences within and between Concepts. *Monographs of the Society for Research in Child Development*, *46*(2), 1–84. https://doi.org/10.2307/1165995

Simon, H. A. (2001). Learning to research about learning. In S. M. Carver & K. D. Mahwah (Eds.), *Cognition and Instruction: Twenty-Five Years of Progress* (pp. 205–226.). Erlbaum.

Sinha, P., Kjelgaard, M. M., Gandhi, T. K., Tsourides, K., Cardinaux, A. L., Pantazis, D., … Held, R. M. (2014). Autism as a disorder of prediction. *Proceedings of the National Academy of Sciences*, *111*(42), 15220–15225. https://doi.org/10.1073/pnas.1416797111

Sivasakthi, M., & Rajendran, R. (2011). Learning difficulties of “object-oriented programming paradigm using Java”: Students’ perspective. *Indian Journal of Science and Technology*, *4*(8), 983–985. https://doi.org/10.17485/ijst/2011/v4i8/30906

Skinner, E., & Belmont, M. (1993). Motivation in the classroom: Reciprocal effects of teacher behavior and student engagement across the school year. *Journal of Educational Psychology*, *85*(4), 571–581. Retrieved from http://psycnet.apa.org/journals/edu/85/4/571/

Smaldino, S. E., Lowther, D. L., Russell, J. D., & Mims, C. (2008). *Instructional Technology and Media for Learning*. Pearson Merrill Prentice Hall.

Smith, A. C. (2007). Using magnets in physical blocks that behave as programming objects. *Proceedings of the 1st International Conference on Tangible and Embedded Interaction - TEI ’07*, 147. https://doi.org/10.1145/1226969.1226999

Smith, B. L., MacGregor, J., Gabelnick, F., & Matthews, R. (2004). *Learning Communities: Reforming Undergraduate Education*. San Franciso: Jossey-Bass, Inc.

Sobel, D. M., & Kushnir, T. (2013). Knowledge matters: How children evaluate the reliability of testimony as a process of

rational inference. *Psychological Review*, *120*(4), 779–797. https://doi.org/10.1037/a0034191
South, M., Newton, T., & Chamberlain, P. D. (2012). Delayed Reversal Learning and Association With Repetitive Behavior in Autism Spectrum Disorders. *Autism Research*, *5*(6), 398–406. https://doi.org/10.1002/aur.1255
Stamper, J. C., & Koedinger, K. R. (2011). Human-machine student model discovery and improvement using datashop. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *6738 LNAI*, 353–360. https://doi.org/10.1007/978-3-642-21869-9_46
Stanfield, A. C., McIntosh, A. M., Spencer, M. D., Philip, R., Gaur, S., & Lawrie, S. M. (2008). Towards a neuroanatomy of autism: A systematic review and meta-analysis of structural magnetic resonance imaging studies. *European Psychiatry*, *23*(4), 289–299. https://doi.org/10.1016/j.eurpsy.2007.05.006
Stankov, L., & Crawford, J. D. (1996). Confidence judgments in studies of individual differences. *Personality and Individual Differences*, *21*(6), 971–986. https://doi.org/10.1016/S0191-8869(96)00130-4
Stankov, L., Lee, J., Luo, W., & Hogan, D. J. (2012). Confidence: A better predictor of academic achievement than self-efficacy, self-concept and anxiety? *Learning and Individual Differences*, *22*(6), 747–758. https://doi.org/10.1016/j.lindif.2012.05.013
Steele, C. M., & Aronson, J. (1995). Stereotype Threat and the Intellectual Test Performance of African Americans. *Journal of Personality and Social Psychology*, *69*(5), 797–811. https://doi.org/10.1037/0022-3514.69.5.797
Steffe, L. P., & Gale, J. (2012). *Constructivism in Education*. Taylor & Francis. Retrieved from http://books.google.es/books?id=f5Up2BpqdSsC
Stoloff, R. H., Taylor, J. A., Xu, J., Ridderikhoff, A., & Ivry, R. B. (2011). Effect of reinforcement history on hand choice in an unconstrained reaching task. *Frontiers in Neuroscience*, *5*, 41. https://doi.org/10.3389/fnins.2011.00041
Stone, R., & Ojika, T. (2000). Virtual heritage: What next? *IEEE Multimedia*, *7*(2), 73–74. https://doi.org/10.1109/93.848434
Sullivan, A., & Bers, M. U. (2016). *Girls, Boys, and Bots: Gender*

Differences in Young Children's Performance on Robotics and Programming Tasks. *Journal of Information Technology Education: Innovations in Practice* (Vol. 15). Retrieved from http://www.informingscience.org/Publications/3547

Tammaro, A. M. (2016). Participatory Approaches and Innovation in Galleries, Libraries, Archives, and Museums. *International Information and Library Review*, *48*(1), 37–44. https://doi.org/10.1080/10572317.2016.1146040

Tan, P. H., Ting, C. Y., & Ling, S. W. (2009). Learning difficulties in programming courses: Undergraduates' perspective and perception. In *ICCTD 2009 - 2009 International Conference on Computer Technology and Development* (Vol. 1, pp. 42–46). https://doi.org/10.1109/ICCTD.2009.188

Tarricone, P. (2011). *The taxonomy of metacognition*. *The taxonomy of metacognition.*

Tauber, S. K., & Rhodes, M. G. (2010). Metacognitive errors contribute to the difficulty in remembering proper names. *Memory*, *18*(5), 522–532. https://doi.org/10.1080/09658211.2010.481818

Tavassoli, T., Hoekstra, R. A., & Baron-Cohen, S. (2014). The Sensory Perception Quotient (SPQ): development and validation of a new sensory questionnaire for adults with and without autism. *Molecular Autism*, *5*(1), 29. https://doi.org/10.1186/2040-2392-5-29

Teong, S. K. (2003). The effect of metacognitive training on mathematical word-problem solving. *Journal of Computer Assisted Learning*, *19*(1), 46–55. https://doi.org/10.1046/j.0266-4909.2003.00005.x

Thomas, A. E. (2017). Gender Differences in Students' Physical Science Motivation. *American Educational Research Journal*, *54*(1), 35–58. https://doi.org/10.3102/0002831216682223

Thomas, J. W. (2000). *A review of research on project-based learning*. Retrieved from http://www.autodesk.com/foundation

Thompson, A., Murphy, D., Dell'Acqua, F., Ecker, C., McAlonan, G., Howells, H., … Lombardo, M. V. (2017). Impaired Communication Between the Motor and Somatosensory Homunculus Is Associated With Poor Manual Dexterity in Autism Spectrum Disorder. *Biological Psychiatry*, *81*(3), 211–219. https://doi.org/10.1016/j.biopsych.2016.06.020

Thuneberg, H. M., Salmi, H. S., & Bogner, F. X. (2018). How creativity, autonomy and visual reasoning contribute to cognitive learning in a STEAM hands-on inquiry-based math module. *Thinking Skills and Creativity*, *29*(July), 153–160. https://doi.org/10.1016/j.tsc.2018.07.003

Tiberius, R. G., & Billson, J. M. (1991). The social context of teaching and learning. *New Directions for Teaching and Learning*, *1991*(45), 67–86. https://doi.org/10.1002/tl.37219914509

Tilden, F. (1977). *Interpreting Our Heritage*. The university of North Carolina press: Chapel hill.

Tunbridge, J. E., & Ashworth, G. J. (1996). *Dissonant heritage : the management of the past as a resource in conflict*. Chichester, New York: J. Wiley.

Turing, a. M. (1950). I.—Computing Machinery and Intelligence. *Mind*, *LIX*(236), 433–460. https://doi.org/10.1093/mind/LIX.236.433

UNESCO. (2015). *Incheon Declaration—Education 2030: Towards inclusive and equitable quality education and lifelong learning for all. World Education Forum*. *UNESCO*. Retrieved from http://uis.unesco.org/en/document/education-2030-incheon-declaration-towards-inclusive-equitable-quality-education-and

Vaghi, M. M., Luyckx, F., Sule, A., Fineberg, N. A., Robbins, T. W., & De Martino, B. (2017). Compulsivity Reveals a Novel Dissociation between Action and Confidence. *Neuron*, *96*(2), 348-354.e4. https://doi.org/10.1016/j.neuron.2017.09.006

Valentine, J. C., DuBois, D. L., & Cooper, H. (2004, March). The relation between self-beliefs and academic achievement: A meta-analytic review. *Educational Psychologist*. Lawrence Erlbaum Associates, Inc. https://doi.org/10.1207/s15326985ep3902_3

van Kesteren, M. T. R., Krabbendam, L., & Meeter, M. (2018). Integrating educational knowledge : Reactivation of prior knowledge during educational learning enhances memory integration. *Npj Science of Learning*, (January), 1–29. https://doi.org/10.1038/s41539-018-0027-8

van Loon, M., de Bruin, A., Leppink, J., & Roebers, C. (2017). Why are children overconfident? Developmental differences in the implementation of accessibility cues when judging concept learning. *Journal of Experimental Child Psychology*,

158, 77–94. https://doi.org/10.1016/j.jecp.2017.01.008
van Rijn, H., van Someren, M., & van der Maas, H. (2003). Modeling developmental transitions on the balance scale task. *Cognitive Science*, *27*(2), 227–257. https://doi.org/10.1016/S0364-0213(02)00116-7
van Swieten, L. M., van Bergen, E., Williams, J. H. G., Wilson, A. D., Plumb, M. S., Kent, S. W., & Mon-Williams, M. A. (2010). A Test of Motor (Not Executive) Planning in Developmental Coordination Disorder and Autism. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(2), 493–499. https://doi.org/10.1037/a0017177
Vanlehn, K., Siler, S., Murray, C., Yamauchi, T., & Baggett, W. B. (2003). Why Do Only Some Events Cause Learning During Human Tutoring? *Cognition and Instruction*, *21*(3), 209–249. https://doi.org/10.1207/S1532690XCI2103_01
Varney, M. W., Janoudi, A., Aslam, D. M., & Graham, D. (2012). Building Young Engineers: TASEM for Third Graders in Woodcreek Magnet Elementary School. *IEEE TRANSACTIONS ON EDUCATION*, *55*(1). https://doi.org/10.1109/TE.2011.2131143
Veenman, M. V. J. (2011). Learning to Self-Monitor and Self-Regulate. In *Handbook of Research on Learning and Instruction*. Routledge. https://doi.org/10.4324/9780203839089.ch10
Veenman, M. V. J., Van Hout-Wolters, B. H. a M., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. *Metacognition and Learning*, *1*(1), 3–14. https://doi.org/10.1007/s11409-006-6893-0
Veenman, M. J. V. (2013). Training metacognitive skills in students with availability and production deficiencies. In *Applications of self-regulated learning across diverse disciplines* (pp. 299–324).
Verschure, P. (2012). Distributed Adaptive Control: A theory of the Mind, Brain, Body Nexus. *Biologically Inspired Cognitive Architectures*. Retrieved from http://www.sciencedirect.com/science/article/pii/S2212683X12000102
Verschure, P. F. M. ., Pennartz, C., & Pezzulo, G. (2014). The Why, What, Where, When and How of Goal Directed Choice:

neuronal and computational principles. *Philosophical Transactions of The Royal Society - Biological Sciences, RSTB-2013*-. https://doi.org/10.1098/rstb.2013.0483
Verschure, P. F. M. J., Voegtlin, T., & Douglas, R. J. (2003). Environmentally mediated synergy between perception and behaviour in mobile robots. *Nature*, *425*(6958), 620–624. https://doi.org/10.1038/nature02024
Veruggio, G., & Operto, F. (2008). Roboethics: Social and ethical implications of robotics. In *Springer Handbook of Robotics* (pp. 1499–1524). Springer.
Veruggio, Gianmarco, Operto, F., & Bekey, G. (2016). Roboethics: Social and ethical implications. In *Springer Handbook of Robotics* (pp. 2135–2160). Springer International Publishing. https://doi.org/10.1007/978-3-319-32552-1_80
Vouloutsi, V., Blancas, M., Zucca, R., Omedas, P., Reidsma, D., Davison, D., … Verschure, P. F. M. J. (2016a). Towards a synthetic tutor assistant: The EASEL project and its architecture. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 9793, pp. 353–364). https://doi.org/10.1007/978-3-319-42417-0_32
Vouloutsi, V., Blancas, M., Zucca, R., Omedas, P., Reidsma, D., Davison, D., … Verschure, P. F. M. J. (2016b). Towards a synthetic tutor assistant: The EASEL project and its architecture. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 9793*, 353–364. https://doi.org/10.1007/978-3-319-42417-0_32
Vouloutsi, V., Munoz, M. B., Grechuta, K., Lallee, S., Duff, A., Llobet, J. P., & Verschure, P. F. (2015). A new biomimetic approach towards educational robotics: the Distributed Adaptive Control of a Synthetic Tutor Assistant. *New Frontiers in Human-Robot Interaction*, 22. Retrieved from https://www.cs.kent.ac.uk/events/2015/AISB2015/proceedings/hri/20-Vouloutsi-anewbiomimetic.pdf
Voyer, D., & Voyer Susan D., D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin*, *140*(4), 1174–1204. https://doi.org/10.1037/a0036620
Vygotsky, L.S., & Cole M. (1978). *Mind in Society: Development*

of Higher Psychological Processes. Harvard University Press.
Vygotsky, Lev Semenovich. (1980). *Mind in society: The development of higher psychological processes*. Harvard university press.
Wadsworth, B. J. (1996). *Piaget's theory of cognitive and affective development : foundations of constructivism* (5th ed.). White Plains N.Y.: Longman Publishers USA. Retrieved from http://www.worldcat.org/title/piagets-theory-of-cognitive-and-affective-development-foundations-of-constructivism/oclc/32310863
Wakkary, R., Muise, K., Tanenbaum, K., Hatala, M., & Kornfeld, L. (2008). Situating approaches to interactive museum guides. *Museum Management and Curatorship*, *23*(4), 367–383. https://doi.org/10.1080/09647770802517423
Wandersee, J. H., Mintzes J, J., & Novak D, J. (1993). Chapter 5: Research on Alternative Conceptions in Science. In *Handbook of Research on Science Teaching and Learning A Project of the National Science Teachers Association* (pp. 177–210). Retrieved from https://studylib.net/doc/5365768/research-on-alternative-conceptions-of-science
Wang, J., Korczykowski, M., Rao, H., Fan, Y., Pluta, J., Gur, R. C., … Detre, J. A. (2007). Gender difference in neural response to psychological stress. *Social Cognitive and Affective Neuroscience*, *2*(3), 227–239. https://doi.org/10.1093/scan/nsm018
Wang, M. C., Haertel, G. D., & Walberg, H. J. (1997). What Influences Learning? A Content Analysis of Review Literature. *The Journal of Educational Research*.
Wang, S. S. H., Kloth, A. D., & Badura, A. (2014). The Cerebellum, Sensitive Periods, and Autism. *Neuron*, *83*(3), 518–532. https://doi.org/10.1016/j.neuron.2014.07.016
Warren, R. D. (2012). *The effect of adaptive confidence strategies in computer-assisted instruction on learning and learner confidence*. Retrieved from http://diginole.lib.fsu.edu/etd
Weller, J. A., Levin, I. P., & Bechara, A. (2010). Do individual differences in Iowa Gambling Task performance predict adaptive decision making for risky gains and losses? *Journal of Clinical and Experimental Neuropsychology*, *32*(2), 141–150. https://doi.org/10.1080/13803390902881926
Wigfield, A., & Eccles, J. S. (1989). Test Anxiety in Elementary

and Secondary School Students. *Educational Psychologist*, *24*(2), 159–183. https://doi.org/10.1207/s15326985ep2402_3
Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology*, *25*(1), 68–81. https://doi.org/10.1006/ceps.1999.1015
Wigfield, A., Harold, R. D., Freedman-Doan, C., Eccles, J. S., Yoon, K. S., Arbreton, A. J. A., & Blumenfeld, P. C. (1997). Change in Children's Competence Beliefs and Subjective Task Values Across the Elementary School Years: A 3-Year Study. *Journal of Educational Psychology*, *89*(3), 451–469. https://doi.org/10.1037/0022-0663.89.3.451
Wilkinson, D. A., Best, C. A., Minshew, N. J., & Strauss, M. S. (2010). Memory awareness for faces in individuals with autism. *Journal of Autism and Developmental Disorders*, *40*(11), 1371–1377. https://doi.org/10.1007/s10803-010-0995-x
Wing, J. M. (2006). Computational Thinking. *Communications of the ACM*, *49*(3), 33–35. Retrieved from http://www.cs.cmu.edu/~wing/
Wing, J. M. (2008). Computational thinking and thinking about computing. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *366*(1881), 3717–3725. https://doi.org/10.1098/rsta.2008.0118
Witherspoon, E. B., Higashi, R. M., Schunn, C. D., Baehr, E. C., & Shoop, R. (2017). Devel-oping Computational Thinking Through a Virtual Robotics Programming Curriculum. *ACM Trans. Comput. Educ*, *18*(4). https://doi.org/10.1145/3104982
Woolfolk, A. (2013). *Educational Psychology*. Pearson Education UK. Retrieved from https://books.google.es/books?id=DBCpBwAAQBAJ&dq=Woolfolk,+A.+(2013).+Educational+psychology&hl=es&sa=X&ved=0ahUKEwiB-bHUio3eAhWl4YUKHdOrDyoQ6AEIQjAD
Yeager, D. S., Hanselman, P., Walton, G. M., Murray, J. S., Crosnoe, R., Muller, C., … Dweck, C. S. (2019). A national experiment reveals where a growth mindset improves achievement. *Nature*, *573*(7774), 364–369. https://doi.org/10.1038/s41586-019-1466-y
Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of

stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, *18*, 459–482. https://doi.org/10.1037/h0073415
Yuan, J., Luo, Y., Yan, J. H., Meng, X., Yu, F., & Li, H. (2009). Neural correlates of the females' susceptibility to negative emotions: An insight into gender-related prevalence of affective disturbances. *Human Brain Mapping*, *30*(11), 3676–3686. https://doi.org/10.1002/hbm.20796
Zhang, J. X., Liu, L., de Pablos, P. O., & She, J. H. (2014). The Auxiliary Role of Information Technology in Teaching: Enhancing Programming Course Using Alice. *International Journal of Engineering Education*, *30*(3), 560–565. Retrieved from http://or.nsfc.gov.cn/bitstream/00001903-5/356833/1/1000013384665.pdf
Zimmerman, B J. (2000). Self-efficacy: An essential motive to learn. *Contemp Educ Psychol*, *25*(1), 82–91. https://doi.org/10.1006/ceps.1999.1016
Zimmerman, Barry J. (2000). Self-Efficacy: An Essential Motive to Learn. *Contemporary Educational Psychology*, *25*(1), 82–91. https://doi.org/10.1006/ceps.1999.1016
Zimmerman, Barry J. (1989). A Social Cognitive View of Self-Regulated Academic Learning. *Article in Journal of Educational Psychology*. https://doi.org/10.1037/0022-0663.81.3.329