PhD-FHSE-2020-31
The Faculty of Humanities, Education and Social Sciences

DISSERTATION

Defence held on 23/10/2020 in Esch-sur-Alzette

to obtain the degree of

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG

EN SCIENCES DE L'ÉDUCATION

by

Carole DORDING

Born on 14th of July 1977 in Pétange (Luxembourg)

GeoGebraTAO: Geometry Learning Using a Dynamic Adaptive ICT-Enhanced Environment to Promote Strong Differentiation of Children's Individual Pathways

Dissertation defence committee

Dr. Charles Joseph MAX, dissertation supervisor
Professor, Université du Luxembourg

Dr. Till DEMBECK, Chairman
Associate Professor, Université du Luxembourg

Dr. Thibaud LATOUR, Vice Chairman
Head at Luxembourg Institute of Science and Technology (LIST)

Dr. Silke LADEL
Professor, Pädagogische Hochschule Schwäbisch Gmünd

Univ. Prof. DI Mag. Dr. Dr.h.c. Markus HOHENWARTER
Professor, Johannes Kepler Universität, School of Education, Linz

Abstract

In our project, we investigate the scientific validity of a specific self-built Adaptive Learning Tool in the field of dynamic geometry with a focus on the individual learning pathways of a highly diverse student population.

A total of 164 children in Luxembourgish elementary schools, aged between 10 and 13 years, acted as the test group and explored elementary geometric concepts in a sequence of learning assignments created with GeoGebra, a dynamic mathematics system which is integrated into the computer-assisted testing framework TAO. They actively built new knowledge in an autonomous way and at their own pace with only minor support interventions by their teacher.

Based on easily exploitable data collected within a sequence of exploratory learning assignments, the GEOGEBRATAO tool analyses the answers provided by the child and performs a diagnostic of the child's competencies in geometry. With respect to this outcome, the tool manages to identify children struggling with geometry concepts and subsequently proposes a differentiated individual pathway through scaffolding and feedback practices. The children can voluntarily watch short video clips aimed to help them better understand any task that they might have difficulty with. A spaced repetition feature is another highly useful component of the tool.

Pre- and post-test results show that the test group (working with the GEOGEBRATAO tool) and a parallel working control group (following a traditional paper-and-pencil geometry course), increased their geometry skills and knowledge through the training program, with the test group performing even better in items related to dynamic geometry. In addition, a more precise analysis within clusters based on similar performances in both pre- and post-tests and the child's progress within the GEOGEBRATAO activities, provides evidence of some common ways of working with our educational technology tool, leading to overall improvement at an individualized level.

Contents

Abstract

Acknowledgements

1 Motivations 1
- 1.1 Springboard for my dissertation topic . . . 1
- 1.2 My personal view of learning and teaching . . . 2
- 1.3 Highly diverse student population . . . 3
- 1.4 Searching for a solution . . . 5
- 1.5 Questions leading the research . . . 6

2 Description of the instruments 7
- 2.1 Pre-and post-test . . . 7
- 2.2 GEOGEBRATAO software . . . 8
  - 2.2.1 General overview . . . 8
  - 2.2.2 Design and layout of the user interface . . . 10
  - 2.2.3 Different kinds of task accomplishments . . . 12
    - 2.2.3.1 Accomplishment through actions in the GeoGebra frame (device A) . . . 12
    - 2.2.3.2 Accomplishment through choices on Likert scales (device B) . . . 14
    - 2.2.3.3 Accomplishment through choices in drop-down menus (device$C$) . . . 15
  - 2.2.4 Integrated 5-in-a-row principle . . . 15
  - 2.2.5 Personalizing the software . . . 18
  - 2.2.6 Scaffolding and feedback practices . . . 19
  - 2.2.7 Video clips . . . 20
  - 2.2.8 Repeating activities . . . 21
  - 2.2.9 Short technical description GeoGebra - TAO . . . 22
  - 2.2.10 Recommendations to teachers for using the GEOGEBRATAO tool in class 23
- 2.3 Instructions for the control classes . . . 24

3 Methodology 27
- 3.1 Sample selection . . . 27
  - 3.1.1 Test group and control group . . . 27
  - 3.1.2 Identification of eligible classes . . . 28
  - 3.1.3 Selection of participating classes . . . 29
  - 3.1.4 Composition of the classes selected . . . 29
  - 3.1.5 Ensuring the quality of the pre-test . . . 30

3.2 Study process and data collection . . . 30
3.2.1 Data collection through the logs of TAO . . . 30
3.2.2 Data collection through the pre- and post-test . . . 30
3.2.3 Auxiliary data: teacher's questionnaire . . . 31
3.2.4 Ethnographic observations in classrooms . . . 32
3.2.5 Restricting post-test data . . . 32
3.3 Hypotheses and their validation . . . 34
3.3.1 Hypothesis (HYP) *one* . . . 34
3.3.2 Hypothesis (HYP) *two* . . . 34
3.3.3 Test of elaborated hypothesis *one* . . . 35
3.3.3.1 Introduction . . . 35
3.3.3.2 Methodology used . . . 35
3.3.3.3 Random coefficients . . . 37
3.3.3.4 Fitting Generalized Mixed-Effects Models . . . 37
3.3.3.5 Specifying random effects structure . . . 38
3.3.3.6 Measurement and insights via Generalized Mixed-Effects Models 40
3.3.4 Test of elaborated hypothesis *two* . . . 42
3.3.4.1 Introduction . . . 42
3.3.4.2 Methodology used . . . 42

4 State of the art / theoretical framework 44
4.1 Broadening the use of ICT . . . 44
4.1.1 Introduction . . . 44
4.1.2 Barriers for the use of ICT . . . 47
4.1.3 Using ICTs in mathematics, specifically GeoGebra in geometry . . . 49
4.1.4 Parallels with the GeoGebraTAO project . . . 51
4.1.5 Gender Differences in ICT use . . . 52
4.1.6 Dynamic Geometry and Computational Thinking (CT) . . . 53
4.1.7 Technology-based assessments, specifically TAO framework . . . 53
4.2 Towards theoretical foundations of the learning approach . . . 54
4.2.1 Teachers' interventions . . . 54
4.2.2 Zone of Proximal Development (ZPD) . . . 55
4.2.3 Scaffolding student learning . . . 56
4.2.4 GeoGebraTAO and blended learning . . . 57
4.2.5 Student-Centered Learning and Classroom . . . 59
4.2.6 Gender Differences in Mathematics . . . 60

5 Field data collection 61
5.1 Written permissions . . . 61
5.2 Data collection from GeoGebraTAO through the logs of TAO . . . 61
5.2.1 Brief review, overview and section purpose . . . 61
5.2.2 Technological aspects . . . 62
5.2.3 Project's logistics, organisational and classroom practice aspects . . . 62
5.2.3.1 Onboarding . . . 62
5.2.3.2 Practice of a first item on choices on Likert scales . . . 63
5.2.3.3 Verbal instructions given to the children and teacher(s) in each class . . . 63
5.2.3.4 Start autonomous work . . . 63

5.2.3.5 Observed problem: lack of watching video clips 64
5.2.4 Teachers' acquiring of expected competences 64
5.2.5 Single table containing data collected from exploratory learning assignments 65
5.3 Data collection through pre- and post-test 66
5.3.1 Detailed encoding of pre- and post-test to categorical data 66
5.3.2 Binary data from the pre- and post-test 66
5.3.3 Restriction of the post-test data set 67
5.3.4 Comparison of the complete binary data set of the post-test and the restricted data set 67

6 Results of the analysis **72**
6.1 Item analysis 72
6.1.1 Pre-test 72
6.1.2 Post-test 74
6.1.3 Comparison between pre- and post-test proportion scores of items having a low *r.drop* value within their domain 75
6.2 Summary of quantitative data at the class level 77
6.2.1 Overview table 77
6.2.2 Summary of the pre-test data set 77
6.2.3 Summary of the post-test_all data set 78
6.2.4 Summary of the post-test_restricted data set 79
6.2.5 Summary of the GEOGEBRATAO tool data set 79
6.2.6 Observations 80
6.3 Test of predicted hypothesis *one* related to the pre- and post-test 81
6.3.1 Comparisons between the pre- and post-test calculated at the children level 81
6.3.1.1 Fitted Generalized Mixed-Effects Model 81
6.3.1.2 Related figure comments and interpretation 83
6.3.1.3 Essential information provided by the Deviance Table 83
6.3.1.4 Fitted model summary statistics and interpretation 84
6.3.1.5 Variability of the random effects of *glmer_Comp_PrePost* . . . 85
6.3.1.6 Summary graph based on factor *Domain* 86
6.3.2 Comparisons between *'gender gaps'* in the pre- and post-test *research question* $RQ1(a)$ *(pg. 6)* 88
6.3.2.1 Fitted Generalized Mixed-Effects Model 88
6.3.2.2 General comments and interpretation 88
6.3.2.3 Interpretation related to the pre-test 89
6.3.2.4 Interpretation related to the progress from pre- to post-test . . . 90
6.3.3 Comparisons between children who *'like'* maths and those who don't *research question* $RQ1(c)$ *(pg. 6)* 93
6.3.3.1 Fitted Generalized Mixed-Effects Model 93
6.3.3.2 General comments and interpretation 93
6.3.3.3 Interpretation related to the pre-test 93
6.3.3.4 Interpretation related to the progress from pre- to post-test . . . 94
6.3.4 Conclusion regarding hypothesis *one* *(in answer to research questions* $RQ1$ *and* $RQ2$ *(pg. 6))* 97
6.4 Test of predicted hypothesis *two* 98
6.4.1 Cluster analysis using the 'complete' binary post-test data set 98
6.4.1.1 Procedure of `PAM` cluster analysis 98

6.4.1.2 Summary of the clusters obtained through PAM cluster analysis 99
6.4.1.3 Visualizations of the performance groups 100
6.4.2 Cluster analysis using the 'restricted' binary post-test data set 102
6.4.2.1 Procedure of PAM cluster analysis 102
6.4.2.2 Summary of the clusters obtained through PAM cluster analysis 102
6.4.2.3 Visualizations of the performance groups 104
6.4.3 Cluster-based comparisons in relation to the different help modi offered by the GEOGEBRATAO tool 106
6.4.3.1 Introduction 106
6.4.3.2 Fitted Generalized Mixed-Effects Models for the *short* video clips 107
6.4.3.3 Analysis and interpretation of the *short* video clips at *Cluster* level 107
6.4.3.4 Summary table 111
6.4.3.5 Analysis and interpretation of the *short* video clips at *Class* level 112
6.4.3.6 Fitted Generalized Mixed-Effects Model for the *overlay* videos 114
6.4.3.7 Analysis and interpretation of the *overlay* videos at the *Cluster* level 116
6.4.3.8 Summary table 118
6.4.3.9 Fitted Generalized Mixed-Effects Model for the *repetition* of prior learning assignments 119
6.4.3.10 Analysis and interpretation of loops and repetitions of prior learning assignments at *Cluster* level 119
6.4.3.11 Summary table 122
6.4.3.12 Fitted Generalized Mixed-Effects Model related to the message to call the teacher 123
6.4.3.13 Analysis and interpretation of the *Call-Teacher-Messages* at *Cluster* level 123
6.4.3.14 Summary table 124
6.4.4 Conclusion on hypothesis *two* *(in answer to research question RQ3 (pg. 6))* 125

7 Barriers to GEOGEBRATAO implementation 130
7.1 Technology barriers 130
7.2 Organizational barriers 131
7.3 Human- and teacher-related barriers 132
7.4 iPad-related barriers 132

8 Conclusion 134

Bibliography 135

A Socioeconomic index by Commune in 2017 161

B Proportion of foreigners per Commune 2018 163

C Children's orientation after fundamental school 164

D Workflows defined by the project GEOGEBRATAO 167

E Reflected post-test data and fitting its distribution 168

F Outputs of model $glmer\_Comp\_Post$ 169
F.1 Analysis of Deviance Table 169
F.2 Summary of fitted model 169

G Outputs of model $glmer\_Comp\_PrePost$ 171
G.1 Analysis of Deviance Table 171
G.2 Summary of fitted model 171

H Outputs of model $glmer\_Comp\_PrePost\_Gen$ 173
H.1 Analysis of Deviance Table 173
H.2 Summary of fitted model 173

I Outputs of model $glmer\_Comp\_PrePost\_LikeM$ 176
I.1 Analysis of Deviance Table 176
I.2 Summary of fitted model 176

J Outputs of model $glmmPQL\_shVid$ 179
J.1 Analysis of Deviance Table 179
J.2 Summary of fitted model 179

K Outputs of model $glmmPQL\_shVid$ **at *Class* level** 180
K.1 Analysis of Deviance Table 180
K.2 Summary of fitted model 180

L Outputs of model $glmmPQL\_ovlayVid$ 182
L.1 Analysis of Deviance Table 182
L.2 Summary of fitted model 182

M Outputs of model $glmmPQL\_repeat$ 183
M.1 Analysis of Deviance Table 183
M.2 Summary of fitted model 183

N Outputs of model $glmmPQL\_mes$ 184
N.1 Analysis of Deviance Table 184
N.2 Summary of fitted model 184

O Entire pre- and post-test 185

P Teacher's questionnaire 217

Q List of the GEOGEBRATAO tool exploratory learning assignments 222

Acknowledgements

Throughout my work on this PhD thesis, I have received a great deal of support and assistance from my thesis committee, family, colleagues and friends. The work presented in this thesis would not have been possible without them.

I would first like to thank my supervisor, **Prof. Dr. Charles Joseph MAX**, who accepted the challenge of guiding me throughout the writing of my thesis over the last year. You have been a tremendous mentor for me and your insightful feedback encouraged me to sharpen my research thinking and writing, thereby elevating my work. Your immense knowledge was invaluable in re-formulating the research questions and deepening the related discussions. I must also express gratitude to **Prof. Dr. Romain MARTIN**, my first supervisor, for enabling this research work and sharing insightful suggestions. It has been a privilege to work with you both.

Dr. Till DEMBECK, thank you for your kindness and help during the last year of my thesis and for spontaneously accepting to take over the task of presiding the Jury.

Apart from my supervisors, I would like to thank my committee members (CET), **Dr. Yves KREIS** and **Dr. Thibaud LATOUR**, for giving the personal and professional support during this project. You provided me with the tools that I needed to successfully complete my dissertation. Yves, I want to thank you for introducing me to this exciting field of science education in primary grades and for lending your help, advice and inspiration throughout my PhD. Thibaud, I am grateful that you enabled cooperation with The Luxembourg Institute of Science and Technology (LIST) and for your encouragement and spontaneous support. You have both played a major role in this work.

Prof. Dr. Silke LADEL and **Univ. Prof. DI Mag. Dr. Dr.h.c. Markus HOHENWARTER**, thank you for taking time out of your busy schedules to read this dissertation and to serve on my dissertation jury. I eagerly anticipate your feedback and hope to enter into further research discussion in the future.

I am also pleased to thank **Olivier PEDRETTI**, engineer from The Luxembourg Institute of Science and Technology, without whom this work would not have been possible. Olivier, I will never forget your valuable and constant technical support, advice and approach towards building our tool. It was a pleasure to work with you during this experience.

Another outstanding member to whom I am more than grateful, **Ulrich KELLER**, provided clear and careful statistical advice. Uli, thanks for sharing your vast knowledge of statistics and in particular for suggesting the statistical model we used. This brilliant idea helped us to optimally present our results and to make data-driven decisions.

Furthermore, I am profoundly thankful for the improvement suggestions of my friend and copy editor, Kate MULLANEY SHIRLEY, and her thorough contribution to the writing of this thesis and the preparation of the defense. Kate, I am delighted to have worked with you and I look forward to working with you again when publishing articles based on this project.

I must also acknowledge Sarah SOBOLE, an American cartoonist and storyboard artist, who helped us to personalize our learning tool. Sarah, the children enjoyed the funny, unique, eye-catching *'cat'* mascot you created. Thank you for making the activities more child-friendly.

I give my special regards to the teachers and children who patiently took part in this research project. I am thankful we walked together on this path. Without permission from the Ministry of Education, Children and Youth, from the directorate of fundamental education, from the communes, and from all participating children's parents, this work would not have been realizable. I am thankful to all of you for such excellent cooperation that allowed me to progress in my pedagogical approach.

Fortunately, I am also privileged to have lovely family and friends who gave me enough moral support, encouragement and motivation to accomplish the personal goals; I'm particularly grateful to my little niece, Laura, who provided happy distractions to rest my mind outside of my research. I would like to thank my parents, who have always been there with a sympathetic ear no matter how busy life is.

In the end, I want to thank all those who supported me in some way, either by providing stimulating discussions, exchanging views, showing interest in our study or simply giving some motivational words; I mention in particular Dr. Catalina LOMOS, Louise Marie STEVENSON, Dr. Bob REUTER, Ben HAAS, Alex CARROLL and my father's friends Renée and Jos SCHOU.

Chapter 1

Motivations

1.1 Springboard for my dissertation topic

The research object of this study evolved from prior research work, beginning with the MATES project during my pedagogical education and further strengthened by the GEOGEBRAPRIM project at the University of Luxembourg, on which I worked as a research associate.

From 2005 through 2007, I worked on MATES, an e-Learning tool on fractions in mathematics. MATES allows students to enter complete questions in natural language, and returns semantically pertinent multimedia results, that explain the answer to the users' questions (Linckels and Meinel, 2011; Linckels et al., 2006, 2007). Relevant improvements were measured in the students' school results over the period during which the students used MATES in an autonomous and explorative way, compared to the school results before using the tool. One of the main reasons for this excellent result may be that the students were more motivated, and therefore put more effort into learning mathematics and acquiring new knowledge.

The MATES project greatly interested me and I was pleased to be involved later in another multimedia project, GEOGEBRAPRIM. This project, which started in January 2007, investigated the use of GeoGebra, an interactive geometry software for learning and teaching mathematics, in primary grade (Kreis et al., 2010a). I considered it as an opportunity allowing me to scrutinize the advantages of GeoGebra, as well as its integration into the computer-assisted testing framework TAO (Latour and Martin, 2007; Ras et al., 2010; Csapó et al., 2012); TAO strives to combine assessment and learning data, and thereby to implement adaptive learning solutions (TAOtesting, 2019; Kreis et al., 2018; Plichart et al., 2008).

Two classes followed a traditional paper-and-pencil geometry course (control group) while the others combined traditional and computer-based learning (test group). The children who were active in their learning process and who explored the geometric concepts (such as parallelism and perpendicularity, area and perimeter calculating, to name a few) on the computer (test group) had better results than those from the control group.

As a research associate on the GEOGEBRAPRIM project, I taught geometry in one of the test classes; there from I got some fixed ideas of this PhD Thesis. As a teacher, I supervised the children and assisted them by giving relaunch instructions so they could enrich their answers and further explore the relevant concepts. I intervened and provided scaffolding (Wood et al., 1976) or feedback (Hattie and Timperley, 2007) to the children's processes of knowledge building, meaning making and spatial thinking. These interventions were a constitutive element of the learning practice and raised a series of interesting pedagogical questions for me regarding mathematics education. The project's value was strengthened by my observation of improvements in the children's understanding of the elementary geometric concepts and further insight into the connection between geometry and algebra.

This existing body of research based at the University of Luxembourg constitutes a promising springboard for the thorough analysis of computer-assisted learning of primary-level dynamic-geometry activities, and thus an ideal springboard for my dissertation topic. It is my belief that the ascertained reasons for the test group's greater success compared to the control group are not satisfying and require further research.

1.2 My personal view of learning and teaching

My research focuses on teacher intervention in the mathematics classroom. However, teaching interventions are highly personalized and oscillate between a teacher-centered approach with a strong focus on transmission processes and a student-centered approach engaging children in processes of participation, dialogue and collaboration.

So, teachers have to find a personal balance between these polarities: teaching explicit methods to their children that they have to apply and encouraging their children to build and use methods and mathematical ideas that they have detected in a largely self-directed manner.

Personally, I support a constructivist learning approach that encourages children to learn new subject matter by building on prior knowledge. Scaffolding and feedback practices are essential in such an approach to stimulate and support processes of knowledge building, meaning making and spatial thinking. However, we have to critically reflect on how exactly to provide successive levels of temporary support and regular feedback during activities without communicating knowledge *directly*:

> *"If both the problem and the information about its solution are communicated by the teacher this deprives the pupil of the conditions necessary for learning and understanding. The pupil will only be able to reproduce the method of handling and solving the problem communicated to her. ··· mathematics is not just a method." (Brousseau and Otte, 1991, pg. 121), as cited in (Johnston-Wilder and Mason, 2004, pg. 82)*

Nor can we opt for letting children reconstruct ideas exclusively for themselves (Falcade, 2006):

> *"If you think that learners ultimately have to reconstruct ideas for themselves, how do you arrange for learners to reconstruct accurately and appropriately without 'telling them'?" (Johnston-Wilder and Mason, 2004, pg. 82)*

In this way, the teachers *"··· have a responsibility to direct and shape the learning opportunities ··· " (Speer and Wagner, 2009, pg. 530)* of their children.

Responding to children without communicating mathematical knowledge *directly* may present a real challenge to teachers. Figuring out how to flexibly help children clarify their own thinking by connection their prior ideas to disciplinary ways of thinking requires a great deal of time and preparation. This dynamic intervention based on prior knowledge, called scaffolding, builds on children's responses with the goal of improving their knowledge and ability in the subject matter. Learning through scaffolding is a *"··· well known approach to facilitating learning, especially the acquisition of and reasoning with abstract concepts ··· " (Wood et al., 1976)*, in which various kinds of learning materials may be used to help children accomplish tasks that initially lie beyond their ability (Masterman and Rogers, 2002). The scaffolds should be short and clear; above all, they should acclimate children to the subject matter. Therefore the teachers' capacity to enact these component practices is of great importance.

According to Judy Olson and Jennifer Platt (2000), assisted activities provided by the teacher related to the new mathematical ideas should be just beyond the level of what the learner can do alone. The children should be able to accomplish (with assistance, as described above) mathematical tasks that they could otherwise not successfully complete. Lew Vygotsky called this learning the *"zone of proximal development (ZPD)"* (Van Der Stuyf, 2002); that is, the distance between what a child can master without help and what they can only achieve with accurate support from a more knowledgeable peer or adult. This development concerns any emotional, cognitive and volitional processes, and might occur in classroom or outdoor projects, oral or writing tasks, computer-based activities, and so on.

1.3 Highly diverse student population

Besides these projects, which have allowed me to gain fundamental knowledge in my chosen field of research, I have also acquired broad experience as a teacher in various secondary and primary school classes in Luxembourg. Thus I am well aware to what extent teachers' beliefs and interventions described in section 1.2 shape learning practices in an environment where external elements impact and complicate the learning-teaching process, e.g. the socioeconomic backgrounds of the children.

The socioeconomic backgrounds of Luxembourg's children (cf. appendix A) are closely linked to the language spoken at home, and of course to any migration background (Fischbach and Hentschel, 2019):

"Wir haben erstens ein generelles Leistungsproblem, das sich durch alle Bereiche zieht und das sich in erster Linie damit erklären lässt, dass wir extreme Disparitäten zwischen unterschiedlichen Subgruppen im Land haben. Das gibt es überall, aber nirgendwo sonst ist die Verbindung zwischen dem sozioökonomischen Hintergrund und dem Einfluss auf die Leistung so ausgeprägt wie in Luxemburg.	**translation:** First of all, there is a general performance issue present in all the domains, that mainly can be explained by the extreme disparities existing between diverse subgroups in our country. Well, that's found worldwide, but nowhere is the relationship between socioeconomic background and impact on performance as significant as in Luxembourg.

· · · Das heißt, dass bei den Leistungen zwischen den 15-Jährigen aus der oberen und denen aus der unteren sozioökonomischen Schicht quasi drei Schuljahre liegen (im Bereich der Lesekompetenz)." (Fischbach and Hentschel, 2019)

· · · This means that, between 15-year-olds of the most advantaged socioeconomic class and those of the most disadvantaged socioeconomic class, there is a difference of nearly three school years (regarding reading skills).

More precise information can be retrieved from two independent studies conducted by STATEC, the National Institute of Statistics and Economic Studies: one on socioeconomic backgrounds of Luxembourg's population and another on foreigners living in the Grand Duchy of Luxembourg:

1. STATEC established an index, composed of a series of socioeconomic variables, to measure social inequality trends at the municipal level. This index is not designed to stigmatize any communes, but to contribute to defining economic and social policies in the future. It is based on the $five$ following socioeconomic variables: 1) the proportion of single-parent households among all the households, 2) the median salary per commune, 3) the proportion of people benefiting from REVIS (social inclusion income designed to help households in the lowest income brackets), 4) the unemployment rate and 5) the proportion of residents having a job and working in low-level professional fields (having a preliminary technical and vocational certificate, CITP). This index is computed according to the methodology used by the United Nations to calculate their human development index. (cf. appendix A)

2. The number of foreigners living in the Grand Duchy of Luxembourg is still rising; residents with a foreign nationality represented about $47,9\%$ of its inhabitants as of the 1st of January 2018. (cf. appendix B)

By analysing these data, i.e. maps and stats, we see that some communes have a socioeconomic index close to one, which means that they are in a *"more socioeconomically disadvantaged situation"* (STATEC, 2017) (at the national level); furthermore, these same communes have a higher proportion of foreigners compared to others.

I taught in two of these communes as a fundamental school teacher and observed that teachers and schools struggle with a variety of social and educational problems. This is due mainly to the complex language situation, the time constraints of the parents balancing work and family life, the family structure and family size. Teachers encounter difficulties delivering the curriculum because of the discrepancy in level within their class population; students in the same class may be one, two or more school years apart in ability. However, fundamental school classes also include high performing children, who must be taught at an adequate level to their educational abilities.

The outcomes of fundamental education in these communes display these difficulties. They show a massive orientation of low-performing students to general secondary education and more specifically to the preparatory route (see appendix C). Often this latter proportion of children is even higher than in all other communes in the country.

It is useful to keep the specificity of this highly diverse student population in mind while reading this thesis. Unfortunately, we were not able to include socioeconomic variables in our statistical analysis due to lack of valid data. The description above is based on data available only at the general commune/municipality level, but not at the specific school or even class level. Data protection regulations make it difficult to collect personal information about the participating children, particularly data concerning their socioeconomic background. Several authorizations and permissions would have been needed from the ministry and from the parents, an insurmountable obstacle for this scientific study.

1.4 Searching for a solution

My objective from the beginning of my research on was and still is to develop a pedagogical approach that allows teachers to address the (wide range of) ability levels in the classroom and ensures that all children are given the same opportunities to actively participate in this learning environment according to their individual needs (Michigan Virtual University, n.d.) and background knowledge. Moreover, children should feel welcomed, motivated, challenged and supported to master required content. Ideally, children would be able to regulate their own learning via a broad range of encouragement and feedback. As it is not possible to multiply the teaching staff within classrooms to reach all ability levels for financial and staffing reasons, a different solution is necessary to free up teaching time while using available resources and to allow personalized learning paths.

While reflecting on this problem, I was struck with the possibility of introducing a *booster* computer software in these classrooms, which would partly fulfill the role of a *twin* teacher and be able to support and assist teachers, as well as increase their effectiveness. This *'booster'* term is analogous to medicine, where a booster dose enhances an effect; in our case, the *booster* software would enhance student-centered learning (Michigan Virtual University, n.d.).

Children should be given the opportunity to drive their own learning by performing a sequence of computer-based activities in an autonomous way, i.e. with as little teacher assistance as possible through scaffolding and clear feedback practices given by the computer software (the *twin* teacher). Operating instructions and response information should be adapted to each child's ability level. The computer should have the ability to assign tasks to the child, interpret the answers given, balance expectations and provide individualized support.

In the meantime, the teachers would be relieved of some classroom duties. While ensuring that all the children, working independently with the computer, are being accurately challenged and dealing with content adequate to their individual abilities and prior knowledge, teachers could devote more time and attention to the learning process of each child. For example, a teacher might assist an individual or group struggling with a particular concept, boost some children's activities, support lower performing children, or if needed, manage other classroom issues. Nevertheless, regular encouragement of the children is absolutely necessary to maintain their engagement, motivation and continuous improvement.

A wonderful saying goes, *"Don't compare your child to others. There's no comparison between the sun and the moon. They shine when it's their time."* A software designed according to the aforementioned criteria wouldn't compare children. It would enable each child to be actively engaged in their own learning process, building new knowledge as autonomously as possible at their own pace. It could progressively complete content-related activities and start a new topic only when the child is prepared - a particular difficulty for teachers with a large class and little support.

At this point, some interesting questions arise which I want to investigate throughout the study.

1.5 Questions leading the research

RQ1: Will we measure varying learning outcomes between children using a software tool as conceived above and children following only a traditional paper-and-pencil course?

Thereupon, several specific subquestions come to mind; for example,

RQ1(a): Do any differences exist among genders?

RQ1(b): Do any differences exist among consecutive classes?

RQ1(c): Do any differences exist among children who like or dislike maths?

RQ1(d): Do any differences exist among cities or children from the same school?

RQ2: Would we find any differences in achievement between a technology-enhanced learning activity and a traditional paper-and-pencil course when comparing the children's results to their outcomes in a pre-test (lower vs higher performing children)?

RQ3: What might we notice about the pupils' ways of working? Can we distinguish between some common ways of working with such a software tool, leading to an overall improvement at an individualized level?

After having

1. explained the chronology of how I arrived at this thesis subject,
2. justified my personal interest and motivation to focus research efforts on this subject and
3. generated the research questions,

I present our research instruments that are being used in studying this topic.

Chapter 2

Description of the instruments

This chapter introduces the $three$ instruments we developed to conduct our research. The two main instruments of our study are the pre- and post-test (1^{st} instrument) for evaluating the software tool called GEOGEBRATAO (2^{nd} instrument), which fosters autonomous and exploratory learning of elementary geometric concepts in a sequence of learning activities [AKTIVITÉIT $1 - 42$]. A $third$ instrument, smaller but of no less importance, is the set of teaching instructions given to the teachers of the control classes on how to conduct their regular classroom activities during elementary geometry lessons.

2.1 Pre-and post-test

When we refer to the pre-test in this section, we also address the post-test as they are strictly identical. Our pre- and post-testing is designed to identify differing starting levels among the children and to measure each child's progress through the specific learning interventions experienced by both test and control groups. This instrument put the children in contact with all geometry topics they would ultimately face in the study-related interventions without echoing the exact same exercises and question-answer combinations.

The pre-test uses different types of exercises with varying levels of difficulty to detect progress and improvement at various levels: a type of gap filling exercise, drawing exercises, true / false exercises, multiple choice exercises and one justifying exercise. Table 2.1 presents an overview of the different exercises [EX. $1 - 18$] (called assignments) and the whole pre- and post-test is appended to this manuscript (cf. appendix O).

The pre-test exclusively invites children to forecast plausible solutions to problems based on prior knowledge without expecting them to be able to provide all answers correctly at this stage. We presume that existing knowledge and understanding of elementary geometry varies widely among the study participants. In previous school years, they participated in nonidentical learning activities taught by different teachers. Some children are able to remember content from these learning activities much better than others in relation to their cognitive strategies, learning capacity and personal involvement (Chi, 1978). To address various competence levels, i.e. to motivate lower performers and to challenge higher performers, our instrument is based on a mix of exercises / items combining easy ones, more difficult ones and some with low discriminating power (extremely easy or difficult).

When students take the same test again at the end of our study (post-test), we expect a higher number of correct answers, i.e. a higher numerical post-score, grounded on an increase in knowledge, ability and understanding.

Taking this into account, our pre- and post-test consists of exercises [EX. $1-18$] with a total of 121 items and contains many pictures and graphic representations. The 121 bilingual items (in German and French) cover $five$ domains: 1) coordinates (24 items), 2) lines and segments (26 items), 3) recognition of symmetry (43 items), 4) drawing of symmetry (20 items) and 5) shapes (8 items). The symmetry topics (domains 3 and 4) of the study number a total of 63 items and constitute the central subject matter investigated. They are the most ambitious elements of the study. To achieve the learning objectives of the symmetry domains, the children must gradually build on their knowledge as they move through the three easier and more basic domains: coordinates, lines and segments, and shapes.

2.2 GEOGEBRATAO software

2.2.1 General overview

The GEOGEBRATAO tool is a specific instrument we developed to foster autonomous and exploratory learning of elementary geometric concepts. It offers a sequence of 90 exploratory learning assignments (activities) grouped in 42 problem sets labelled AKTIVITÉIT 1 to AKTIVITÉIT 42. An overview of the assignments is appended to the manuscript (cf. appendixQ). These activities were created with the dynamic mathematics system GeoGebra and incorporated into the computer-assisted testing framework TAO (Kreis et al., 2010b). This combination is a high guarantor that no technical issues would occur during the test phase in the different classes.

According to Mitchell Kapor's software design manifesto (Kapor, 1990), a software program should be free of bugs that inhibit its function; with this in mind, we want to emphasize that the GEOGEBRATAO tool is firm and reliable. In addition to firmness, good software should also exhibit commodity and delight; this is addressed in further depth below (see subsection 2.2.2). Commodity means that a program should be suitable for the purposes for which it was intended and delight refers to a pleasurable user experience.

GEOGEBRATAO collects data about the performance and progress of each child through the sequence of assignments as it leads children from one learning task to the next - mainly forwards, but with the possibility of repeating prior assignments as needed. Moreover, it consistently offers help in form of feedback, hints, scaffolds and video clips based on recorded user-specific and progress-related data. In some cases, the tool even adds supplementary tasks to facilitate understanding of a certain geometry topic.

Hence, high-performing pupils progress more quickly through the sequence of exploratory learning assignments and trigger less support and remediation. The knowledge and skills they develop about the geometry topics and the learning tool itself enable them to auto-regulate their further advancement through the program. The GEOGEBRATAO tool quickly empowers high-performing pupils to seek individual challenge through level-appropriate geometry tasks in their ordinary school environment.

Table 2.1: Overview of the pre-test / post-test assignments

Assignment	**Domain**	**Number of items**	**Type**
EX. 1	coordinates *(treasure hunt)*	8 items	gap filling exercise *(a letter plus a numeral)*
EX. 2	coordinates	8 items	gap filling exercise *(a numeral plus a numeral)*
EX. 3	coordinates	8 items	drawing exercise *(drawing points on a grid)*
EX. 4	lines and segments	8 items	multiple choice exercise *(in the form of a table, each item has 2 choices)*
EX. 5	lines and segments	8 items	drawing exercise *(drawing lines and segments on a grid)*
EX. 6	lines and segments	2 items	drawing exercise *(completing figures)*
EX. 7	recognizing symmetry *(cartoons)*	4 items	multiple choice exercise *(each item has 2 choices)*
EX. 8	recognizing symmetry *(cartoons)*	12 items	true / false exercise
EX. 9a EX. 9b	shapes recognizing symmetry	8 items 8 items *(same as* EX. 9a*)*	gap filling exercise *(filling in the blanks with the appropriate shape name)* multiple choice exercise *(true / false items)*
EX. 10	recognizing symmetry *(real life picture)*	3 items	gap filling exercise *(writing given terms on the right line)*
EX. 11	recognizing symmetry *(real life pictures)*	4 items	drawing exercise *(drawing after recognizing axis of symmetry)*
EX. 12	recognizing symmetry	4 items	multiple choice exercise *(true / false items in the form of a table)*
EX. 13	recognizing symmetry	8 items	drawing exercise *(- drawing after recognizing axes of symmetry,* *- paired items, one with grid, one without grid)*
EX. 14	drawing symmetry	8 items	drawing exercise *(- practicing symmetry drawings of points, lines and shapes,* *- the origin shape is situated <u>on only one side</u> of the axis of symmetry,* *- paired items: one with grid, one without grid)*
EX. 15 & EX. 16	drawing symmetry	2 items & 2 items	drawing exercise *(- practicing symmetry drawings of points, lines, circles (discs) and shapes,* *- the origin shape is situated <u>on both sides</u> of the axis of symmetry,* *- paired items: one with grid, one without grid,* *- different choice of words in the exercise instructions)*
EX. 17	lines and segments	8 items	multiple choice exercise *(in the form of a table; each item has 3 choices)*
EX. 18a EX. 18b	drawing symmetry drawing symmetry	4 items 4 items *(same as* EX. 18a*)*	multiple choice exercise *(true / false items)* justifying exercise *(justifying correctness in form of a short sentence which may be similar to the given sentence)*

Pupils experiencing greater difficulty grasping a geometric concept get repeated remediation instructions and support from the tool. They are invited to repeat steps or entire activities until they have developed the required knowledge and skills to move forward. Children are given the necessary time and freedom to explore the activities and to revise a topic free of pressure or competitive struggle.

This exploratory learning approach was clearly explained to all the test-class teachers, who first had to get acquainted with the GEOGEBRATAO tool so they could appropriately implement the tool in their lessons. (see below for further details)

2.2.2 Design and layout of the user interface

The design and the layout of the user interface required serious conceptual reflection as they have the ability to maximize the quality of pupils' interactions with and perceptions of the GEOGEBRATAO tool. Interactivity (along with user trust) is of high importance in order to generate accurate affordances for topic exploration and student-centered support among children in the test classes, i.e. to *"provide different types of action possibilities to the child" (Kaptelinin, 2014)*.

> *"The term affordance is broadly linked to an opportunity for action; to afford an action is to facilitate or enable it. Affordances as a concept have been employed in the analysis of game-related phenomena in a variety of different contexts, such as games for education (Linderoth* 2010*, Spires, et al.* 2011*, Meluso* 2012*)··· " (Cardona-Rivera and Young, 2013, pg. 1)*.

This aspect refers to research outcomes of user experience (UX), which aims to understand the particularities of a certain software by investigating the user's needs, abilities, beliefs and motives and by taking into account users' specific limitations.

First, it is essential to our experiment that our software is easy to use. Manipulation features negatively impact students' motivation as soon as the handling of the site is too complex or difficult to apprehend. For this reason, we opted for a simple visual design and developed a consistent layout for recurring elements so that they would be highly intuitive and user-friendly. According to the *usability.gov* site, visual design spotlights *"the aesthetics of a site and its related materials by strategically implementing images, colors, fonts, and other elements"*. The site also states that successful design *"does not take away from the content on the page or function"*, but *"enhances it by engaging users and helping to build trust and interest ··· " (U.S. General Services Administration, n.d.)*. For example, as seen in Figure 2.1, we selected just a few simple colors for the basic layout.

GEOGEBRATAO strategically combines *three* different devices in one digital learning environment, which creates different types of affordances (or action possibilities) for the child (see subsection 2.2.3 for more details). This intentional design decision also relies on Kapor's dimensions of commodity (a program should be suitable for the purposes for which it was intended) and delight (using the program should be a pleasurable experience) (Kapor, 1990).

The basic working screen of the tool is split into *two* main columns presenting the core elements of an exploratory geometry learning assignment [AKTIVITÉIT $1-42$]. The title of the respective activity is displayed in a bar above the two basic columns. The title bar includes a

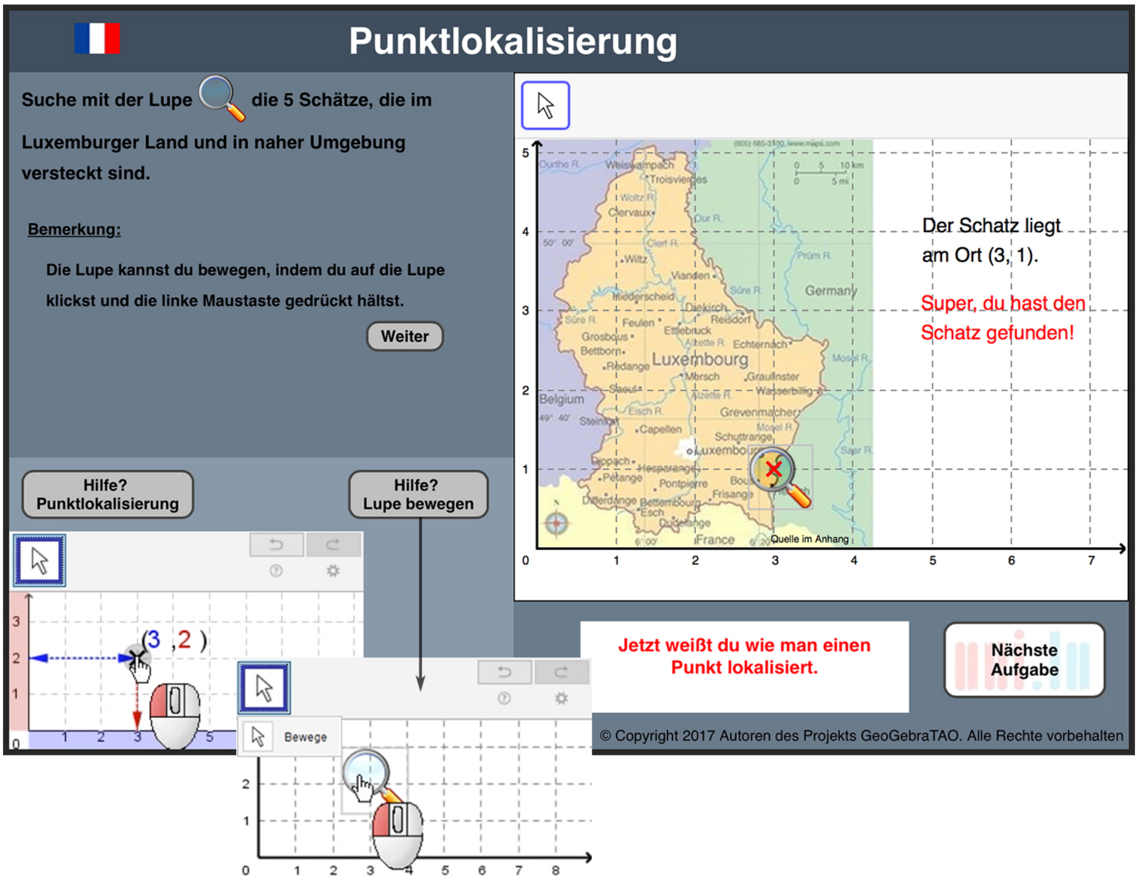


Figure 2.1: User interface design and layout of the software

French flag icon allowing children to switch to the French version of the program once the cursor moves to it. This option was conceived to enable Romance-language speaking children to understand any given assignment.

The upper half of the left column is the block where an assignment's objective and teaching instructions are clearly stated. Adhering to the principle, 'less is more', we tried to keep the text of this block brief and simple to facilitate effortless reading. According to Barbara L. McCombs, providing clear objectives and instructions are essential features for student motivation during any activity (Viau, 2009). Through the sequence of assignments, the children learn to localize and identify the operating directives in the layout screen and perform the necessary steps according to these guiding instructions. For some assignments, further elements might also pop up in this area, such as drop-down menus with multiple-choice options, Likert Scale questions or statements, all based on cursor movements.

The lower half of the left column features a section where the child can get help if needed (on a voluntary basis). We decided to provide visual support to sustain the learning process of GeoGebra. In the assignment displayed in Figure 2.1, help is provided by *two small* video clips in the lower left area (instant aid), one explaining how to localize a point in a coordinate system (mathematical help) and another demonstrating how to the move magnifying glass (technical help). To create these video clips, we used a digital toolkit, which includes screen capture software to record a GeoGebra area, a mouse pointer, and a keyboard action visualisation software. There was no need to add sound to these clips due to environmental classroom noise. Moreover, most of the schools lacked the necessary number of headphones. The children were supposed to watch the video(s) in order to understand and possibly imitate mouse movements and their respective actions in the GeoGebra area. Provided help could also consist of a static picture or a short text explaining a specific concept.

The right column of the basic screen hosts two areas. The GeoGebra area is located in the upper half of the right column. This dynamic worksheet has been developed according to the KISS principle, which stands for *'Keep It Short and Simple'*. It allows children with only rudimentary computer skills to follow given instructions and carry out an operation by modifying a dynamic figure (Hohenwarter et al., 2007; Nguyen, 2017), e.g. to observe the movements of the mirror image by interactively modifying the dynamic original image.

The GeoGebra's toolbar is built up step by step and customized according to each child's needs. It starts with the display of the *Move Tool* which allows the child to drag (and drop) Free Objects, followed by the display of a second tool called the *Point Tool* (which creates a new point), and continues until the inclusion of more complex tools, i.e. the *Relation Tool* or the *Reflect about Line Tool*. Visual scaffolds are also provided by the GeoGebra tool when their inclusion is logical.

The white text box at the bottom of the right column forms the box for feedback, scaffolding and purpose stating. It is quite small, so the text must be short, precise and effective. Next to this box is the button for moving to the next exploratory learning assignment.

2.2.3 Different kinds of task accomplishments

2.2.3.1 Accomplishment through actions in the GeoGebra frame (device A)

The GeoGebra area (device A), displayed in the upper-right part of the computer screen, provides affordances by allowing users to manipulate dynamic objects so they can comfortably explore and grasp mathematical concepts. The device is developed according to the KISS principle (see above) so that a simple instruction stimulates the user to execute an operation such as, for example, modifying a dynamic figure or simply moving a point (Hohenwarter et al., 2007; Nguyen, 2017). The assignment shown in Figure 2.1 deals with locating points in a coordinate system and invites children to find $five$ treasures hidden in the country of Luxembourg and its environs within a limited amount of time. The achievement of this task is facilitated by a magnifying glass (attached to a point) that the child might move over the map (affordance) to identify a point's precise coordinates.

Figure 2.2 shows an assignment in which the children have to properly move a line such that it transforms into the symmetry axis of a real-life, familiar image. Since incorrect moves are possible, the student receives feedback in the small white box below the GeoGebra area. At any time, another affordance is available in form of an *instant aid* video clip. The clip displays a similar situation and demonstrates the successful completion of the task. In this assignment, we imported photography into the GeoGebra software what could possibly tackle the child's mathematics anxiety, if any; according to Joseph M. Furner, *"When math has a purpose, then students are willing to spend time in exploring and understanding new concepts. Real-life photographs that are inserted into GeoGebra will provide the basis to observe relationships with* *different and similar shapes"* (Furner, 2019, abstract).

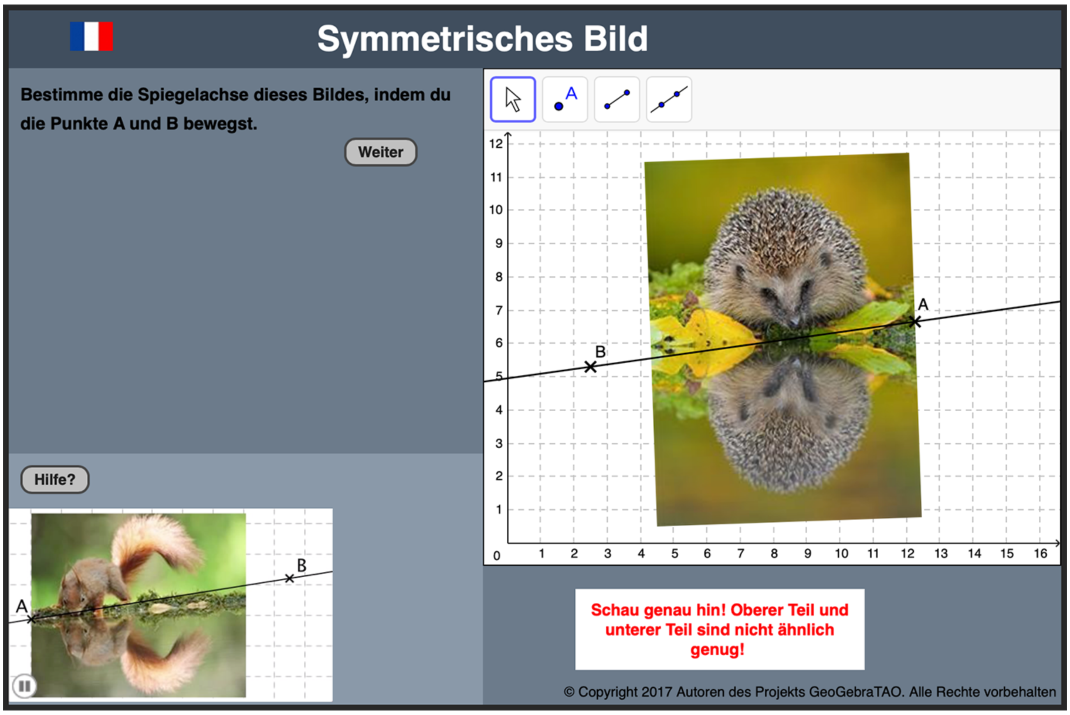


Figure 2.2: Example of an assignment containing a task to be fulfilled in the GeoGebra area itself (properly moving a line such that it transforms into the symmetry axis of a real-life, familiar image)

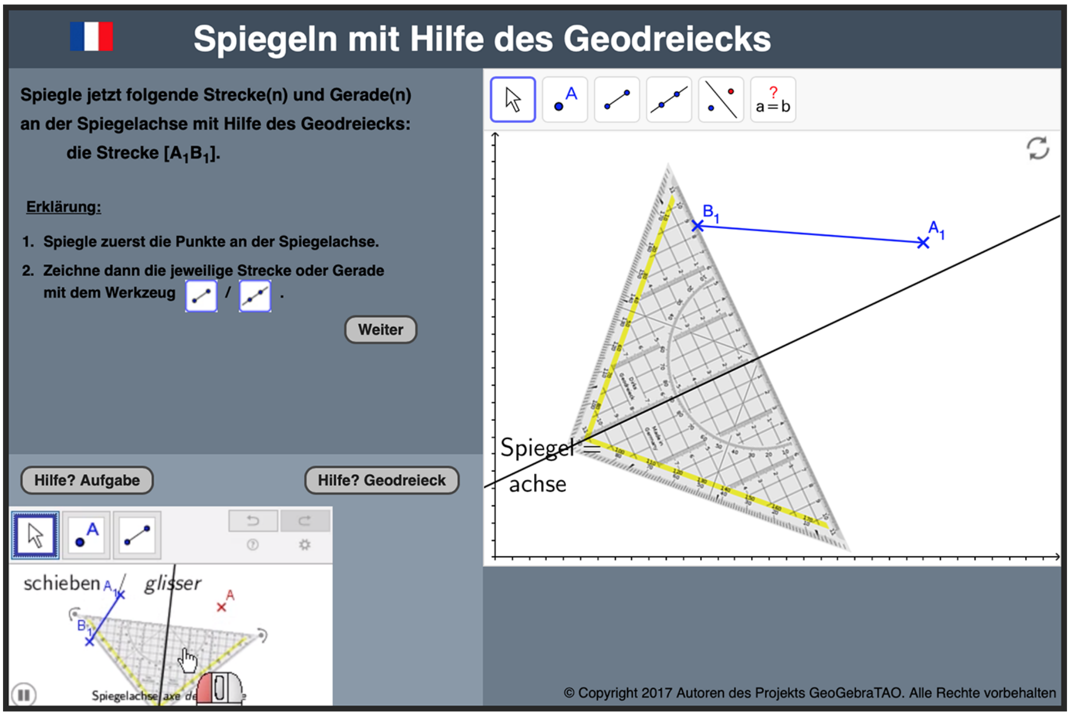


Figure 2.3: Example of a strict geometry assignment containing a task to be fulfilled in the GeoGebra area itself (mirroring a segment with reference to a given line using a dynamic set square)

Figure 2.3 illustrates a strict geometry assignment which asks the child to mirror a segment in the GeoGebra area with reference to a given line. Affordances the child might draw upon are the use of a dynamic set square (created by us) and the appropriate GeoGebra drawing tools. *Instant aid* is provided by a *short* video clip showing how to proceed.

2.2.3.2 Accomplishment through choices on Likert scales (device B)

The GEOGEBRATAO tool frequently invites children to answer Likert scale questions or analyse Likert scale statements (located in the upper-left part of the screen) by using the GeoGebra area for experimentation. Therefore, the tool uses a $(two$ x $two)$-point Likert-Type scale (device B) to measure certitudinal (attitudinal (Likert, 1932; Boone and Boone, 2012)) scales. To avoid central tendency and neutral responses, the original Likert scale (a $five$-point scale) has been reduced to a $(two$ x $two)$-point scale ($four$ phrasings instead of $five$).

A certain number of learning assignments use this device in presenting the learner with $three$ questions or statements with $four$ response choices each. The questions and statements are formulated with identical phrasing across all assignments. More precisely, it is a combination of two two-point Likert-Type scales; the first based on a 'yes - no' response choice crossed with the second based on a 'sure (!) - not sure (?)' response choice. The $four$ phrasings are:

Table 2.2: Likert scale matrix and phrasings

	!	?
Yes	**Ja! / Yes!** *means:* I am completely sure the answer is *yes*.	**Ja? / Yes?** *means:* I think the answer is *yes*, but I am <u>not</u> sure.
No	**Nein! / No!** *means:* I am completely sure the answer is *no*.	**Nein? / No?** *means:* I think the answer is *no*, but I am <u>not</u> sure.

Figure 2.4 illustrates a geometry assignment coupled to a Likert scale rating task. The children are invited to imagine how the symmetry point of a given original point will behave in case the original point starts moving (both points are initially depicted as static in the GeoGebra frame). $Three$ answers are stated and coupled to Likert scales. The child is invited to select the option that seems correct to them. However, if the child is uncertain about his decision (as in the present case), the GEOGEBRATAO tool offers a scaffold, which allows the child to verify their answers autonomously by moving the original point in the frame (by activating the points). This assignment offers no *instant aid short* video clips.

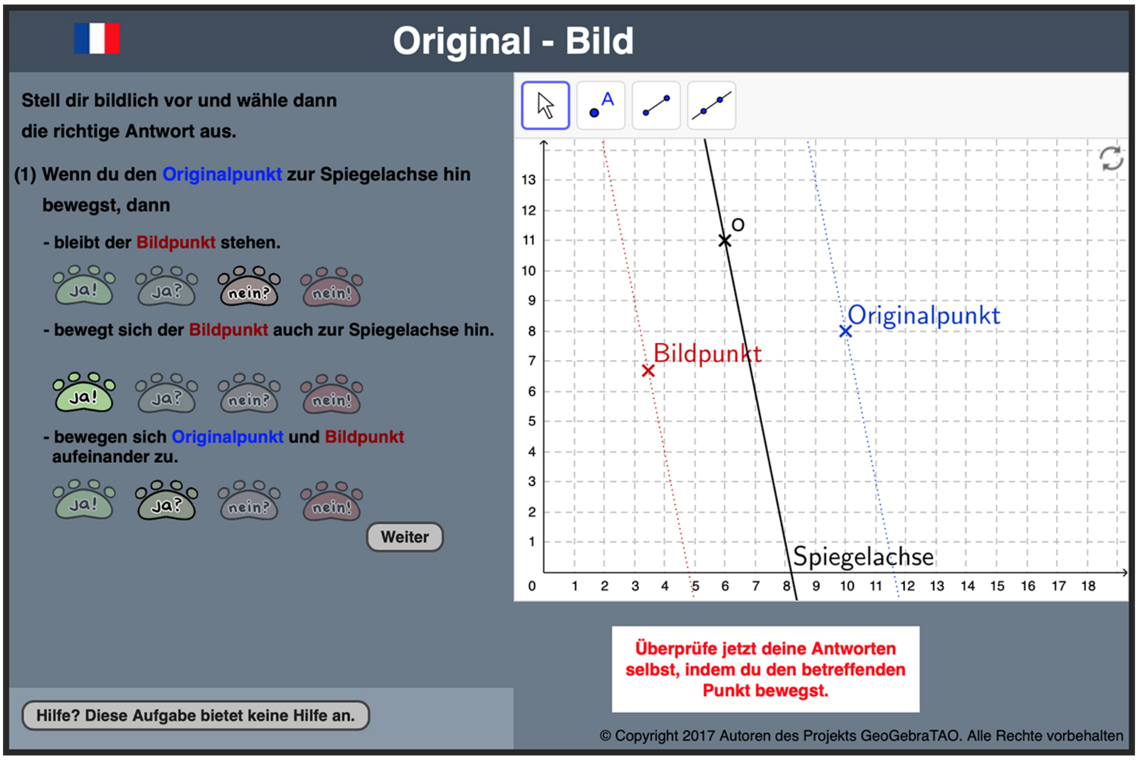


Figure 2.4: Example of an geometry assignment (about symmetry) containing Likert scale statements

2.2.3.3 Accomplishment through choices in drop-down menus (device$C$)

In this case, the children complete the task by selecting the right answer from a drop-down menu, situated in the upper-left part of the computer screen, after having explored the topic (e.g. properties) within the GeoGebra frame.

Figure 2.5 illustrates an exploratory learning assignment involving a drop-down menu option. The figure shows a task that compels the child to investigate a symmetry feature related to a bird's flight over a lake and its mirror image (on the water surface). The child is asked to predict the direction of the mirror image when the bird flies to the right. In this figure, the child has made the wrong selection, so that they receive help in form of a scaffold (similar to the one of Figure 2.4).

2.2.4 Integrated 5-in-a-row principle

Like the Stanford-Binet intelligence test, the GEOGEBRATAO tool emphasizes a basal and a ceiling level defined as a fixed number of either successfully completed learning tasks or consecutive errors for a given task: *"A basal is a performance on a sequence of items that is sufficiently good to justify an assumption that all easier items would have been passed if they had been administered. $\cdots$ A ceiling is the converse: a series of item failures that allows one to predict failure on all more-difficult items." (Embretson and Hershberger, 1999, pg. 39)*

To fix the number of consecutive errors or successful actions, we took the *Test of Early Mathematical Ability*-3 (TEMA-3) as a basis. TEMA-3 measures many aspects of

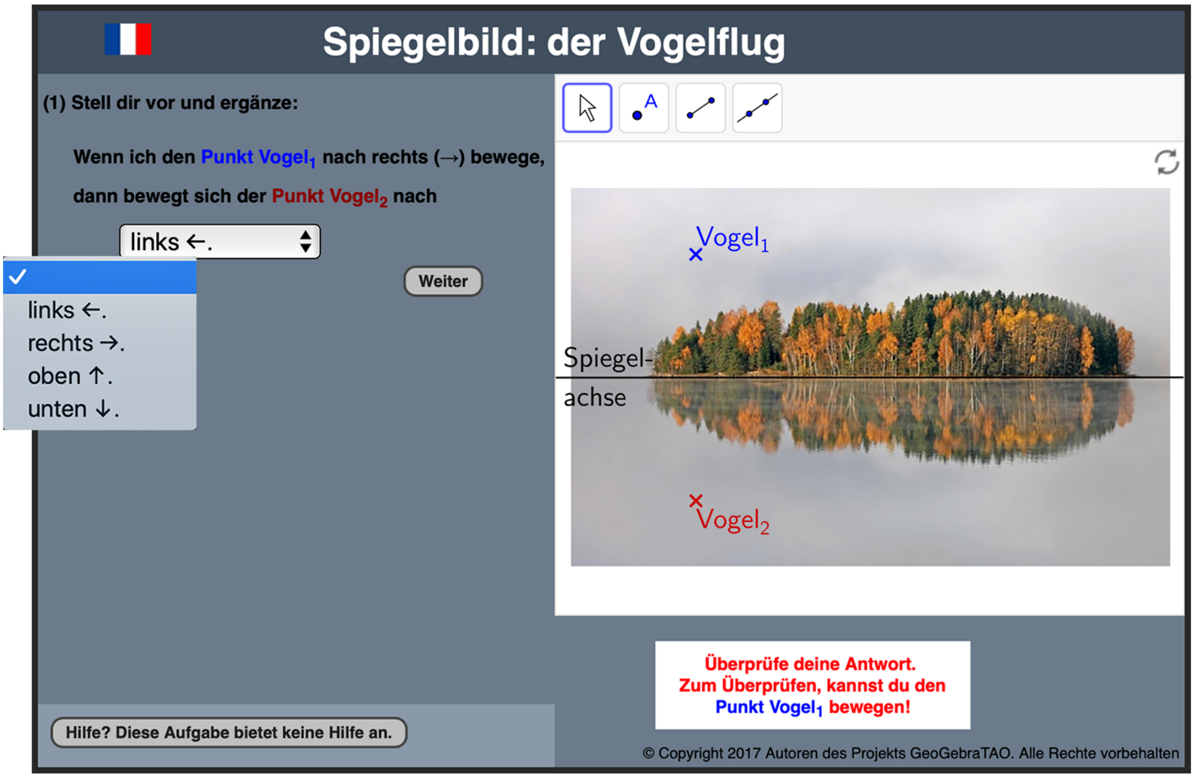


Figure 2.5: Example of a real-life symmetry situation containing a drop-down menu

mathematical performance and ability in childhood, e.g. numeracy skills, calculation skills, number concepts and number-comparison skills. It continues testing until the child has passed $five$ consecutive items (tasks) or missed $five$ consecutive items. Children's performance on TEMA-3 is highly correlated with their performance on other math achievement tests (Li et al., 2018; Chu et al., 2013).

Thus, the GEOGEBRATAO tool applies the 5-*in-a-row* principle in assignments using device A and in some assignments using device C. The principle is enacted as follows for the different assignment types:

- For tasks related to device A and some to device C,
  - as soon as a child successfully completes five consecutive tasks of an exploratory learning assignment, they progress to the next assignment.
  - as soon as a child fails to complete the same task $five$ times in a row, a specific help feature is inserted into the child's activity sequence. This action aims to overcome the child's particular difficulty in comprehending the topic. This aid is provided in form of an *overlay* video clip display (cf. Figure 2.6) or a series of recurrent tasks. However, the support system begins even before the $fifth$ unsuccessful attempt occurs. From the $third$ consecutive failure onward, the child receives some minor help in form of enhanced scaffolds, e.g. short sentences in the white box or visual scaffolds in the GeoGebra area. Moreover, the child's activity sequence is complemented by an additional assignment called '*external prim*' [e.g. AKTIVITÉIT $1'$], which is assigned only after the child manages to complete the regular assignment with which they experienced difficulty.

- For tasks related to device B and the remaining tasks related to device C,

  we can't use the 5-*in-a-row* failure principle due to the risk of the activity becoming a guessing game. So, for these tasks, we put a ceiling level of two in place. After the first failure or uncertainty (for tasks related to device B), the child gets some help in the form of scaffolds (see Figures 2.4 or 2.5). After the second failure, more help is proposed in form of an *overlay* video clip display (cf. Figure 2.6) or a series of recurrent tasks. Additionally, supplementary items are introduced into the child's learning path (on an *enforced* basis). The supplementary items, called *'internal prim'* items [e.g. AKTIVITÉIT 6INT$'$], are inserted into the exploratory learning assignment the child had difficulties with.

- For all tasks,

  at the child's $third$ passage, which means after the following sequence: '1) errors - overlay video clip display, 2) errors - overlay video clip display, 3) errors -', the software locks automatically and displays a message asking the student to call the teacher. This same procedure is activated for returns to any previous assignments from an assignment X onwared, i.e. '1) errors - return to previous assignment(s) - assignment X, 2) errors - return to previous assignment(s) - assignment X, 3) errors -'. By entering a password only the teacher knows, they are able to unlock the tool. However, the teacher should act according to the pre-defined mediational intervention procedures taught during the GEOGEBRATAO-launch training session.

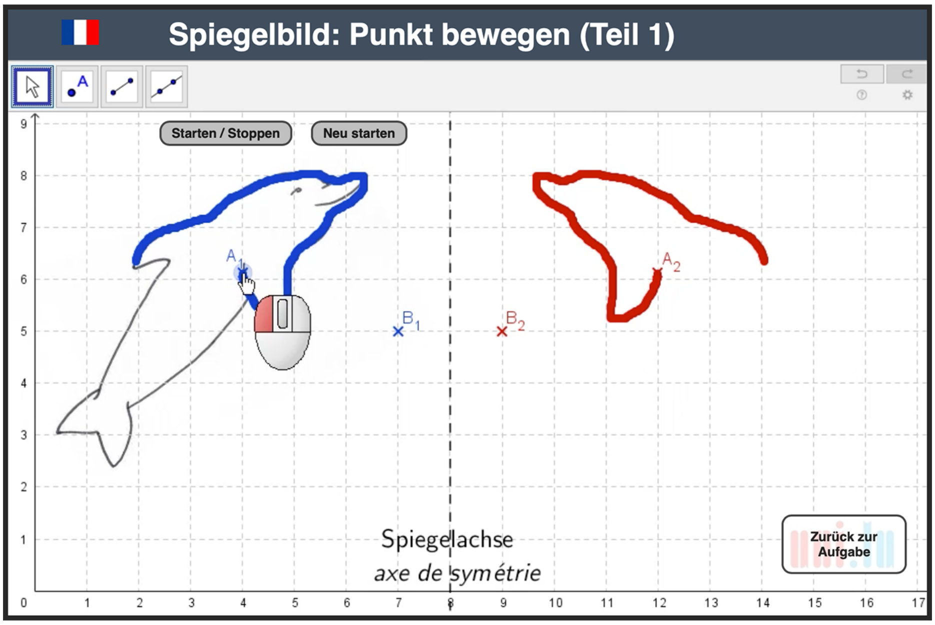


Figure 2.6: An *overlay* video clip display, studying the movement of the mirror image of a jumping dolphin

It should be noted that the 5-*in-a-row* principle is also used in a literature-related curriculum project, called 5-*in-a-row*, in which parents or teachers spend $five$ consecutive days (a full week) reading the same story to their child: *"$\cdots$ children enjoy repeating books because the experience imbues them with feelings of competence and mastery; because, with each reading, they understand a bit more of what they're seeing and hearing." (Gurdon, 2019), as cited in (Five in a Row - My FIAR, n.d.)*

2.2.5 Personalizing the software

To personalize the software, a mascot has been created by the cartoonist SARAH SOBOLE from Cleveland, Ohio: a cat able to communicate different facial expressions and body positions. Being alternately static and animated, and located nearby the feedback box (see Figure 2.7), this mascot is meant to capture the child's attention and create a pleasurable experience in keeping with Kapor's feature of delight. For instance:

- a sad static cat indicates that an error number (errors in a row) of $three$ has been reached for achievements related to device A and for some achievements related to device C, and thus represents an alert symbol in addition to the given feedback,
- a sad animated cat indicates that the ceiling level of consecutive errors has been reached for a certain activity task,
- a happy animated cat provides positive visual feedback and joyfully indicates the successful achievement of an exploratory learning assignment.

Figure 2.7 is a snapshot of a GEOGEBRATAO Type A achievement, which targets the differentiation of lines and segments. We see in the screenshot the precise moment when the user failed to correctly answer the same item $three$ times in a row. The sad static cat helps the child understand their progress (idea about their error number) on the current assignment item while targeted scaffolds are displayed on screen. Scaffolds are provided in the form of both a short sentence in the white box and visual cues in the GeoGebra area; the visual cues consist in the use of identical colors, which draw the child's attention to the geometric objects concerned. Furthermore, the child can also take advantage of a *short* video clip (an affordance provided on a *'volunteer'* basis).

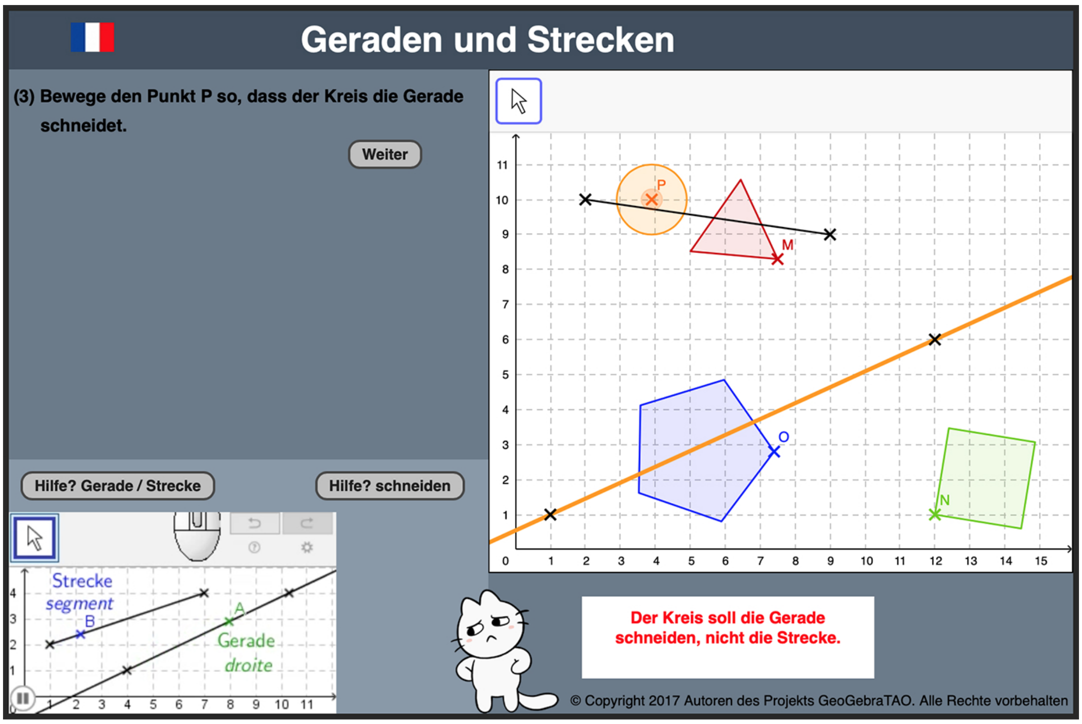


Figure 2.7: The GEOGEBRATAO tool screenshot with a sad software mascot (cat) for $three$ consecutive failures

2.2.6 Scaffolding and feedback practices

GEOGEBRATAO provides individual support to each child within a diverse classroom community through practices of scaffolding and feedback. This approach has been conceived to facilitate the exploration and understanding of new geometric concepts, and consequently to enable child-specific differentiation or differentiated learning (Belland, 2017). The following definition, used by Brian R. Belland from Utah State University (USA), has been adopted:

"Instructional scaffolding can be defined as support provided by a teacher / parent, peer, or a computer- or a paper-based tool that allows students to meaningfully participate in and gain skill at a task that they would be unable to complete unaided." (Belland, 2014, pg. 505)

As we have already discussed, some degree of freedom in individual children's timing and pace is necessary to address the broad range of ability levels present in a classroom community. The GEOGEBRATAO tool allows classrooms to tackle this challenge through automated feedback - more precisely through computer-based scaffolds:

"Computer-based scaffolds emerged as a solution to the dilemma that teachers in $\cdots$ classrooms cannot be expected to provide adequate one-to-one scaffolding to all students in a classroom. $\cdots$ can be used to supplement one-to-one scaffolding (Saye and Brush, 2002)", as referred in (Belland, 2014, pg. 510) and, in this case, can act as a *booster*.

Scaffolds in text form, given in the white box at the bottom-right of the screen, are brief and simple so as not to burden the children with too much challenging reading (Viau, 2009). Simultaneously, in some learning assignments, visual scaffolds are displayed in the GeoGebra area to complement the text-based support cues, pointing to the relevant graphic elements to enhance the text-visual connection: *"$\cdots$ visual scaffolding may be particularly important in instructional settings, in which students' comprehension is often challenged by new concepts and unfamiliar terms" (Alibali, 2006).* These scaffolds are also useful in the case of language or reading comprehension difficulties. For example, the scaffold provided in the GeoGebra area (see Figure 2.7) quickly pinpoints the relevant geometric objects emphasized by identical color: *"(Reiser, 2004) for principles on how to balance simplifying tasks and drawing attention to particularly important content", as referred in (Belland, 2014, pg. 513).*

Through scaffolds, children should be able to perform the necessary steps to succeed in any activity on their own. According to Brian R. Belland, *"$\cdots$ scaffolding serves to both simplify processes and highlight their complexity (Reiser, 2004) $\cdots$ scaffolding can address more complex processes and knowledge (Pea, 2004; Puntambekar and Hubscher, 2005) $\cdots$ scaffolding is designed to be used temporarily while students gain skill at the scaffolded task (Wood et al., 1976)", as referred in (Belland, 2014, pg. 506).*

Simple elementary feedback based on the child's answer is provided after each error. Once the number of consecutive errors reaches $three$ for achievements related to device A and for some achievements related to device C, the GEOGEBRATAO tool mascot turns static and sad. The occurring scaffolds are enhanced by visual effects or by supplementary instruction, e.g. emphasizing objects which should be given particular attention. This scenario takes place when the child gets stuck on a task or fails to grasp the geometry concept taught by the exploratory learning assignment.

For Type B and for the remaining Type C achievements, scaffolding support is only given once (cf. subsection 2.2.4), including visual scaffolds whenever possible.

As it is *"crucial to develop scaffolds"* (i.e. computer-based scaffolds) *"that better enlist student interest, maintain student direction, and control frustration" (Belland et al., 2013)*, feedback is best phrased positively and the software mascot creates a lighthearted environment in the GEOGEBRATAO tool.

When the pupil successfully progresses in a learning assignment, they receive feedback announcing the content that has been learned during the assignment accomplishment.

2.2.7 Video clips

As previously described, the GEOGEBRATAO tool includes

1. *short* video clips (cf. Figures 2.1, 2.2, 2.3, 2.7) aimed at helping the child better understand any task as needed. This is an affordance available on a **voluntary basis**,
2. *overlay* video clips (cf. Figure 2.6) aimed at drawing the child's full attention to the concept of the ongoing learning assignment. They are imposed in some specific error scenarios on an **enforced basis**.

Short video clips show the main ideas of an assignment item in a dynamic manner; the child is stimulated to imitate the respective movements in the GeoGebra area while watching the video clip. This option does obviously not exist for *overlay* video clips. During their transmission the child is supposed to stay focused on the content of the clip as much as possible without getting distracted by other tasks to perform in parallel (e.g. by clicking on objects in other areas). In this case, the GEOGEBRATAO tool effectively takes over the regulation of the activity and the child becomes a silent watcher.

As most of the participating schools didn't have enough headphones, only silent video clips are used to present (visual) explanation, i.e. without sound or spoken language (Kristinsdóttir et al., 2018, 2019). Further characteristics are:

- They are very short, typically less than 3 minutes. However, when parallel working in the GeoGebra area is intended, they may be slightly longer.
- They present mathematics dynamically based on recordings of animated mathematical objects or concepts in the GeoGebra area.
- *Short* video clips have a tighter focus compared to *overlay* video clips due to the small area they occupy within the entire GEOGEBRATAO tool screen.
- They draw attention to a precise theme or topic. Some deal with a mathematical concept (mathematical help), while others offer technical help for a specific feature of the tool.
- *Short* video clips can be replayed as often as necessary. *Overlay* video clips may be paused and then restarted.
- Regarding written text used in the clips, only occasional short sentences or single words (mathematical terms) are displayed.

We used BANDICAM (Bandicam Screen Recorder, n.d.), a screen recorder software, to record a specific area on the PC screen, typically the GeoGebra area. Any mouse click effect was added to the recording through the SPOT ON THE MOUSE SOFTWARE (Welz, n.d.), a mouse pointer and keyboard action visualization software created by Markus Welz which only runs on Windows: *"Mouse graphic: A picture of a mouse moves along next to the pointer. It indicates the button clicks and wheel rotations."* Thus, the children are able to see what they should do with the mouse.

2.2.8 Repeating activities

"Brain research indicates that repetition is of vital importance in the learning process." (Saville, 2011, pg. 69)

The spaced repetition feature (Desai and Gaglani, 2018) is a highly useful component of the GEOGEBRATAO tool. It addresses *two* learning-related problems in our study. First, the lasting retention of mathematical information and skills in long-term memory varies among learners, according to Atkinson-Shiffrin's multi-store memory model (Atkinson and Shiffrin, 1968). After a short break such as a weekend or a longer break such as a school holiday, students' memorized content typically declines exponentially to varying degrees, so concepts and their associations need to be strengthened again through repetition of prior learning assignments.

Second, the prior knowledge of students varies regarding which concepts they can recall from memory and the gains they had made in developing sustainable understanding in prior learning assignments. As soon as children do not show mastery of the relevant subject matter to complete a specific geometry activity, they must be redirected to more basic learning assignments that enhance their understanding and they need additional activities to reinforce their knowledge and familiarity with the concepts under scrutiny.

According to *How to Study in Medical School*, *"Spaced repetition is not only about revisiting the material more often, but also about reviewing it at the right time" (Desai and Gaglani, 2018)*. In the GEOGEBRATAO tool, the time dedicated to re-strengthening understanding of a geometry concept in students' long-term memory is governed by the integrated 5-*in-a-row* principle, described above in subsection 2.2.4. This principle is a computer algorithm that operates as follows:

1. If a child fails to pass the same task a fixed number of times in a row (critical moment), the algorithm automatically adapts the information displayed on screen by prioritizing an *overlay* video clip or a series of preceding learning assignments for repetition.

2. If a child successfully passes the essential items of an assignment, the algorithm automatically refreshes the screen by directing the child to the next exploratory learning assignment or possibly to the corresponding *'external prim'* activity (for Type A and for some Type C achievements). Thus, this algorithm reduces the child's overall study time by avoiding any waste of time in unnecessary review of concepts already mastered by the child.

In the GEOGEBRATAO tool, the study of new concepts is spaced out in (short) intervals with repetitions foreseen within each of these intervals. Figure D.2 (see appendix D) is an extract of the workflow of the GEOGEBRATAO tool and visualizes two of those intervals.

- The first contains $four$ main assignments ($Akt13$, $Akt15a$, $Akt15b$, $Akt13bis$) [Akt abbreviation for AKTIVITÉIT], two *'external prim'* activities ($Akt13prim$, $Akt15bprim$) and two possible loops to repeat preceding assignments.
- The second interval is shorter, and only contains two main assignments, one *'external prim'* activity and one possible loop.

One benefit of these procedures is that they avoid boredom by eliminating unnecessary repetition; conversely, they prevent the children feeling overwhelmed by information during comprehension difficulties. However, these procedures might not completely prevent repetitions from being perceived as frustrating. To minimalize such negative experiences, we reiterate that the software is locked at the child's third passage, i.e. after the following behavioral sequence: *'1) errors - return to previous assignment(s) - assignment* X*, 2) errors - return to previous assignment(s) - assignment* X*, 3) errors -'*. A message asks the student to call the teacher, who can use a password to unlock the software and engage in clarifying dialogue with the child (in a similar way for the display of *overlay* video clips).

As repetitions are of vital importance in any learning process (Saville, 2011), even in cases where no prior errors and failures make a loop mandatory, easy tasks are regularly integrated into the activity sequence of any child. These tasks are conceived similarly to the ones the child has already passed and are used to introduce a new concept/goal.

2.2.9 Short technical description GeoGebra - TAO

By default the Question & Test Interoperability ®(QTI) item specification on which TAO (version 3.1.0-RC3) (User Guide TAOtesting, 2016) relies allows a finite list of interactions, e.g. Multiple Choice Questions, matches and ordering: *"The QTI standard enables the exchange of item, test, and results data between authoring tools, item banks, test constructional tools, learning systems, and assessment delivery systems" (TAOtesting, 2018)*. However, a custom interaction, such as one that embeds a GeoGebra file, can be created through Portable Custom Interaction (PCI): *"Portable Custom Interactions (PCIs) are interactions which are developed for a specific scenario, generally to fulfill a particular need of a customer, hence are not classical QTI interactions. PCIs represent a best practice for defining and packaging custom interactions" (User Guide TAOtesting, 2017)*. Thus, thanks to a dedicated PCI and a modification of TAO to allow the GeoGebra file type, it is possible to combine the use of GeoGebra and the mechanics of QTI tests.

The QTI test describes the branching rules that create links between items (AKTIVITÉITEN). These rules are processed each time an item submits its result(s). This makes it possible to implement the workflow defined by the GEOGEBRATAO tool, i.e. to provide users (children) the appropriate next item (AKTIVITÉIT) based on their previous results (see extract of workflows, appendix D).

The data collected within each exploratory learning assignment is a combination of user inputs, result(s) and some metrics related to the assessment. It is sent to the server and stored on it as JSON data.

TAO allows the delivery of the test to the users (children), and provides user management features (creating, adding and deleting a test-taker (child) into/from the system, and its adding to a group) (User Guide TAOtesting, 2015) and test delivery management features: *"Assembled Deliveries provide the means of publishing and administering Tests. These govern when a test is taken by selected individual or Groups of Test-takers and how long tests will be. A delivery can only be assembled after the creation of Interactions, the assembly of the test, the creation of test-taker profiles, and the gathering of test-takers into formal groups." (User Guide TAOtesting, 2019)*

2.2.10 Recommendations to teachers for using the GEOGEBRATAO tool in class

The core recommendations concern the language barriers in classroom activities. When teaching a highly diverse student population, including many children with an immigration background, language issues are a day-to-day challenge. Understanding the learning task is the sine qua non of success in its achievement. This is particularly important as we know that children struggle more with mathematics (Vukovic et al., 2010) when they also experience reading difficulties, not only regarding foreign languages but also in the case of dyslexia or any other specific reading impairment.

To make sure that every child is able to draw upon their full potential while working with the GEOGEBRATAO tool and experiences equal opportunity for success, teachers are allowed to explain, simplify and translate some expressions on the screen when they perceive this action as absolutely necessary. The software itself is in German and offers access to an inbuilt French version through a simple mouseover move.

When a child faces language difficulties, the teacher is asked to:

- read the sentence(s) aloud with or to the child in the child's most fluent language (either German or French),
- inform the child that they should immediately stop reading aloud when confused,
- let the child explain their understanding (meaning check),
- help the child first understand what they read and start translating only after this first step (if still necessary).

The same procedure is used when children are stuck on an activity, i.e. when the message to call the teacher is displayed or when the teacher simply notices that the tool is blocked or the child is confused. Here, the teacher should prompt the child to

- explain what they have done or tried so far through 'preliminary questions',
- describe their reasoning and mental processes, e.g. the current state of the solving process, knowledge check, $\cdots$

For the child, being asked to explain a process, communicate steps or indicate understanding is a sort of supplementary but meaningful exercise. Children are often so eager to get the answer that they rush, and hence read maths texts quickly or even just skim them. By doing so, they miss significant details, connections and underlying logic (Barton and Heidema, 2002). This request for 'supplementary explanation' should both avoid the 'rush to save time' attitude of the children and help them gain competence in reading a maths text.

The explanations and descriptions provided by the child allow the teacher:

- to fill gaps in the child's understanding;
- to invite the child to read the corresponding text again;
- to sharpen the child's skills to watch processes displayed on screen;
- to show the child how to activate prior knowledge of the topic;
- to explain the reasoning behind the solution.

In a couple of activities, the teacher might practice a kind of visualisation or prediction (Taylor, 2012) to make a child aware how to improve in an activity despite difficulty reading the corresponding text. In that case, the teacher can guide the child to use the GeoGebra file clues, for example by asking the child to anticipate what might happen when they move the point(s) of the given GeoGebra file and to test the move right after (an advantage of the drag mode in a Dynamic Geometry System). Further interventions to enhance the child's comprehension might focus on making connections between the GeoGebra file and the text of the respective assignment more comprehensive.

2.3 Instructions for the control classes

The control-class teachers had greater freedom to run their familiar classroom activities than the test-class teachers. In fact, the only imposed requirements were:

1. the pre-test (on the $8th$ of January 2019),
2. the post-test (by the end of March 2019),
3. 500 minutes of geometry training (comparable duration to the computer-enhanced learning activity in the test classes) between the pre- and post-tests.

The control-class teachers were given the following list of the main geometry topics that were, if possible, to be taught in this period:

1. coordinates,
2. segments and lines, including parallel and perpendicular lines,
3. recognizing symmetry in pictures / symmetrical pictures,
4. basic knowledge (mathematical properties) of symmetry,
5. drawing symmetrical pictures (symmetry of points, segments, lines, circles, shapes, parallel and perpendicular lines),

6. drawing a symmetical axis / symmetrical axes of shapes,

7. knowing the names and mathematical properties of the main shapes (e.g. square, rectangle, rhombus, parallelogram, triangle, $\cdots$).

In addition to this list, the teachers (of both the control and test classes) were allowed to see the pre-test. Moreover, they were told that the pre-test and the post-test were identical. However, they were neither allowed to talk with their pupils about any pre-test exercises nor use them during preparatory work in their classes. We assumed that this would not be a problem for them, since this is a procedure common to the national 'standardized tests' or 'national common tests' conducted recurrently throughout the country.

The teachers of the control classes were expected to structure their lessons as usual, prioritizing more traditional paper-and-pencil manner (standard method), to cover main topics and enable children to solve exercises similar to those encountered in the pre-test. If the teacher couldn't complete all topics at the end of the 500 minute period, this would not present a problem; the research would take these circumstances into account.

After the completion of the research, our 5 control-class teachers all reported that they completed instruction on all required topics. Similarly, the test classes were not required to make children complete all GeoGebraTAO tool activities by the end of the 500 minute period. To create similar classroom conditions in the control and test groups, it was important for us to account for the possibility that all topics might not be covered or completed.

The teachers of the control classes filled in a sort of lesson plan to document the topic, date, lesson time and instructional mode of each geometry lesson. The total duration obviously could not exceed the 500 minute limit. Table 2.3 is a part of such a lesson plan.

Table 2.3: Part of a lesson plan documenting geometry lessons in the control classes

date	**lesson time**	**topic plus instructional mode**
$30th$ of January 2019	15 minutes	finding the x-coordinate of a point *(explanations on the blackboard)*
$30th$ of January 2019	40 minutes	finding the x-coordinate of a point *(writing exercises down in the children's notebook)*
$30th$ of January 2019	10 minutes	finding the x-coordinate of a point *(children's homework)*

We limited any detail in the control-classes' instructions to avoid the possibility of us influencing their pedagogical approach. The teachers of the control classes were supposed to exercise autonomy and work as usual, with their normal school manuals and personal worksheets. For example, any lesson plan or worksheet(s) designed by us and imposed on the control classes could have made some teachers feel less comfortable than working with their conventional and familiar classroom tools, in which case our results could have been influenced by our own actions.

For the study, we finally chose the mixed design that comprises the following two independent variables (Gliner et al., 2003):

1. The between-groups independent variable (intervention) is exclusively the exploration of elementary geometric concepts in a sequence of learning activities, created with the dynamic mathematics system GeoGebra integrated into the computer-assisted testing framework TAO (Kreis et al., 2018, 2010b) (by the test group). The control group received no intervention.

2. The within-subjects independent variable is the *'change over time from pre-test to post-test'* (in both the test group and control group).

In this chapter we gave an insight into the instruments we developed especially for this project with the highlight being the GEOGEBRATAO tool. A more detailed description of this exploratory learning tool was added (cf. appendixQ). Now we will explain the way we design our study to ensure valid and reliable results.

Chapter 3

Methodology

This chapter provides an overview of the research process, including the sample selection, the data collection and analysis, and how we establish the validity of our results and the reliability of our research, of both the GEOGEBRATAO instrument itself and its use in classrooms. Therefore, a quasi-experimental control-group design based on pre- and post-tests was applied in this study (Martella et al., 2013; Gliner et al., 2003; Bonate, 2000), which is one of the most common research designs used in education. In addition, data collected through the logs of the TAO platform are of paramount importance.

3.1 Sample selection

3.1.1 Test group and control group

164 children ages 10 to 13 acted as the **test group**. They came from $twelve$ different classes (labeled 01, 02, 03, $\cdots$, 12) within six different schools located in six different villages (labeled A, B, C, $\cdots$, F) in the southwest of the Grand-Duchy of Luxembourg. Hence, the test classes are labeled $XA01$, $XA02$, $XB03$, $XB04$, $XC05$, $\cdots$, $XF12$. These children worked exclusively for 500 minutes with the GEOGEBRATAO tool in an autonomous way and at their own pace, with as little help as possible from the teacher.

For our controlled experiment, a **control group** of 72 children of similar ages was selected for the purpose of comparison. They were affiliated to $five$ different classes (labeled 21, 22, $\cdots$, 25) within $three$ different schools located in $three$ different villages (labeled G, H, I) in the same region of Luxembourg. The control classes are labeled $XG21$, $XG22$, $XH23$, $XI24$ and $XI25$. These children received standard instruction (no experimental stimulus). They exclusively followed a paper-and-pencil geometry course for 500 minutes based on traditional didactic teaching methods. The course was delivered by each control-class's teacher and dealt with the same material as the exploratory learning assignments of the test group.

3.1.2 Identification of eligible classes

In the Grand-Duchy of Luxembourg, the inspectorate of fundamental education is divided into 15 regional directorates. Each regional directorate is managed by teams composed, among others, of one director and two to $four$ assistant directors.

For this project, we chose $four$ regional directorates located in the southwest area of the country. In this region, some communes or municipalities have a socioeconomic index close to 1, which points towards *greater socioeconomic disadvantage* than other communes at national level. In addition, these same communes also display a strong immigration background, evidencing a greater diversity in their local population than in other regions of Luxembourg. This diversity is mainly fueled by the large Portuguese community living in the southwest of Luxembourg. The high percentage of non-national residents in this area is further boosted by significant groups of Italians, Africans and non-EU-28 Europeans (STATEC, 2019).

As previously described, teachers working in this socioeconomically disadvantaged area are often overstrained. Their class populations generally have poorer educational outcomes than those with populations from a higher average socioeconomic level (Thomson, 2018). According to OECD, *"Immigrant students* (children) *often face the double disadvantage of coming from immigrant and disadvantaged backgrounds" (OECD, 2016, pg. 244)*, known as lower socioeconomic backgrounds (cf. section 1.3).

Our aim was to investigate whether a tool such as GEOGEBRATAO would allow self-regulated learning in highly diverse classes and free some teaching time to be devoted to children in particular need.

The recruitment of project classes took place with the help and under the control of the local directors or assistant directors. We shared a brief description of the project by e-mail and asked for schools that suited our study objectives: classes with high student diversity, adequately equipped with ICT (to implement the GEOGEBRATAO tool) and not already overloaded with other projects. Furthermore, we preferred to work with two classes in each school. Interested teachers could either apply through their directorate or directly contact us.

The main targets of our study were ISCED *Level* 1 [1] fifth grade classes (noted ISCED 1.5). They had been taught fewer geometric topics at that time than ISCED *Level* 1 sixth grade classes (noted ISCED 1.6). ISCED 1.6 classes, however, were not formally excluded from our study. They could participate in the project when both teacher and pupils were motivated in using the computer as a medium for exploring geometric concepts.

Concerning the schools' ICT equipment, we established only one precondition for participating schools: each child needed to have access to a computer with an Internet connection from which it could connect to the GEOGEBRATAO tool with a personalized username ($xxyy$,

[1] International Standard Classification of Education (ISCED) *Level* 1 represents the education at the first level (primary or elementary education), which usually begins at age 5, 6 or 7 and continues for about 4 to 6 years. For the Grand-Duchy of Luxembourg, the first level starts at the age of 6. Two following school years are tied up to a two-year-spanning learning cycle, starting with cycle 2 ($6-7$ years old) and ending with cycle 4 ($11-12$ years old). We obviously refer to the regular age of the children.

xx = class number, yy = child number) and password for the whole duration of the project.

Once potential classes were identified, we first checked whether they met all requested criteria and then confirmed their eligibility.

3.1.3 Selection of participating classes

The basic assumption of a **control-group design** requires comparable classes in both the test and the control groups. To respect this precondition, both groups had to incorporate a comparable number of ISCED 1.5 and ISCED 1.6 classes, proportional to their respective number of classes. Any equivalence in knowledge between both groups could not be known before the pre-test was completed and therefore did not impact the selection of the classes.

We used a **quasi-experimental design** to select the participating classes and children with respect to the specific preconditions mentioned above. The students could not be randomly selected from the regional school population (they were attached to their class), nor could the participating classes be randomly assigned to either the test or control group (as they had to be equipped with the appropriate ICT).

Quasi-experimental research designs, in spite of some well identified possible biases, *"are extremely useful"*. They *"enable researchers to conduct representative research"* closer to *"real-world"* conditions (Martella et al., 2013, pg. 159-160) compared to the restrained school context alone.

3.1.4 Composition of the classes selected

$Twelve$ test classes and $five$ control classes fulfilled our conditions and were selected to participate in our project. The classes were paired (same group) within the separate selected schools, except for a single control class in the village school H. The test classes and the control classes came from different schools.

Table 3.1 lists the final composition of our test and control groups:

Table 3.1: Test- and control-group final composition

	test group	**control group**
number of classes (cl.)	12	5
number of children (n)	164	72
number of ISCED 1.5 classes	9	4
number of ISCED 1.5 children	125	54
name-code of the ISCED 1.5 *cl.*	$XA01$, $XA02$, $XB03$, $XC06$, $XD07$, $XD08$, $XE09$, $XE10$, $XF11$	$XG21$, $XG22$, $XH23$, $XI25$
number of ISCED 1.6 classes	3	1
number of ISCED 1.6 children	39	18
name-code of the ISCED 1.6 *cl.*	$XB04$, $XC05$, $XF12$	$XI24$

3.1.5 Ensuring the quality of the pre-test

In order to ensure the quality of our pre- and post-test instrument, we included two additional review procedures in our approach. It is essential that any study *"entails a critical examination of the understanding of each question and its meaning as understood by a respondent" (Kumar, 2010, pg. 158)*. So, we decided that

1. researchers should review the pre-test beforehand to identify potential difficulties before the test would be first given in a pilot class,

2. a pilot class with 12 children should first take the pre-test under field conditions as close as possible to the real pre-test situation later on.

We were present in the pilot class testing and observed the children, wrote down their questions, and discussed multiple features of the test items afterwards with the children and their teacher(s), including the ways the exercices were formulated, specific terminology, glitches in wording, ambiguities, difficulty levels, clarity of instructions, measured skills, and so on. This pilot phase of the pre-test enabled us to analyse, evaluate and correct it by using simple descriptive statistics. Thanks to this review and pilot phase, we were able to eliminate pre-test impediments.

3.2 Study process and data collection

The main data collection sources of our study are the TAO logs and the pre- and post-test. They are selected according to the study's core objective, i.e. investigating technology-enhanced learning with the GEOGEBRATAO tool; however, they also define the nature and quality of the data required to test our hypotheses (Kumar, 2010, 2018). Auxiliary sources are the teacher's questionnaire and some personal field notes.

3.2.1 Data collection through the logs of TAO

For this project, we built an interactive digital tool offering a sequence of dynamic learning assignments with successive levels of temporary support, called GEOGEBRATAO. The children were invited to explore elementary geometric concepts generated through the dynamic mathematics system GeoGebra, which is itself integrated into the computer-assisted testing framework TAO. TAO allowed us to connect assessment with learning, i.e. to collect rich and meaningful data about the children's learning activities that served as means to:

1. continuously assess each child's progress through the sequence of assignments (Kreis et al., 2018, 2010b),

2. partly mimic each child's approach and course of action in the GeoGebra area of these activities (more information in subsection 2.2.9).

3.2.2 Data collection through the pre- and post-test

As all the children had some prior knowledge about the geometry content from previous school years, we were able to use a **pre- and post-test** design in our quasi-experimental research. Between both tests, we imposed a time span of at least *two* months to minimize the 'memory effect' and 'practice effect'. Indeed, a substantial literature in experimental psychology asserts that short-term and intermediate memory decays exponentially with time (Wickelgren, 1970). The same test was given as a pre-test to both the test and control groups on the $8th$ of January 2019, and as a post-test to both groups by the end of March 2019.

As a general rule, the pre- and post-test sessions lasted for 90 minutes, during which all children worked individually and at their own pace. If a child did not finish in time, the teacher could allow more time by adding a comment on the child's test sheet next to the unfinished exercise(s). To take the test, the children needed a sharpened pencil, a set square and a compass and could choose to use a ruler.

In section 2.1, the precise structure of the pre-test / post-test was already presented in detail. As you could particularly see in Table 2.1, half of the exercises consisted of 8 items, and almost a quarter of the exercises consisted of 4 items, which allowed us to establish comparisons between certain exercises.

To eliminate ambiguous or misleading items in the pre- and post-test, an **item analysis** was performed as described in (Office of Educational Assessment, n.d.). The quality of the items was assessed by comparing the children's item responses to their total test score.

Obviously, for our study, the item difficulty is relevant in order to ascertain whether the children learned the geometry concept under scrutiny. Moreover, an important function of an item difficulty is its ability to discriminate between the children who know the subject matter being assessed and those who don't.

We listed the 121 pre- and post-test items according to their degree of difficulty (very low, low, medium, high, very high). This varying difficulty of the items (ranging from very easy ones to quite challenging ones), with items encoded to binary data, allows us to measure the reliability of the pre- and post-test with the **Kuder-Richardson Formula** 20 (Ebel, 1967):

$$r = \frac{k}{k-1} * \left(1 - \frac{\sum pq}{\sigma^2}\right) \tag{3.1}$$

where k is the number of items in the test; $\sum pq$, the sum of the item variances; and σ^2, the variance of the test scores. p is the proportion of the test takers who pass an item and q is the proportion of test takers who fail an item.

3.2.3 Auxiliary data: teacher's questionnaire

After the post-test, conducted at the end of March 2019, we sent the class-related outcomes of our study to the respective teachers. This information allowed them to give feedback to their children and, if they would like, to the parents. Attached to these outcomes, every teacher received a personal questionnaire about the use, viability and perceived effectiveness of the GEOGEBRATAO tool in their class. The explicit version of the teacher's questionnaire is appended to the manuscript (cf. appendix P).

3.2.4 Ethnographic observations in classrooms

To introduce the GEOGEBRATAO tool to the classes and to provide the best possible help in case of major problems, we decided to visit the participating classes. A first visit was scheduled to launch the GEOGEBRATAO tool in each test class (for the first GEOGEBRATAO lesson). This visit allowed us to witness first reactions from both the students and their teachers.

In the event of software-related or other technical problems (e.g. use of iPads, use of older computers, WiFi access, $\cdots$), we joined the classes again on request during the GEOGEBRATAO lesson(s) to reduce any bias that might have arisen from the use of the technology in our study. During our visits, we took care to interact as little as possible with the class population.

These school visits also allowed us to make some ethnographic observations concerning the use of our tool in class, the usability of the technical equipment, and the teachers' attitudes to this special learning approach. Ultimately, we considered these observations a fourth data source.

To sum up: The data for our study were collected from *four* sources: a) the logs from the GEOGEBRATAO tool, b) the pre- and post-test, c) the teacher's questionnaire and d) field notes from the classrooms.

In the following sections, we will first describe the methodology used to restrict post-test data as some test-group children were not able to complete the entire activity sequence, then, in a second section, we will define a set of hypotheses for our present research and specify the methodology used to test each of these hypotheses.

3.2.5 Restricting post-test data

Before developing the methodology that will be used for each individual testing of the elaborated hypotheses, we will describe a general procedure that recurs in the statistical analysis.

The pre- and post-test design accomplishes measurements before and after the intervention to assess whether the children's scores changed: *"Assuming that"* the test group and the control group *"are equivalent from the start of the investigation, any observed differences at the conclusion of the investigation may reasonably be attributable to the independent variable, along with measurement and sampling error." (Martella et al., 2013, pg. 136)*

However, we anticipated that some of the children would not complete the entire geometry sequence. As a consequence, zero-to-moderate progress would be evidenced in the pre- and post-test activities dealing with topics which these children were not able to explore within the GEOGEBRATAO phase. Nevertheless, they might still progress despite this lack of training by establishing links to other topics they encountered with the GEOGEBRATAO tool. Scientific literature shows that interactive software has the potential to provide a kind of mental scaffolding (Somekh, 1996) allowing participants to build new knowledge. This guidance leads children progressively towards a stronger understanding and, ultimately, to greater independence in the learning process.

That's why, we calculated *two* different post-test scores for each child:

- one score for **all** the items of the post-test (called post-test_all) and
- one score **only** for the items each individual child encountered when working with the GEOGEBRATAO tool (called post-test_restricted).

Accordingly, we were interested in whether there were any significant differences between these *two* scores.

To answer this question, we used **Generalized Mixed-Effects Models** throughout our statistical analysis. We contrasted models based on the restricted data with those based on the unrestricted data, and subsequently compared general trends. In this way, significant differences between post-test_restricted and post-test_all data could be identified.

In other words: We looked whether those children who did not finish the entire programmed geometry learning assignment sequence nevertheless built relationships between the subjects or activities that they did and did not explore within the GEOGEBRATAO phase.

So far, we explained the research strategy we used to collect data from the children and their teachers. Furthermore, we allowed the reader to get a good basic understanding of the study process and the different types of data we used, i.e. the primary data and the restricted data. Now we will illustrate what we *expect* to happen during our study.

3.3 Hypotheses and their validation

In the present project, we investigate the scientific validity of a very particular Adaptive Learning Tool in the field of dynamic geometry with a particular focus on the individual learning pathways of a highly diverse student population. In section 1.1 we already described our own observations and experiences from previous studies:

1. the MATES project in which we were able to demonstrate that students' school results improved substantially while using MATES compared to their achievements before working with this e-Learning tool;
2. the GEOGEBRAPRIM project in which children who explored geometric concepts actively on the computer (test group) had better learning results than children who followed a traditional paper-and-pencil course (control group).

On the basis of these promising results, generated thanks to a technology-enhanced autonomous and active learning approach, we created another software tool called GEOGEBRATAO. Concretely, we integrated the dynamic mathematics system GeoGebra into the computer-assisted testing framework TAO (Kreis et al., 2018, 2010b; TAOtesting, 2019).

Basically, the aforementioned outcomes also served as the rationale for defining a set of hypotheses for our present research that will be investigated with our methodological toolkit in order to bring clarity, specificity and reliability to our study (Kumar, 2010, 2018).

3.3.1 Hypothesis (HYP) *one*

HYP1: Children who follow a paper-and-pencil geometry course with traditional didactic teaching methods will not evidence statistically significant higher learning outcomes compared with children who learn the same geometric topics via the GEOGEBRATAO tool.

3.3.2 Hypothesis (HYP) *two*

HYP2: Student-centered software allows teachers to create a blended learning environment in which children learn geometric concepts autonomously, thereby reducing teacher interventions for the children in need of (individualized learning) support.

Thus, this project needs to test *two* hypotheses to determine the validation of the GEOGEBRATAO tool and of its implementation. The effects for children who were building new geometric concepts via the GEOGEBRATAO tool and those for children who were following a traditional paper-and-pencil geometry course are investigated based on their achievement.

3.3.3 Test of elaborated hypothesis *one*

3.3.3.1 Introduction

Hypothesis *one* predicts that there is no statistically significant difference regarding children's learning outcomes *after* an initial testing phase, in favour of the control group (following exclusively a paper-and-pencil geometry course taught with traditional didactic methods) compared to the test group (taught the same geometric concepts via the GEOGEBRATAO tool).

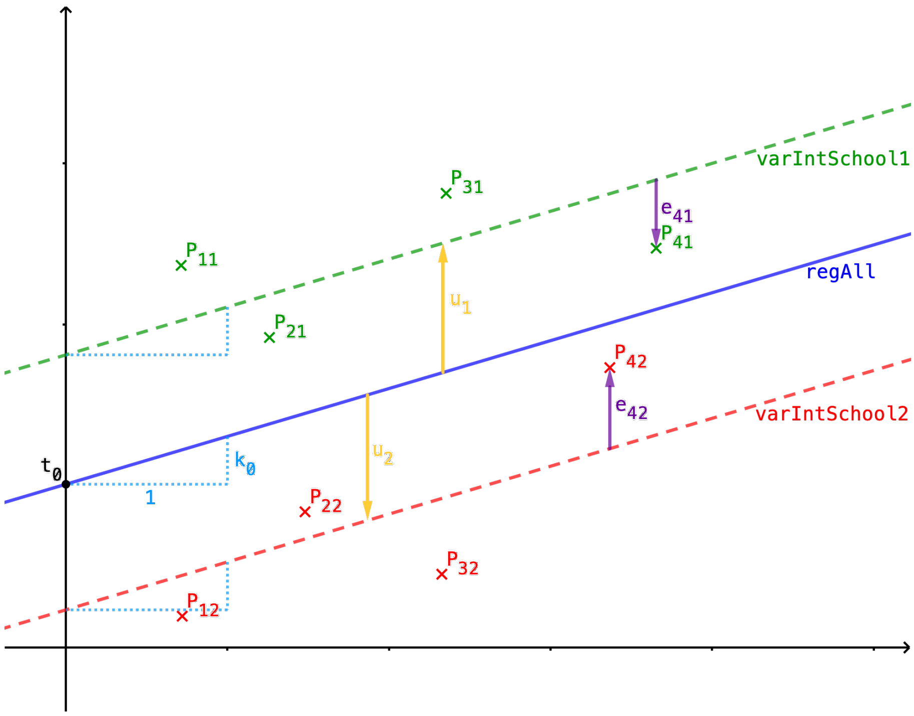
3.3.3.2 Methodology used

To explore and ultimately accept or reject this hypothesis, data from the pre- and post-test are analysed using **Generalized Mixed-Effects Models** (Stroup, 2012), whose *"··· most important advantage ··· is that they allow the researcher to simultaneously consider all factors that potentially contribute to the understanding of the structure of the data" (Baayen et al., 2008, pg. 409).*

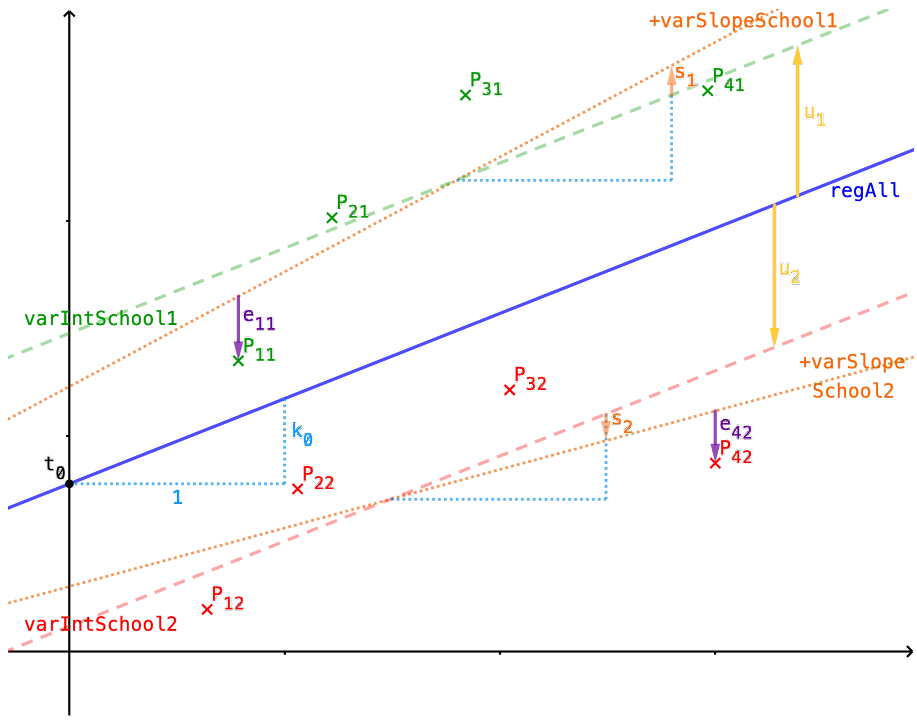
In this study, we are interested in children's learning of geometrical concepts through a sequence of computer-based activities versus a traditional paper-and-pencil geometry course (for children ages 10-13). The children came from 9 different schools, 17 different classes and 2 different grade levels (ISCED 1.5 and ISCED 1.6).

If a traditional linear model were to be fitted to the collected data, the assumption would implicitly be that learning would be the same regardless of the class or school the child attends. This is unlikely to be true; schools reflect the demographics of an area and the school ethos (cf. very diverse student population), and individual teachers might perform better or worse in teaching geometry to children. These realities imply that children within the same school are likely to be more similar to one another than to children from different schools. Moreover, children within the same class are likely to be more similar to one another than to children in the school's parallel class. In this way, data should be nested such that children are defined as part of a specific class in a specific school, and such that the different statistical models reflect this ordering. (Williams, 2018)

Generalized Mixed-Effects Models allow us to directly model this dependency in the collected data. Moreover, Generalized Mixed-Effects Models make fewer strict assumptions compared to other tests; in particular, they allow data distorted from normality, incomplete data, violation of the homogeneity of regression slopes assumption and violation of the independence assumption. Separate intercepts and slopes for each group (test group, resp. control group, school, class) in the study can be modeled as in the made-up examples shown in Figure 3.1.



(*a*) Random coefficient solution with varying (random) intercept; each school having its own line, parallel to the overall regression line



(*b*) Random coefficient solution with varying (random) intercept **and** varying (random) slope

Figure 3.1: Examples of *Mixed Model – Random coefficients* showing a relationship for an initial ranking on test outcomes across all children, from *two* different schools, participating in a project (Votta, 2017; Williams, 2018)

3.3.3.3 Random coefficients

Model (3.1b) of Figure 3.1 has the following structure:

$$y_{ij} = t_0 + (k_0 + s_j) * x_{ij} + u_j + e_{ij} \tag{3.2}$$

where

(x_{ij}, y_{ij})	: coordinates of child P_{ij} (i^{th} child of j^{th} school);
x_{ij}	: initial ranking of child P_{ij};
y_{ij}	: test outcomes of child P_{ij};
$t_0(fixed)$	: intercept of the overall regression line (all children involved);
$k_0(fixed)$	: slope of the overall regression line (all children involved);
u_j	: (level 2 residuals, random intercepts), difference between the intercept of the overall regression line and the intercept of the j^{th} school line (fixed slope);
e_{ij}	: (level 1 residuals, random errors), vertical distance between child P_{ij} and the j^{th} school line (fixed slope for model (3.1a), resp. varying slopes for model (3.1b));
s_j	: (random slopes), difference between the varying slope for j^{th} school line and the slope for the overall regression line.

In models of Figure 3.1, from the color of the points, 2 distinct groups of children - children coming from 2 different schools - can be identified. The schools have different intercepts and in model (3.1b), school 1 has a steeper slope than school 2. This difference across groups (test group / control group, schools, classes) will be modeled later throughout the statistical analysis of the thesis.

3.3.3.4 Fitting Generalized Mixed-Effects Models

Generalized Mixed-Effects Models are defined similarly to traditional linear models, except that random effect(s) are specified in addition to fixed effect(s); hence the name *mixed models*. According to Henrik Singmann and David Kellen, *"The most important concept for understanding how to estimate and how to interpret mixed models is the distinction between fixed and random effects." (Singmann and Kellen, 2019, pg. 5)* Fitting variables as fixed effects generally allows us to statistically **compare means** of groups (test group / control group, schools, classes). Random effects, for their part, **estimate the variance** between the groups, not the mean of each group (Theobald, 2018). In some contexts, fixed effects *"are referred to as the population average effect" (Clark, 2019)*. Random effects enable us *"to factor out the idiosyncrasies of our sample and obtain a more general estimate of the fixed effects of interest" (Singmann and Kellen, 2019, pg. 6)*.

> *"In theory, the random effects in mixed models can have any plausible distribution. Lee and Nelder* (1996) *and Lee et al.* (2006)*, for example, discuss doubly generalized linear models, in which random model effects as well as the response variable may have non-Gaussian distributions." (Stroup, 2012, pg. 11)*

- Model fitting random intercept(s), corresponding to (3.1a) of Figure 3.1:

$$\text{intercepts\_model} \leftarrow glmer(\underbrace{\text{testOutcomes}}_{\text{response variable}} \sim \underbrace{\text{initialRanking}}_{\text{fixed effect}} + \underbrace{(1 \mid \text{school})}_{\text{random effect}}, \quad \text{data} = \cdots) \tag{3.3}$$

The intercept differs across the 2 schools, but the slope of the school lines is the same. Random effects structure with random intercepts only is used for data on groups (test group / control group, schools, classes) expected to differ from one another at a starting time (testing), but for whom the effect of initial ranking (e.g.) on test outcomes is the same.

- Model fitting random intercept(s) **and** random slope(s), corresponding to (3.1b) of Figure 3.1:

$$\text{intercepts\_slopes\_model} \leftarrow glmer(\underbrace{\text{testOutcomes}}_{\text{response variable}} \sim \underbrace{\text{initialRanking}}_{\text{fixed effect}} + \underbrace{(1 + \text{initialRanking} \mid \text{school})}_{\text{random effects}}, \quad \text{data} = \cdots) \tag{3.4}$$

The intercept and the slope of the school lines vary; groups (test group / control group, schools, classes) are expected to differ from one another at a starting time (testing) **just like** the effect of initial ranking (e.g.) on test outcomes across the groups.

3.3.3.5 Specifying random effects structure

According to Dale J. Barr (et al.) from the Institute of Neuroscience and Psychology, University of Glasgow (UK), the maximal random effects structure *justified by the design* should be included in the fitted model; i.e. the maximal random effects structure should fit random intercepts, random slopes, and their correlation for any fixed effect(s) and for any interaction(s) of fixed effects in the model (Barr et al., 2013; Williams, 2018). In particular, **Nested and Crossed Random Effects** are used in the present study.

- Models with Nested Random Effects:

> *"Traditional multilevel models involve hierarchical data structures whereby lower-level units such as students are nested within higher-level units such as schools and where these higher-level units may in turn be nested within further groupings or clusters such as school districts, regions, and countries. With hierarchical data structures, there is an exact nesting of each lower-level unit in one and only one higher-level unit." (Leckie, 2019b, pg. 2)*

In our research, a variety of children (girls, boys, those who like maths, or those who don't) were nested within separate classes, which were nested within separate schools, which were nested within the test group / control group (a four-level nested structure): *"Each of these nested levels, $\cdots$, is important to account for with random effects." (Theobald, 2018, pg. 4)*

These nested random effects can be formulated as follows:

$$\underbrace{(1 \mid \text{group/school/class/child})}_{\text{nested random effect}}, \quad \text{data} = \text{nested\_data} \tag{3.5}$$

where the random effects structure defines random intercepts for children, adjusting for similarities for children within the same class, school and group (Williams, 2018).

Regarding the learning assignment items, they are nested within 5 domains (cf. Figure 3.2).

- Models with Crossed Random Effects:

> *"In cross-classified data there is not an exact nesting of each lower level unit in one and only one higher level unit. Rather, lower level units belong to pairs or combinations of higher level units formed by crossing two or more higher level classifications with one another." (Leckie, 2019a, pg. 3)*

In the case of our study, some random effects are crossed as all the children took the same test twice, once as a pre-test and again as a post-test (repeated-measures design). Since all the covered domains appear both in the pre-test and in the post-test (cf. Figure 3.2), all participating children passed all 5 domains twice, first during an initial testing phase and later during the post-test session (Baayen et al., 2008). Child 01 in the pre-test was not independent from child 01 in the post-test.

The crossed random effects can be formulated as follows:

$$\underbrace{(1 \mid \text{child}) + (1 \mid \text{test}) + (1 \mid \text{domain})}_{\text{crossed random effects}}, \quad \text{data} = \text{crossed\_data} \tag{3.6}$$

We have the usual extensions here of the already provided formulations (3.5), (3.6); for both nested and crossed random effects, random slopes could also be specified.

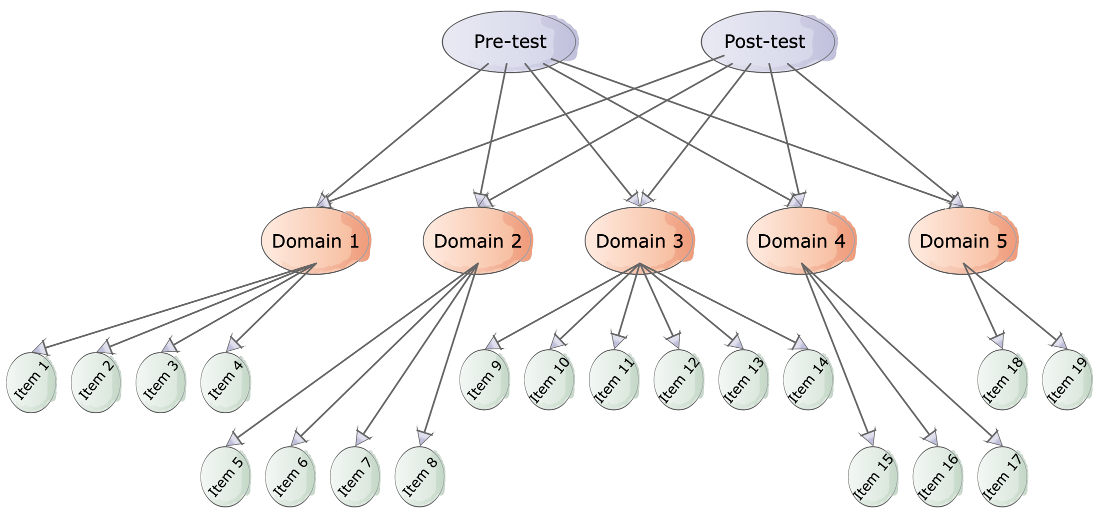


Figure 3.2: Learning assignment items are nested within the domains; every domain belongs to every test, i.e. to both the pre-test and post-test (crossed design)

3.3.3.6 Measurement and insights via Generalized Mixed-Effects Models

> *"Mixed-effects models may offer substantially enhanced insight into how subjects are performing in the course of an experiment, for instance, whether they are adjusting their behavior as the experiment proceeds to optimize performance." (Baayen et al., 2008, pg. 409-410)*

Is there any (significant) difference between subjects belonging to the test group and those belonging to the control group, within gender groups, children from consecutive classes, children who do or do not like maths, children taught in different cities or children from the same school?

> *"⋯ mixed-effects modeling allows researchers not only to control for individual variability when identifying relationships between predictors and outcomes but also to identify how the variability in individual characteristics, learning material, and training strategies specifically contributes to learner outcomes (Misangyi, LePine, Algina, & Goeddeke,* 2006*). This is vital to improving ⋯ educational practice given the variability in individual responses to interventions and classroom instruction (Kelley, Leary, & Goldstein,* 2018*)." (Gordon, 2019, pg. 508)*

For example, we provide measurements of the degree to which individual factors, like belonging to the test group or control group (different teaching methods), being in a ISCED 1.5 or ISCED 1.6 class (different package of basic knowledge of geometric concepts), liking or not liking maths and being tested on domains 1, 2, 3, 4 or 5, affect learner outcomes.

As *"potential differences in achievement between female and male students have always been an interesting topic in educational research,"* and particularly the so-called *'gender gap'* in mathematics and science (Meinck and Brese, 2019, pg. 1), the gender example is used for the explanations of Table 3.2. This table shows the main comparisons made between the levels of different variables in achievement (e.g. gender includes *two* levels: girls and boys). To test whether the levels of a variable are different by means (e.g. whether boys and girls have different learning outcomes), the variable (gender, with girls and boys as levels) needs to be specified as a fixed effect in the relevant multilevel model. Random effects group variables or factors (variance components between the levels) as the part of the design over which a researcher wants to generalize (Singmann and Kellen, 2019) (for gender, these variables are test, domain, group and group/class (cf. Table 3.2)). This means we could, for instance, compare the differences in achievement between girls and boys in the pre-test to those in the post-test, also termed comparing the *'gender gap'* in the pre-test to that in the post-test (generalization over the variable *test*). It should be noted that fixed effects can also serve as random effects, and vice versa.

Similar analyses are run, in this case for differences in city, grade (ISCED 1.5 / ISCED 1.6) and the personal statement *'do / do not like maths'*.

Figures presenting overall trends (predictions from different models), like differences in children's pre- and post-test performance, class or test vs. control group, will illustrate the outcomes and conclusion supporting either the rejection or acceptance of hypothesis *one*. Furthermore, plots of estimated random effects (e.g. for each class) and their interval estimates are shown. The complexity of the models may increase throughout the statistical analysis (Singmann and Kellen, 2019); features of each model will become clear.

Table 3.2: Hypothesis *one*: comparisons made between the levels of different variables in achievement, and the corresponding sources that help us measure how much variation between the levels is attributable to random effects

variables	comparisons of means made between the levels, [a]	variance components between the levels (random variability in the data coming from different sources)
test	between pre-test and post-test [b]	- group (test group, control group) and - group/class (class c, $c \in \{1, 2, \cdots, 17\}$) and - group/class/gender (girls, boys) and - grade class (ISCED 1.5, ISCED 1.6) and - domain (domain d, $d \in \{1, 2, \cdots, 5\}$)
group	between test group and control group [c]	- test (pre-test, post-test) and - domain (domain d, $d \in \{1, 2, \cdots, 5\}$) and - group/class (class c, $c \in \{1, 2, \cdots, 17\}$) and - group/class/gender (girls, boys)
school	between the s schools, $s \in \{1, 2, \cdots, 9\}$	- test (pre-test, post-test) and - school/class (similarities for classes within the same school)
gender	between girls and boys	- test (pre-test, post-test) and - domain (domain d, $d \in \{1, 2, \cdots, 5\}$) and - group (test group, control group) and - group/class (class c, $c \in \{1, 2, \cdots, 17\}$)
grade class	between ISCED 1.5 and ISCED 1.6	- test (pre-test, post-test) and - domain (domain d, $d \in \{1, 2, \cdots, 5\}$) and - group (test group, control group) and - group/class (class c, $c \in \{1, 2, \cdots, 17\}$)
maths	between do or do not like	- test (pre-test, post-test) and - group (test group, control group) and - group/class (class c, $c \in \{1, 2, \cdots, 17\}$) and - group/class/gender (girls, boys)

[a] All the comparisons were made for post-test_all and for post-test_restricted.

[b] As there is strong evidence that the children gained additional knowledge during the 500 minutes of geometry instruction, either via the GEOGEBRATAO tool or the more traditional format (paper-and-pencil), there must be a significant difference in the scores of the pre-test and the post-test.

[c] We wish **no** significant difference(s) in the scores between the test group and the control group in favour of the control group for every comparison under discussion.

In the next subsection, we'll get on with the methodology for testing hypothesis *two* that exclusively relates to the test group, unlike hypothesis *one* that focuses both on the test group **and** the control group.

3.3.4 Test of elaborated hypothesis *two*

3.3.4.1 Introduction

Hypothesis *two* predicts that student-centered software allows teachers to create a blended learning environment in which children learn geometric concepts autonomously, thereby reducing teacher interventions for the children in need of (individualized learning) support. It mainly refers to our software data collected through TAO (Kreis et al., 2018, 2010b) and is thus entirely related to the test group.

3.3.4.2 Methodology used

To explore and ultimately accept or reject hypothesis *two*, we use **clustering methods** to *"classify observations, within a data set, into multiple groups based on their similarly" (Kassambara, 2017, pg. 16)* in combination with **Generalized Mixed-Effects Models** (cf. subsection 3.3.3).

We use `PAM` (Partitioning Around Medoids) clustering, also called `K-medoids` clustering, in which, *"each cluster is represented by one of the objects in the cluster" (Kassambara, 2017, pg. 16)*. `PAM` clustering is less sensitive to outliers compared to `K-means` clustering, which is one of the most popular partitioning algorithms in clustering. For our clustering strategy and method selection, we base our research on a chapter in the Handbook of Cluster Analysis (Hennig, 2015).

`PAM` cluster analysis allows us to classify the children into subsets based on similar performance on both the pre- and post-test and based on the child's progress within the sequence of computer-based activities. Each child will belong to exactly one subset, called a cluster.

According to Christian Hennig, the ideal number of clusters is not uniquely defined and *"there is no unique 'true' number of clusters"*. Even though we have chosen `PAM` clustering, the number of clusters is still ambiguous.

> *"Ultimately, how strong separation between different clusters is required and a partition into how many clusters is useful in the given situation cannot be decided by the data alone without user input. It is often suggested in the literature that the number of clusters needs to be 'known' or otherwise it needs to be estimated from the data." (Hennig, 2015, pg. 26)*

In this study, the desired number of clusters K will be specified using the `NBCLUST` library (Charrad et al., 2014), probably equal to 4 or 5, so that we might have one cluster containing the higher performing children, another containing lower performers and 2 or 3 clusters containing the average ones.

Thereafter, we will fit various Generalized Mixed-Effects Models in order to provide evidence of some common ways of working with our dynamic geometry tool that lead to overall improvement at an individualized level. To analyse the children's common ways of working within each cluster and to compare the different clusters, we focus on the affordances available to the children, i.e.

- the *short* video clips that the children can consult voluntarily;
- the *overlay* video clips that are imposed in some specific error scenarios;
- the loops, repetitions of prior learning assignments, that are also imposed in some specific error scenarios;
- any specific learning assignments for which the teacher had to be called, which is imposed by the GEOGEBRATAO tool.

A thorough analysis within the clusters will enable us to discuss the rejection or acceptance of hypothesis *two*. Are the children able to work independently with our software and to learn autonomously? Is there any cluster in which the children are particularly stuck (significantly more loops and more viewings of *overlay* video clips)? Or just the opposite: is there any cluster in which the children just ‘soar through’ the activities (no loops, significantly fewer viewings of *short* video clips)? In the first case, we would have overwhelmed children working with our software, and in the second case we might have bored or under-challenged children, though we are not expecting this to be the case. And finally, does this have an effect on the children’s post-test scores?

This chapter described the methodology for our sample selection (cities, schools, children), the data collection methodology, along with the methodology used to test, assess and validate our elaborated hypotheses. The methodology used for testing hypothesis *one* is fairly common, whereas that for hypothesis *two* requires a more in-depth analysis and many interpretations of possible conclusions within the different clusters of children.

Let us now set our pedagogical approach in relation to existing research regarding learning and the use of ICT.

Chapter 4

State of the art / theoretical framework

This chapter first takes a look at some current aspects of Information Communication Technologies (ICTs) related to our project. Thereafter, we explore the theoretical foundations which serve as a valuable basis to our thoroughly designed learning approach. In this chapter, we do not provide any theoretical foundations for the methodology of our statistical analysis; for convenience, this will be done in the relevant section(s).

4.1 Broadening the use of ICT

4.1.1 Introduction

ICTs have significantly changed how people interact with each other (Gebhardt et al., 2019). Almost all schools have made use of ICTs as teaching tools in the 21^{st} century (Hardman, 2019), so there is extensive research literature on the use of ICT in education. The study of a UK primary school carried out in 2007, *'Fitting it in'*, already demonstrated effective use of ICT to support teaching and learning (Cartwright and Hammond, 2007).

Battelle for Kids, a national not-for-profit organization committed to collaborating with school systems and communities, strives to *"realize the power and promise of* 21^{st} *century learning for every student — in early learning, in school, and beyond school — across the country and around the globe" (Battelle for Kids, n.d.).* Figure 4.1 represents the different components of the framework for the 21^{st} *Century Learning* (overview of some funded research projects for the *Learning and Innovation Skills (noted* $4Cs$*)* from $2005 - 2016$ in (Abdulla and Runco, 2018)). Kurt F. Geisinger presents a brief but broad view of some models and approaches existing for identifying and addressing 21^{st} century skills in educational settings (Geisinger, 2016).

We are interested in the $4Cs$ skills, which include, among other things, Critical Thinking and Creativity. Children need to be *"able to look at problems in new ways and link learning across subject areas"* and *"willing to try new approaches to get things done" (Zalewski and Cairns, 2011 & 2015, pg. 3).* The paper *'Assessing* 21^{st} *Century Skills: Integrating Research Findings'* synthesizes research evidence pertaining, i.a., to critical thinking and creativity (Lai and Viering, 2012). Professor Stephen Lamb et al. generically define critical thinking as a skill that *"refers to the ability to assess the value of a claim or information and come to a conclusion about what to believe or to do about it." (Lamb et al., 2017, pg. 19)* Among other things, *"instructional interventions support the development of critical thinking skills" as referred in (Lamb et al., 2017, pg. 20).* Frequently related to critical thinking is creativity in discussions on skills: *"Developing creative skills is accessible to all students when adequate didactical and pedagogical conditions exist." (Lamb et al., 2017, pg. 21)*

> *"The Partnership for 21^{st} Century Skills presented the 4Cs of critical thinking and problem solving, $\cdots$, and creativity plus innovation, as the super skills in the 21^{st} century because they are foundational essentials for success in college, $\cdots$, and life outside educational institutions." as written in the overview paper that explores the pedagogical meaning and implications of the 4Cs skills (Kivunja, 2015, pg. 235)*

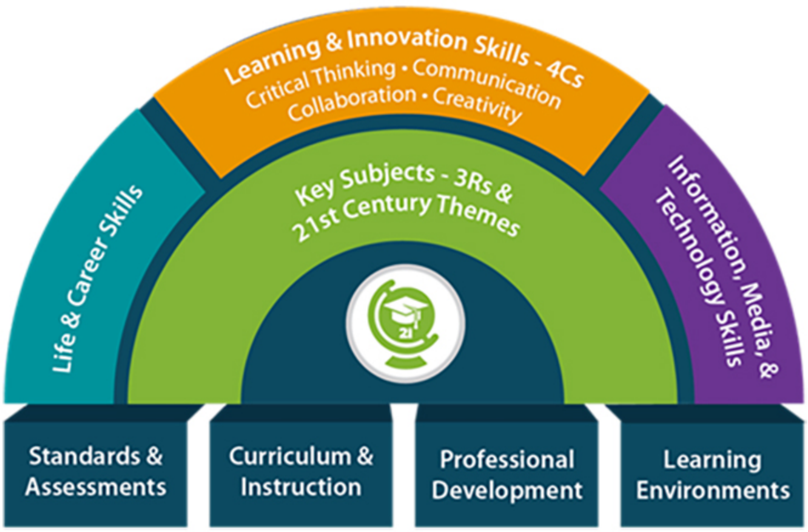


Figure 4.1: Framework for the 21^{st} *Century Learning* (Battelle for Kids, n.d.)

The report of the *European Digital Competence Framework for Citizens* presents its proficiency levels (from basic with guidance to most advanced and specialized), for each competence, defined through learning outcomes. Each level represents a progression in students' acquisition of a digital competence in accordance with the cognitive challenge of the level, the complexity of the tasks the students can handle and the students' autonomy in completing the task. For example, at basic level and with autonomy and appropriate guidance where needed, the child should learn to choose from a list of digital maths resources prepared by their teacher an educational game that helps them to practice their mathematics skills. At an intermediate level, independently, according to their own learning needs, the child should be able to tell their teacher which digital activities and pages they surf in order to keep their digital competence updated so that they can profit the most from digital learning platforms. (Carretero et al., 2017)

The *Digital Education Action Plan* 2018 − 2020 from the European Commission (European Commission, 2018) tries to support the use of technology in education and the development of digital competences, i.e. it presents measures to help education *"to reap the opportunities and meet the challenges presented by the digital age", as referred in (Glezou, 2020, pg. 6). Three* of the plan's main priorities are how to make better use of digital technology for teaching and learning, how to develop digital competences and skills, and how to improve education through better data analysis and foresight.

The International Association for the Evaluation of Educational Achievement (IEA) also acknowledges that *"Education is changing. No longer just focused on the acquisition of a knowledge-based curriculum, countries worldwide are expanding their educational vision to include what are often termed 21^{st} century skills ($21CS$)." (IEA, 2021, pg. 1)* Unfortunately, *"there is no consensus on what is meant by $21CS$ or how these skills are included in countries' curricula"*. Therefore, IEA is *"implementing a new, comparative, curriculum mapping study"*, called the $21CS$ MAP. Let us mention how Luxembourg approached / approaches $21CS$ integration.

In 2015, the Luxembourg Ministry of Education, Children and Youth, presented *Digital4Education*, a national digital strategy in the context of the transformation of a country into a digital nation (Linckels et al., 2019). It promotes new learning strategies and innovative educational projects, using digital in schools, thereby developing skills and know-how fit for the 21^{st} century. *Digital4Education*'s goals are to create equal opportunities for innovation and equal access to ICT tools [1], become hubs for the future (integrate ICT-related skills into the curriculum) and open a new sphere in education (modern, multimedia-based lessons) (Digital Luxembourg innovative initiatives, n.d.). Similar to our project, one of the educational challenges consists of using the potential of ICT to address the many different needs of children. The Ministry tended/tends to focus on training future specialists in the ICT sector. (eduSphere, 2015)

Most recently, in February 2020, the Luxembourg Ministry of Education, Children and Youth, announced a new strategy for digital education from primary school to secondary general and classic education, i.e. a comprehensive digital education strategy for all pupils, called *'Simply digital - future competences for strong children'*. It aims to strengthen the digital competences of the students. We are particularly interested in the primary grades, in which the children should be able to understand, from an early age, how machines *'think'* and react. Furthermore, they should acquire basic programming knowledge. The ministry's approach emphasizes that

> *"the importance of focusing efforts on creativity and cooperation, proper human skills that machines are incapable of, while mastering new technologies. Humanistic education has more relevance than ever." (CEDEFOP, 2020)*

According to Marian Henry, *"It is important to teach children about, with and through* ICT *not simply because it 'enchants the disenchanted' child or because it makes learning 'more fun' (even though these are important), but because the children we teach live in an age where* ICT *is a core element of how we learn, work, play and connect with and contribute to society" (Younie et al., 2014, pg. 10).*

[1] *"reduce the digital divide, granting all young people, independent of their social background, access to quality educational resources" (eduSphere, 2015)*

4.1.2 Barriers for the use of ICT

A series of barriers prevent teachers from making full use of ICT in their daily work; some barriers are material or organisational, while others may be psychological or have to do with the personal characteristics of the teacher. Carless broadly divides the main barriers into *three* categories: 1) teacher-related, 2) system-related and 3) school-related (Carless, 2012). A most recent paper, among other things, looks into the barriers and facilitating factors for innovation in education that is often related to technology and new inventions (Rahmat, 2020). Examining barriers about technology integration from Kindergarten to Grade *Six* (midwestern United States), Pi-Sui Hsu observes: *"students' lack of computer skills, teachers' lack of training in technology, teachers' lack of time to implement technology-integrated lessons, and teachers' lack of technical support" (Hsu, 2016).*

The constant and fast-moving development of technology, its rapid change of the everyday world and the classroom, and the fact that many pupils begin primary school as highly skilled and experienced users of technology (Beauchamp, 2014, pg. 210) constitutes a psychological barrier for teachers. They need to quickly and better understand *"how new and emerging digital tools can be used effectively"*, while realizing that access to technological education can advance or impede pupils' educational successes (Sinclair et al., 2017, pg. 282; Schilmoeller et al., 2018).

Pedagogy does indeed change with ICTs, but the exact nature of this change remains opaque (Hardman, 2019). Teachers must, among other things, prepare their pupils to flourish in this new learning environment by *"giving them the opportunity to develop critical, creative and collaborative skills" (Henry, 2014, pg. 10).* In the book *'Teaching and Learning with* ICT *in the Primary School'*, Marian Henry describes why teachers should help their pupils to develop these skills and how teachers can confidently foster them in the classroom; what can act as a barrier to teaching and learning with ICT.

According to Gina Blackberry and Deb Woods, it is *"unrealistic to expect that teachers will be able to transform their classrooms with* ICT *and develop a co-constructed pedagogy quickly or easily" (Blackberry and Woods, 2014, pg. 139)*, i.e. a pedagogy in which teachers rather than seeing themselves as the teachers-in-charge they relinquish some of their control in order to integrate ICTs to *support* the learning and genuinely engage children in their classrooms into their approach (Williams, 2014, pg. 278; Kervin and Mantei, 2010); thereby the teachers' and the children's creative energy should be unleashed. The teachers need to *"reflect on their practice, together with their thoughts and feelings about pedagogy and* ICT*"* and to *"be supported to make small, incremental changes"* (Blackberry and Woods, 2014, pg. 139). Gina Blackberry and Deb Woods argue that *"teachers' thinking and self-determination about* ICT *and pedagogy are powerful predictors of their preparedness to change"*, and thereby help to overcome some barriers. Clearly, teachers' attitudes and beliefs towards ICT are fundamental for successful implementation of new technologies in European countries (Eickelmann and Vennemann, 2017) may, however, represent a barrier to overcome for some teachers (Hohenwarter et al., 2017).

Another barrier is effective management of technology in a multicultural classroom (Chisholm, 1995), since *"the field of technology operates somewhat differently where multicultural educational issues are concerned" (Schilmoeller et al., 2018, pg. 205).*

Assessment practices tend to be adapted, radically changed or replaced with the use of ICT; this can also represent a barrier for some teachers. They must involve parents and help them gain a better understanding of how their child is performing (Beauchamp, 2014). In particular, parents must become aware that in the primary school, children's *"achievements can often be spontaneous and transitory, and it is only through the use of technology that some of these can be captured and shared" (Beauchamp, 2014, pg. 222)*. There are studies that analyse the potential impact of converting from paper-and-pencil to computer-based assessment, such as *The TIMSS* 2019 *Item Equivalence Study* (Fishbein et al., 2018).

In their literature review paper (LRP), Marthese Spiteri and Shu-Nu Chang Rundgren stress different factors which affect primary teachers' use of digital technology in their teaching practices, and provide a concept map to achieve greater understanding of this topic and the associated connections (cf. Figure 4.2):

> *"four influencing factors were identified: teachers' knowledge, attitudes and skills, which are also influenced by and influence the school culture" (Spiteri and Rundgren, 2018, pg. 115).*

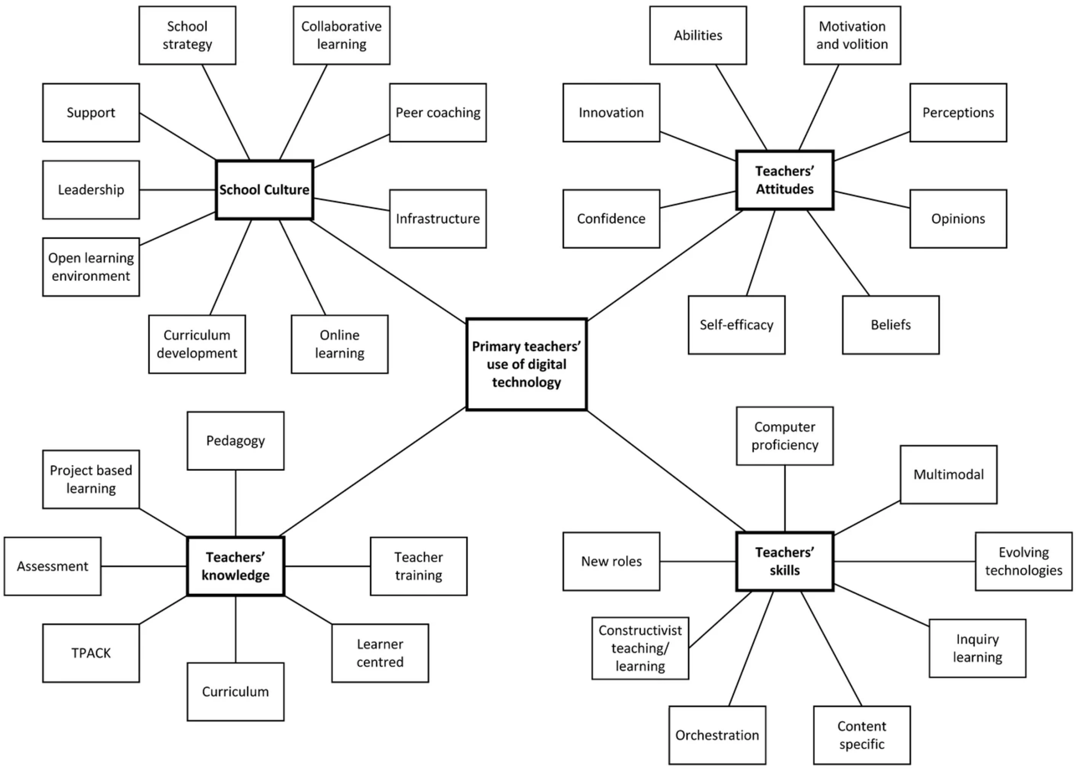


Figure 4.2: Concept map to categorize emerging factors which affect primary teachers' use of digital technology integration (Spiteri and Rundgren, 2018, pg. 119)

It should be noted that by 2004, BECTa ICT research identified the most successful factors in overcoming some barriers and encouraging the uptake of successful ICT practices by teachers (Scrimshaw, 2004). Hohenwarter et al. are referring to Hew and Brush's (Hew and Brush, 2007) description of strategies to overcome several barriers of ICT integration into teaching and learning of mathematics, e.g. by having a shared vision and technology integration plan, by changing attitudes and teachers' pedagogical beliefs (in particular, about how to teach align with how the teachers were taught at school), by reconsidering assessments (Hohenwarter et al., 2017).

Integrating ICTs in **mathematics** classrooms poses its own unique problems, and besides, mathematics independent barriers (Mulenga and Marbán-Prieto, 2018); recent reviews stress difficulties, challenges and barriers associated with technology integration into mathematics education (Viberg et al., 2020). These difficulties include *"altering the role of the teacher from instructor to facilitator"* and *"stressing a need for a structured research-based approach"* to achieve the full potential of technology (Bray and Tangney, 2017, pg. 257-258), as mentioned in (Viberg et al., 2020, pg. 2). Teachers are challenged; they have to design effective learning experiences that sound mathematical pedagogical approaches in a coherent educational context.

Spangenberg E. and de Freitas G. investigated the mathematics teachers' levels of technological pedagogical content knowledge (TPACK, cf. teachers' knowledge factor, on the lower left of Figure 4.2) and the barriers to integrating ICT in the classroom affecting the continuous professional development of mathematics teachers (Spangenberg and De Freitas, 2019). According to Morsink P. M. et al. (Morsink et al., 2011), *"developing expertise in* ICT *integration is a time-consuming, long-term process that requires commitment and ongoing effort from teachers" (Spangenberg and De Freitas, 2019, pg. 3)*. A couple of study participants also revealed that curriculum-related time constraints serve as a persistent barrier to an effective integration of ICT in the classroom.

The project *"Maths Teachers' Adventure of* ICT *Integration" (*MTAII, *http://www.mtaii.com)*, funded by the Erasmus+ Programme of the European Union, has as purpose to help teachers overcoming the barriers of knowledge and skills (cf. lower part of Figure 4.2) in relation to the use of digital technology in mathematics teaching (Hew and Brush, 2007). This should be achieved by creating a professional development environment for teachers that includes some strategies, such as changing attitudes and beliefs, or conducting professional development. (Hohenwarter et al., 2017)

As ICT affordances in mathematics education *"enhance access to core mathematical concepts through dynamic representations and classroom connectivity" (Hegedus and Moreno-Armella, 2020, pg. 380)*, it's worth overcoming the barriers for the use of ICT, including the barriers to implementing a dynamic geometry software (DGS) in the classroom (Little, 2009). The next subsection illustrates, among others, the important role of ICT in mathematics (especially in geometry) teaching.

4.1.3 Using ICTs in mathematics, specifically GeoGebra in geometry

ICTs for mathematics attainment can positively impact primary school performance. However, a constructivist pedagogy must be used as opposed to a traditional transmission-based pedagogy (Hardman, 2019). This conclusion is based on a review of *ten* years of studies conducted in the field of ICT impact(s) on mathematics outcomes in elementary school and related pedagogical practices. Therefore, *"a shift in school routines when incorporating technology into the mathematics classroom"* should be observed (Dockendorff and Solar, 2018, pg. 66).

Several studies focus on the use of DGS (dynamic geometry software) in education; a notable example is the use of GEOGEBRAPRIM software in primary schools in the Grand Duchy of Luxembourg (Kreis et al., 2010a, 2018, 2010b); others include findings that movement fosters understanding (Haftendorn, 2005) and that "DGS *deeply changes geometry if it is taken as a* *human activity integrating the use of modern instruments like* DGS*" (Straesser, 2002, pg. 319).* The use of the GEOGEBRAPRIM software, which represents the springboard for my thesis, was highlighted in section 1.1.

Kyeong-Sik Choi (2010) investigates educating gifted students in mathematics via computer in the context of student-centered GeoGebra classes (Choi, 2010). Daisy E. Vasquez (2015), a master's student from California State Polytechnic University, Pomona, studies how to enhance student achievement using GeoGebra in a technology rich environment and the effects that GeoGebra has on the student comprehension and retention of maths concepts (Vasquez, 2015). Vasquez writes,

> *"Qualitatively, I, ⋯, found that using GeoGebra to learn geometric transformations greatly increased students' motivation and engagement in learning the material. ⋯ I did not feel that the students in the control group had the same deep visual understanding ⋯. Students became more interested in their learning with the use of the software because it provided a dynamic, hands-on, and discovery learning environment." (Vasquez, 2015, pg. 48,50,52)*

Using GeoGebra-based dynamic applets in an exploratory way promotes mathematical processes and thereby captures student attention (Dockendorff and Solar, 2018).

Another study shows that Geogebra is an amazing scaffolding tool in developing students' mathematical thinking about analytic geometry, both for higher and lower performing students, because it allows them *"to visualize and explore the insight of mathematics with optimum possibilities"* (Khalil et al., 2019, pg. 427). GeoGebra is useful in enhancing the mathematics learning process with dissimilar spatial visualization. Experimental students performed better in the post-test compared to the control group students, who followed traditional teaching methods, but there was no significant difference in the performance of students having different types of spatial visualizations (Bakar et al., 2015).

GeoGebra dynamic software use has an impact on teachers' conceptions about teaching and learning mathematics. In fact, when teachers appropriate and familiarize themselves with GeoGebra tools and applications, they somewhat rediscover school mathematical content through technology. This allows them to introduce a more exploratory methodology into their school teaching practice. The use of GeoGebra *"impacts practicing teacher's paradigms, perceptions and performance on how teaching and learning mathematics takes place"*. Using GeoGebra tends to enable *"a more visual approach to mathematics, generating an increasing acknowledgement of the importance of representations for learning"*. A prospective teacher *"also emphasizes the contribution of ICT in promoting for teachers a mediating role between technology and knowledge acquisition"*. (Dockendorff and Solar, 2018, pg. 82)

> *"The ubiquitous presence of dynamic mathematics learning technologies has pushed mathematical understanding to the forefront of mathematical education practice, research, and development." (Bu and Hohenwarter, 2015, pg. 372)*

Beyond the simple use of GeoGebra on a two-dimensional screen, modeling with dynamic technology becomes highly interesting and pedagogically powerful in mathematical education. This empowers both students and their teachers *"to experience genuine mathematics in the real* *world"*. Furthermore, through the meaningful combination of Science, Technology, Engineering, Arts and Mathematics, called 'STEAM fields', students and teachers are able to *"construct, diagnose, evaluate, and appreciate mathematical meaning in a thoughtful, emotional, and ultimately aesthetic manner"* (Bu and Hohenwarter, 2015, pg. 373). In particular, some potential implications for both teachers and students were shown when connecting physical (real-world situations) and digital ($3D$ dynamic geometry) worlds through the integration of new technologies ($3D$ printing), creating meaningful mathematical tasks within and, in a certain sense, outside classrooms (Lavicza et al., 2020).

This recent technological emergence in mathematics education offers new directions for mathematics teaching and learning that leave increasingly less room for traditional paper-and-pencil manipulations (Bu and Hohenwarter, 2015); this is sometimes still difficult for teachers and parents to accept. Instead, there is more room for developing skills to acquire deeper understanding of students' environments and everyday situations (Lavicza et al., 2020).

Therefore, we can say that the pallet of technical possibilities in school mathematics which help students to foster mathematical understanding is greatly increasing by incorporating the ability to infuse aesthetic experiences into students' mathematical performance (Bu and Hohenwarter, 2015). When children is given the possibility to develop the capacity of mathematical problem-solving with technology (e.g. with GeoGebra), a key issue is, how this can be tackled to promote the children's success in mathematics in our digital era (Carreira et al., 2016; Jacinto and Carreira, 2017).

4.1.4 Parallels with the GEOGEBRATAO project

Gerry Stahl and The VMT (Virtual Math Teams) Project Team have created dynamic-geometry activities with GeoGebra intended to show teachers how to promote collaborative learning using the tool: *"Collaborative learning involves a subtle interplay of processes at the individual-student, small-group and whole-classroom levels of engagement, cognition and reflection."* Matching these processes, Gerry Stahl et al. design and structure the activities to *"seek a productive synthesis of collaboration, discourse, visualization, construction, and argumentation skills applied in the domain of beginning geometry"*, thereby enhancing motivation, extending attention and spreading understanding. (Stahl and The VMT Project Team, 2013, pg. 100)

Just as we did when creating the GEOGEBRATAO tool, Gerry Stahl et al. rely on *van Hiele's Model of Geometric Thinking* (Vojkůvková, 2012), i.e. on the principle that children *"who are at a given level cannot properly grasp ideas presented at a higher level until they reach that level"*. Children pass through the levels 'step by step'. Thus, both in Stahl et al.'s study and in our project, *"a developmental series of activities pegged to the increasing sequence of levels is necessary to effectively present the content and concepts of geometry"*. According to Stahl et al., *"Failure to lead students through this developmental process is likely to cause student feelings of inadequacy and consequent negative attitudes toward geometry."* (Stahl and The VMT Project Team, 2013, pg. 100-101)

Stahl et al. also refer to the mathematician de Villiers who also embraces letting students pass through levels 'step by step'. Furthermore, de Villiers uses constructions that can initially be highly scaffolded through instruction and collaboration so students can be guided in their work.

Scaffolds make tasks easier than they would be if offered with no such guidance. The provided scaffolds are gradually reduced so that students can be brought to a stage where they are ready to perform on their own based on their exploratory experiences. (Stahl and The VMT Project Team, 2013; de Villiers, 2003; Cantrell, 2000)

Other researchers, Radović et al., focused in their study on the design of an interactive learning textbook (eBook), created with GeoGebra applets and up-to-date Web Technologies. According to them: *"The eBook is designed to meet the pedagogical and didactic needs of learners, allowing a high degree of interactivity and feedback during the learning process."* Hence, some parallels between this eBook and the GeoGebraTAO tool can be drawn. It should also be noted that in Radović et al.'s project, students emphasized that *"tasks with interactive applets and new kind of learning materials inspired them to learn more, both in school and at home"*. (Radović et al., 2020, pg. 32)

4.1.5 Gender Differences in ICT use

'Gender gaps' among students in ICT use are of great research interest, because ICT is a core element of how we learn, work and live in modern societies. (Gebhardt et al., 2019)

In particular, a series of in-depth analyses based on IEA [2] data show that in general, *"female students performed relatively better on tasks that involved communication, design, and creativity, and male students generally performed relatively better on more technical tasks"*. Furthermore, female students lack more self-confidence in their own ability to perform specialized ICT tasks than male students. Also of interest is that no appreciable differences in the pedagogical use of ICTs between female and male teachers were detected. (Gebhardt et al., 2019, pg. 69-70)

The TIMSS [3] study *"provides an overview of the so-called 'gender gap' in mathematics and science knowledge, based on an in-depth analysis of both extremes of student ability distributions"*. It notes that while *"the gender gaps that existed 20 years ago have persisted, gender equality in education has increased"*; there seems to be a *"persisting trend of more male students in the group of high-achievers for both mathematics and science in many educational systems"* (Meinck and Brese, 2019, pg. 1&20). These observations are not ICT-related, but will be of great interest in the next TIMSS study since, from the 2019 cycle onwards, an innovative computerized version of TIMSS will be used to investigate more complex areas in mathematics and sciences. (IEA Study TIMSS, n.d.)

Other studies highlight, among others, that *girls perform better than boys on performance-based* ICT *literacy assessments*, that *gender differences are larger in primary schools than in secondary schools* (relevant to our project), that *overall, the gender differences in* ICT *literacy are significant but small* (Siddiq and Scherer, 2019) and that *males still hold more favorable attitudes toward technology use than females* (Cai et al., 2017).

As already mentioned in subsection 4.1.4, it seems to be *"very appropriate to divide the teaching of geometry into different levels according to van Hiele theory"*. It should be noted that Haviger J. and Vojkůvková, I. show in their paper that *"van Hiele levels are equally suitable for*

[2] IEA, International Association for the Evaluation of Educational Achievement
[3] TIMSS, Trends in International Mathematics and Science Study, flagship study of IEA

both genders". (Haviger and Vojkůvková, 2014, pg. 977)

Another study, which focused on *'primary school students' interest, collaboration attitude, and programming empowerment in computational thinking education'* also found that boys show more interest in programming than girls (Kong et al., 2018). Let us further illustrate another important change in education: computational thinking.

4.1.6 Dynamic Geometry and Computational Thinking (CT)

Steve Easterbrook claims, *"If information and communication technologies (*ICT*) are to bring about a transformational change to a sustainable society, then we need to transform our thinking" (Easterbrook, 2014, pg. 235)*. This is why computational thinking is emerging in education, a trend which is *"apparent not only in computing science settings but also in* STEAM *fields* [4]*" (Chytas et al., 2019, pg. 1179)*, (Angelopoulos et al., 2020). Studies are carried out to promote CT skills in primary school children, enhancing geometry learning. (Echeverría et al., 2019; Sinclair and Guyevskey, 2018)

According to Wing, *"Computational thinking is taking an approach to solving problems, designing systems and understanding human behaviour that draws on concepts fundamental to computing. ⋯ It shares with mathematical thinking in the general ways in which we might approach solving a problem" (Wing, 2008, pg. 3717)*. (Wing, 2006)

It has been observed that, in addition to supporting the teaching of geometry, tasks based on dynamic geometry effectively support the teaching of many CT concepts. This seems to provide sufficient grounds for integrating mathematics (geometry) and CT in an interdisciplinary way (Sinclair and Guyevskey, 2018).

Further papers that we find relevant and interesting to the current initial stage of this very promising research field

- study the use of $3D$ technologies to support CT in STEM education, which asks whether *"$3D$ technologies can support the development of CT skills in students" (Angelopoulos et al., 2020, pg. 425)*;
- examine computational design project *"that are intended to be $3D$ printed and could potentially introduce coding learning actions" (Chytas et al., 2019, pg. 1173)*.

Considering that, up to now, quite little research has been done in the field of STEAM learner engagement, especially promoting CT skills to enhance geometry learning, we expect to see continued growth in these areas. We also hope to see increased research interest in the teaching and learning of geometry to improve mathematical understanding.

4.1.7 Technology-based assessments, specifically TAO framework

To complete this section, we describe the TAO framework, an open-source project, *"that provides a very general and open architecture for developing and delivering tests for the purpose of technology-based assessments" (Ras et al., 2010, pg. 644)*. The authors of the paper *'Using tangible user interfaces for technology-based assessment – Advantages and challenges'* agree

[4] already mentioned in subsection 4.1.3

with (Binkley et al., 2012) who argue that technology-based assessment

> *"has the potential to support educational innovation and development of 21st Century skills, such as complex problem solving, communication, team work, creativity and innovation" (Binkley et al., 2012), as cited in (Ras et al., 2012, pg. 8).*

Technology-based assessments offer a lot of advantages and will certainly replace paper-based testing in a number of areas over time (Csapó et al., 2009); Open Assessment Technologies explain in a white paper why online testing tools are the future of assessments (OAT, Open Assessment Technologies, 2019). When researchers and teachers integrate adaptive technology with a digital assessment platform in the classroom, then they are able to personalize the testing experience to meet their students' unique needs. Online assessment platforms help to provide deeper insights into the children knowledge and to identify possible gaps, as outlined in (OAT, 2020c).

Plichart et al. illustrated computer-based school system monitoring, *"ongoing assessment based on the TAO platform"*, with *"feedback to teachers done in Luxembourg"*. This feedback enables teachers to *"early detect shortcomings of their courses"* and decision makers of the educational system to receive *"information on the educational system steering efficiency"* (e.g. through the PISA studies and national initiatives) (Plichart et al., 2008, pg. 5065).

Further papers (marketing white papers) explain how digital assessment tools make it easier to prove learning outcomes (OAT, 2020d) and how to use adaptive testing in digital assessment to support learning (OAT, 2020a). Samantha Leonard describes empowering online education with digital testing technology (Leonard, 2020b) and building engaging online assessments with Technology-Enhanced Items (TEIs) (Leonard, 2020a); TEIs should be examined in light of the need by test makers to develop a strategy for incorporating them into large-scale assessments (Bryant, 2017). Technology-Enhanced Items may allow students to *"engage with content and show their understanding of the material in a range of ways" (OAT Practical Guide, n.d., pg. 5).*

Most recently, Open Assessment Technologies take a theoretical approach to building and designing digital assessment systems from the ground up (OAT, 2020b).

After having broadened and deepened our knowledge about the use of ICT, we will provide a theoretical understanding of our learning approach in the next section.

4.2 Towards theoretical foundations of the learning approach

4.2.1 Teachers' interventions

According once again to Mike Lansdown, *"the notion of challenge is fundamental to improvements in learning"*. We were driven by this belief while building the GEOGEBRATAO tool and tried to structure it accordingly. In fact, when children are not challenged appropriately, i.e. when the level is not high enough to avoid boredom and disinterest or so high as to cause feelings of anxiety and/or despair, children will make less or no progress in their understanding, knowledge and skills. Sometimes they show symptomatic behaviors reflective of being underchallenged. (Lansdown, 2020, pg. 18)

Teachers' interventions consist of attracting and retaining the children within the 'challenge zone' (Lansdown, 2020). Therefore, we refer to the principle used by Marjorie Siegel and Raffaella Borasi (Siegel and Borasi, 1992): *"Research (Siegel and Borasi, 1992) has shown that 'in order to acquire mathematical expertise in a durable and useful form, students need to construct mathematical knowledge and create their own meaning of the mathematics they encounter'* $(p.19)$*" (Barton and Heidema, 2002, pg. 3).* That's why, in our project, we stress the importance of forcing children to build new knowledge autonomously and at their own pace. The teachers were required to intervene as little as possible during the implementation of the GEOGEBRATAO tool and only to do so effectively (e.g. when the child makes very little progress toward a solution) to enable children to progress on their own. (cf. subsection 2.2.10)

4.2.2 Zone of Proximal Development (ZPD)

As described in section 1.2, *the Zone of Proximal Development (*ZPD*) is an area of learning sandwiched between what a child can do unaided and that which they can only achieve with accurate support from a more knowledgeable peer or adult who negotiates the gradual transfer of responsibility to their child* (Van Der Stuyf, 2002; Lansdown, 2020, pg. 12; Meyer and Turner, 2007). In fact, it is the zone where a child is able to tackle problems which they deem challenging, but which are achievable with the help and guidance of a 'more knowledgeable' person, like their teacher, another child or their parents. To provide clarity in his research, Mike Lansdown roughly calls this area 'the **challenge** zone', which is a more colloquially understood term (Lansdown, 2020). Furthermore, he defines two more areas:

1. 'The comfort zone', where a child faces problems that are currently too easy;
2. 'The panic zone', where a child faces problems that are currently too hard.

As already mentioned several times, our objective when building the GEOGEBRATAO tool was to keep children within 'the challenge zone'; this should likewise be the goal of any teacher intervention. It is important to realize that each child has their own 'challenge zone' / Zone of Proximal Development (ZPD). Therefore, every teacher, including the GEOGEBRATAO tool in its role as a twin teacher, should be able to provide instruction in accordance with each child's ZPD. (Isrokatun et al., 2019)

According to Charles Max, Figure 4.3 *"serves as a conceptual device and heuristic tool for improving the learning culture and managing its setup" (Max, 2011, pg. 70).* We have used our imagination to adapt the original figure (following Gibbons 2009) to our ways of thinking. As you see on Figure 4.3, there are $four$ clusters:

- a first cluster containing the child's moments of panic, frustration and anxiety;
- a second cluster containing the child's moments of comfort;
- a third cluster containing the child's moments of boredom (this cluster has not been specifically defined by Mike Lansdown);
- a fourth cluster containing the child's moments of challenge, learning and engagement.

Our goal is to minimize surface size of the first three clusters and to maximize surface size of the fourth cluster, i.e. to minimize the moments of panic, comfort and boredom and to maximize the child's moments of challenge (ZPD). The moments of panic, frustration and anxiety should be minimized to the greatest extent.

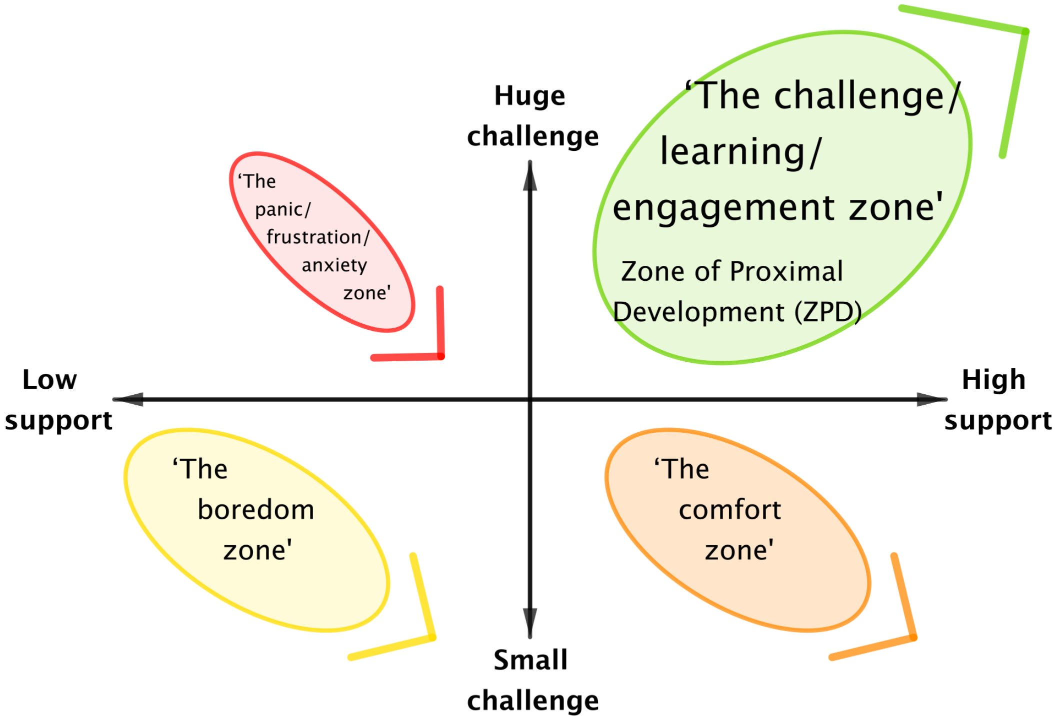


Figure 4.3: Locating a supportive and challenging learning culture within *four* kinds of learning environments (following Gibbons 2009) (Max, 2011, pg. 70)

4.2.3 Scaffolding student learning

First of all let us remember that scaffolding and feedback practices were described in detail in subsection 2.2.6. However, for those who are interested in scaffolded student learning (Hogan and Pressley, 1997) we will now provide some more relevant references.

Both children and their teachers benefit from judicious scaffolds provided in mathematical investigations or in another learning environment (McCosker and Diezmann, 2009), particularly considering that, all along, scaffolding *"aligns closely with the thinking of Vygotsky who saw children's learning as mainly resulting from social interaction" (Lansdown, 2020, pg. 16).*

Scaffolding, widely considered to be an essential element of effective teaching, teacher assistance and guidance, *"refers to a variety of instructional techniques used to move students progressively toward stronger understanding and, ultimately, greater independence in the learning process" (Scaffolding, 2015)*, provided in a *one*-to-*one*, peer or computer-based form (Belland, 2017). It is a style of instruction through which students are given *"the intellectual support to function at the cutting edge of their individual development" (Hogan and Pressley, 1997)*. Pupils would not be able to perform given tasks that are slightly beyond their ability without the provided scaffolds.

Scaffolding techniques *"can foster students' creative and divergent thinking skills, and enhance their independence, sense-making and self-confidence in mathematics" (McCosker and Diezmann, 2009, pg. 7-8)*. Scaffolding is identified as temporary teacher support provided only when necessary. During the scaffolding phase, the teacher and their students *"move from a position of shared responsibility to one in which the student takes ownership" (Meyer and Turner, 2007, pg. 236).*

To provide judicious scaffolding, it is essential that teachers are well aware of and responsive to the children's thinking (McCosker and Diezmann, 2009). Therefore, they must constantly ask themselves in each case how to provide just the right amount of support and which scaffolding form is most appropriate (Belland, 2017). Scaffolding can be useful for all children, both for children who need support in learning to avoid failure and for higher performing children who are trying to master difficult material (Hogan and Pressley, 1997).

Arthur Bakker et al. look at the differences between scaffolding and dialogic teaching in mathematics education; for example, dialogic teaching does not require *"withdrawal of support" (Bakker et al., 2015, pg. 1048).* Combining both might be a good balance; conversation between the children and the teacher could help to *"encourage children to think for themselves in an open-ended way such that new things can be learnt that are not known in advance" (Bakker et al., 2015, pg. 1057)*, i.e. to think somewhat more 'creatively'. Scaffolding, for its part, only describes how to teach the correct use of mathematical concepts. It is still worth noting that Arthur Bakker et al. established a table summarizing the scaffolding studies reviewed from 2010 through 2015.

According to Brian R. Belland, it is quite beneficial to pair strong computer-based scaffolding with effective *one*-to-*one* scaffolding; this *"can promote high levels of achievement among students" (Belland, 2017, pg. 44).* The GEOGEBRATAO tool can be considered a computer-based support in which *one*-to-*one* scaffolding has been integrated. Therefore, it may be seen as an example of computer-based scaffolding helping children to *"engage in and gain skill at tasks that are beyond their unassisted abilities" (Belland, 2017, pg. 26).* This computer-based scaffolding is complemented by *one*-to-*one* scaffolding through teacher intervention.

Scaffolding student learning thus enables educators to reduce the negative emotions and self-perceptions that children may experience when they get frustrated, intimidated, or discouraged by an activity that is slightly beyond their ability and do not have the assistance, direction or understanding they need to achieve it (Scaffolding, 2015).

Finally, we would like to compare and contrast scaffolding and differentiated instruction. Both approaches try to move children's learning and understanding from where it is to where it needs to be, possibly blended together in the same classroom to the point of being indistinguishable. When **differentiating** instruction, the teacher can also differentiate the content, process or learning environment; this is definitely not the case for the **provided scaffolds**. (Scaffolding, 2015)

4.2.4 GEOGEBRATAO and blended learning

During the current decade and especially now in the time of COVID-19, the growth of online learning in all grades is occurring both **remotely** via virtual schools and within the real classroom via **blended learning** methods, increasingly normal and emerging technologies (Dziuban et al., 2018; Staker and Horn, 2012).

The term ‘blended learning’ is used to describe a learning environment in which learning and teaching is made possible through the mediation of Information Communication Technologies (ICTs), i.e. in which *"the integration of technology in the learning process"* exists together *"with traditional pedagogical practices"* (Stacey and Gerbic, 2007), (Yeop et al., 2016, pg. 37).

In the GEOGEBRATAO project, children learned elementary geometric concepts through the GEOGEBRATAO tool, a kind of online delivery of content and instruction with some element of student control over pace and learning pathways customized to their needs. This fulfills the $first$ component of blended learning. For convenience reasons relating to the evaluation of our study, we determined that this would take place in the school / classroom; the intervention of a parent or other person associated primarily with the child might have led to truncated data. So, the criterion of *distance education* applying to blended learning was not respected, and consequently the concept of blended learning not used effectively. The $second$ component of the ‘blended learning’ method was somewhat *‘supervised’* by the teacher, who was allowed to intervene when absolutely necessary (cf. subsection 2.2.10, subsection 4.2.1), obviously also in the classroom (away from home as required by the ‘blended learning’ method). This component represents face to face learning. (Staker and Horn, 2012; Stacey and Gerbic, 2007)

Globally, the GEOGEBRATAO project is characterized by the **Flex model** of blended learning defined by Norm Friesen (report 2012 (Friesen, 2012)), i.e. a model under which the geometric concepts to be studied are delivered primarily by the GEOGEBRATAO tool. The teacher, who is physically present, provides traditional face-to-face support on a flexible and adaptive as-needed basis through individual tutoring. (Staker and Horn, 2012; Hwang et al., 2019)

The effectiveness of a blended learning environment / approach in general, and in primary grades for some studies, is investigated through learner outcomes. In particular, researchers try to establish a relationship between student characteristics / backgrounds, design features and outcomes. Furthermore, they examine whether student characteristics / backgrounds and design features are (significant) predictors for student learning outcomes in blended learning. Thereby, they ask themselves what *"the student characteristics and blended learning design features for an effective blended learning environment"* are (Kintu et al., 2017, pg. 8).

Other papers discuss how blended learning can be successfully implemented in schools (Yeop et al., 2016) and how to improve learning achievement in science education for elementary school children via blended learning (Hwang et al., 2019). Children’s computer skills and learning motivation are measured to understand how the blended learning environment affects their learning achievement.

The purpose of another research study is to determine the effects of traditional science instruction and blended learning on STEM achievement of elementary school children. Among other things, Steven J. Seage and Mehmet Türegün are interested in *"the differences between science, technology, engineering and mathematics achievement scores of students of low socioeconomic backgrounds who received traditional instruction and those who received blended learning instruction" (Seage and Türegün, 2020, pg. 134).*

Finally a last paper focuses on the efficacy of a **highly structured**, **adaptive**, **personalized** blended learning approach implemented in primary grades to support reading skills of instruction through **constant assessment without testing**, which has a series of interesting similarities with the GEOGEBRATAO tool. The interactive computer on-line program (Lexia Reading CORE5 (Lexia CORE5, 2017)) provides *"**immediate** corrective **feedback** and explicit instruction when students struggle with an online activity"* among others through **scaffolded support** (Kazakoff et al., 2018, pg. 433).

4.2.5 Student-Centered Learning and Classroom

According to Catlin R. Tucker, the ultimate goal of using technology to complement work achieved in class is to attach greater importance to student-centered classrooms, i.e. to shift the focus in the classroom from the teacher to the pupils; therefore, **blended learning** seems to be the **ideal tool** (Tucker, 2012).

The pupils have to construct and reconstruct knowledge in order to learn effectively; their learning is most effective when the pupil *"experiences constructing a meaningful product"* in each learning assignment. They need to take a more active role in their education and learning process to have more *"responsibility and accountability"* and to develop their ability to work autonomously and independently. The teachers for their part have to serve as *'facilitators'* who appreciate different learning styles and attract their pupils' interests to effectively engage them. Teachers must rely on their pupils and vice versa. Furthermore, it is important that both teachers and their pupils collaborate adequately and take time to reflect on the teaching and learning process. These are only a few principles of student-centered learning highlighted by Catlin R. Tucker through which pupils *"enjoy more freedom and control over the direction of their learning"*. (Tucker, 2012, pg. 6)

Kim Huett investigates how to design student-centered blended learning environments in primary grades to allow these educational approaches and to empower the children. Her paper focuses on the extent to which

1. the core values and assumptions of student-centered learning environments and

2. the design components of student-centered learning environments (context, tools, resources, scaffolds)

are supported or limited within the observed blended learning environment. (Huett, 2018)

The book *'Theoretical Foundations of Learning Environments'* also provides some broad information about student-centered learning environments, in particular about the theoretical foundations, assumptions and design methods that often underlie these environments. In particular, it shows that significant advances have been made in utilizing *"technological capabilities"* and in designing research methodologies *"to address complexities in studying interactions among teachers, learners, technology and learning processes"*. (Land et al., 2012, pg. 20-21)

Over the GEOGEBRATAO period, we try to minimize the teacher's role and interventions while giving the children the opportunity to self-guide their learning in a dynamic adaptive ICT-enhanced environment with easy-to-use technology [5]. This allows student-centered learning to occur (Eronen and Kärnä, 2018); *"GeoGebra lends itself to being a student-centered program" (Nguyen, 2017, pg. 47).*

4.2.6 Gender Differences in Mathematics

There are significant differences between boys and girls in primary grade performance. Girls tend to perform better in languages than boys, whereas boys outperform girls in mathematics. Bart H.H. Golsteyn and Trudie Schils of the Netherlands observe that boys are better equipped with several important non-cognitive resources than girls. In contrast, girls take more advantage of their IQ than boys. Moreover, boys and girls employ their skills differently. As the researchers note, *"from the data we learned that the gender gap in math performance is mostly an 'endowment difference' in terms of non-cognitive factors"*. Social skills, instrumental skills and openness to experience are the main important factors. (Golsteyn and Schils, 2014)

Another project implemented in primary grades in Switzerland shows similar findings, i.e. gender has a significant effect on mathematics results. Boys are more likely than girls to achieve a high result and are perceived as more active and participative. Girls, in contrast, tend to conform more to school rules, but are less involved and self-confident. (Zanolla et al., 2018)

The ABC of Gender Equality in Education published by OECD provides detailed information about the differences in performance between girls and boys. It gives particular attention to underperformance among boys and lack of self-confidence among girls. (OECD, 2015)

So we have provided an overview of what has been done and what is most recently done in the field of our project. The combination of the GEOGEBRATAO educational technology tool, that largely fulfills the role of a *twin* teacher, and the learning approach described in this section demonstrates the **novelty** of our research results; it should suit pupils of all learning styles. With the advances in technology, the project's field of research is currently developing rapidly at the ICT level and the children's learning style level. In the following two chapters, our focus will be on the statistical part of the project.

[5] GeoGebra is particularly developed according to the KISS principle, i.e. *'Keep It Short and Simple'*.

Chapter 5

Field data collection

This chapter mainly describes the necessary permissions we obtained to retrieve data, the editing of the primary data sources, the project's logistics, and the organisational and classroom practice aspects extremely important for a successful launch of the GEOGEBRATAO tool in our classes. The teachers' competency aspects are also outlined, in particular the way we *'trained'* the test-class teachers.

5.1 Written permissions

For the implementation of our project in schools and as part of the evaluation of the GEOGEBRATAO tool, a number of written permissions had to be obtained: 1) from the Ministry of Education, Children and Youth, 2) from the president of the college of the directorate of fundamental education, 3) from the directors of the $four$ selected regional directorates, 4) from the respective communes or municipalities, and 5) from all parents whose children were participating.

5.2 Data collection from GEOGEBRATAO through the logs of TAO

5.2.1 Brief review, overview and section purpose

As mentioned in subsection 3.2.4, data were collected from two **different main sources**, plus two auxiliary ones: a) the logs from the GEOGEBRATAO tool, b) the pre- and post-test (the most common design used in education), c) the teacher's questionnaire, and d) field notes from the classrooms during participation in GEOGEBRATAO lessons. The last two sources are the auxiliary data sets.

The purpose of this section is to provide general information about the classroom technology and organizational structure designed for a smooth launch of the GEOGEBRATAO tool in each test class, thus validating its status as a primary data source. This also includes knowledge and competencies appropriated from teachers. Finally, we will describe how the GEOGEBRATAO data collected through the TAO logs within each exploratory learning assignment, stored as JSON data (one file per child), was combined and put into one single table ready for statistical analysis.

5.2.2 Technological aspects

Each child had access to a computer with an Internet connection with which they connected to the GEOGEBRATAO tool with a personalized username ($xxyy$, xx = class number, yy = child number) and password for the whole duration of the project.

In our study, we had no control over the schools' equipment. Establishing other conditions, e.g. the type of devices to use, was not possible due to the limited number of classes available in the selected area of our project. As a consequence, the schools were equipped quite differently in terms of technological quality and availability. To balance this technical gap, we remained available as much as possible to assist the teachers and join the classes on request during the GEOGEBRATAO lesson(s), particularly to reduce any bias that might have arisen from the use of the technology in our study. Table 5.1 presents short information about each school's equipment.

Table 5.1: Schools' equipment

classes	schools' equipment
$XA01$, $XA02$	computers in computer room with a large touch screen for each computer
$XB03$, $XB04$	laptop carriage
$XC05$, $XC06$	iPads, poor Internet connection
$XD07$, $XD08$	computers in computer room
$XE09$, $XE10$	laptops (MacBook), occasionally poor Internet connection in one classroom
$XF11$, $XF12$	computers and laptop carriage

5.2.3 Project's logistics, organisational and classroom practice aspects

5.2.3.1 Onboarding

As contact with the children and teachers was essential for proper implementation of the GEOGEBRATAO tool, we scheduled a first visit to launch our tool in each test class (first GEOGEBRATAO lesson). During this visit we introduced ourselves and explained the context of our project and its proceedings. Curious children could ask any questions based on what attracted their interest, which also sparked their classmates' interest, and enhanced their motivation for using the GEOGEBRATAO tool.

Then, the children received their password (their login was the same one as in the pre-test) and logged in for the first time. In order to preserve privacy, the entire project was **unnamed**, which we repeated several times during our visits to reassure all participants. As a consequence, only the teachers distributed the passwords and helped the children to log in.

Once logged in, the children were invited to 'snoop' around the first exploratory learning assignment consisting of a short introductory task unrelated to mathematics so they could explore the tool. The objective was to let them collectively (as a class team) discover the functions of the different buttons (e.g. display of video clips as help, 'continue' button), image overlay effects (switching languages), how to make choices in drop-down menus (device C) and how to move a point in the GeoGebra frame. Each child said everything they observed aloud. To ensure equality with respect to the use of technology in the 12 test classes, we made sure that all the useful functions of the GEOGEBRATAO tool were noticed by each class, and thus well understood by each child.

5.2.3.2 Practice of a first item on choices on Likert scales

During the GEOGEBRATAO launch lesson, in a later part of the first exploratory learning assignment, the children were invited to study an introductory item on Likert scales. They did this through choices on Likert scales (device B) with our active support (cf. subsubsection 2.2.3.2). A certain number of learning assignments use this device for measuring certitudinal (attitudinal (Likert, 1932)) scales.

As the children would have to actively explore geometric concepts on the computer and acquire new knowledge autonomously during their 500 minutes with the GEOGEBRATAO tool, it was most important that they thoroughly understood the meaning of each of the $four$ response choices in this type of assignment.

In the introductory item, the children had to click on each of the $four$ response choices at least once; the meaning of the corresponding response choice was displayed. We asked one child per class to read the displayed text aloud in German and again in French.

5.2.3.3 Verbal instructions given to the children and teacher(s) in each class

Before the children officially started the sequence of 90 exploratory learning assignments autonomously and at their own pace, we read aloud the following information and requirements to the children and their teacher(s) in each class:

1. The teacher should only be called if absolutely necessary or if the message to call the teacher was displayed.

2. The children were reassured not to worry when an *overlay* video clip was displayed, if they had to repeat an activity, or if they got a different activity than their classmate (a supplementary one). This would promote a better understanding of the concept. We explained that all the children had to go from a learning assignment 1 to a learning assignment 2, but that several children might have to make a short ‘detour’. Maybe the subject had not been treated at all in the past years, and so the child needed more support, the child was off sick, misunderstood a concept or a method, or had simply forgotten it.

3. We also mentioned that the tool uses a principle based on a certain number of right (or wrong) answers (the previously described 5-*in-a-row* principle).

5.2.3.4 Start autonomous work

Once the practice phase was finished and all the necessary information and requirements were given, the children were instructed to start working and exploring the tasks autonomously and at their own pace by using the GEOGEBRATAO tool as a kind of personal teacher in a sort of blended classroom; the physical co-presence of teachers and children was required.

5.2.3.5 Observed problem: lack of watching video clips

The software data collected through TAO (Kreis et al., 2018, 2010b) about the children's learning activities served as a means of quickly assessing each child's progress as well as whether, and for how long, the child watched the individual video clips. At the very beginning of the test period, we noticed that many children watched the clips only partially, or not at all. The video clips are silent and in slow motion to give each child enough time to watch and *'understand'* them.

In order to solve this problem as early as possible, a mail had been sent to all the teachers asking them to emphasize the utility of watching the video clips in class and to encourage the children to be patient during the clips' loading and playing process. The children had to recognize that every single object shown or displayed on the computer screen had a purpose.

5.2.4 Teachers' acquiring of expected competences

Our first visit, scheduled to launch the GEOGEBRATAO tool in each test class, should be considered an additional **training within the teacher's class**. The visit was intended to help the teacher to acquire and develop the adequate competencies for using the GEOGEBRATAO tool in their class.

We connected a sort of training based on real classroom situations (a rather practical approach), that might occur while using the GEOGEBRATAO tool in geometry lessons. Just like the children, the teachers were made aware of basic requirements during the first part of the introductory lesson. Thereafter, by observing OUR approaches helping the children pass the first exploratory learning assignments of the sequence (cf. section 1.2, subsection 2.2.10), the teachers became more aware of the competences we expected from them during the 500 test minutes. In particular, they observed how WE strove to promote learning by giving appropriate information to the children to foster their understanding without *directly* communicating answers to possible questions or difficulties.

Once the teachers felt ready, they could assist us in supporting their children during the introductory lesson. Some teachers already had experience using specific learning tools in mathematics or another subject, whereas other rarely used computers during classes. Clearly, some teachers felt ready to assist earlier than others.

The objective of this training was to make our expectations accessible and more deeply understood by all the teachers so they could support our shared goal.

During the entire testing phase, we also remained available to support or assist the teachers as much as possible with technical issues, pedagogical questions or general questions about the implementation of the GEOGEBRATAO tool. This help was actually most necessary in the *two* iPad classes, which will be further discussed in the chapter concerning the project's inconveniences and challenges (cf. section 7.4).

5.2.5 Single table containing data collected from exploratory learning assignments

As mentioned in subsection 2.2.9, the data collected within each exploratory learning assignment is a combination of the child's inputs, the result(s) and some metrics related to the assessment. It is stored on the server as JSON data; there is one JSON file for each child:

"JavaScript Object Notation (JSON) is a text format for the serialization of structured data (Crockford, 2006a). $\cdots$ Design of JSON is simple and concise in comparison with other text based formats, and it was originally proposed by Douglas Crockford as a 'fat-free alternative to XML' (Crockford, 2006b). The syntax is easy for humans to read and write, easy for machines to parse and generate and completely described in a single page at http://www.json.org. $\cdots$ For R (R Core Team, 2013), several packages that assist the user in generating, parsing and validating JSON are available through CRAN, including rjson (Couture-Beil, 2013), $\cdots$, and jsonlite $\cdots$" (Ooms, 2014, pg. 1)

We converted the 164 JSON files (for the 164 test group children) into one single csv table by using R, and implementing `toJSON` and `fromJSON` functions. Only JSON data of particular relevance for the statistical analysis are stored in the csv table. All data appear appropriate encoding, i.e. according to child and exploratory learning assignment. The relevant data are:

- the duration the child worked on an assignment,
- the number of attempts,
- whether the subsequent activity is a repeating one,
- whether the subsequent activity is an additional assignment called *'external prim'*,
- the number of wrong answers,
- the answers provided for Likert scale questions or Likert scale statements,
- the answer(s) selected from drop-down menu(s),
- the number of *short* video clip views,
- the number of *overlay* video clip views,
- the number of times the child switched to the French version of the program (i.e. mouse pointer moved over French flag),
- whether the child acted in the GeoGebra frame to accomplish a task related to device B or device C.

Let us describe now the data collection methods used through traditional paper-and-pencil-based pre- and post-test. The structure of the pre- and post-test was described in greater detail in section 2.1. We will first describe the detailed data encoding of the pre- and post-test; then, the switch from detailed data to binary data, the restriction of the post-test data (cf. subsection 3.2.5), and the comparison between the complete binary data of the post-test and the restricted ones.

5.3 Data collection through pre- and post-test

5.3.1 Detailed encoding of pre- and post-test to categorical data

The pre- and post-test was delivered to 236 children in total (164 in the test group, 72 in the control group). To evaluate their answers and encode them into categorical data, we strictly followed the set of established criteria listed below:

- 0 for completely wrong answers,
- 1 for completely right answers,
- 2 for confusions, such as confusion between lines and segments, or between x-coordinate and y-coordinate,
- 3 for imprecise drawings off by $1-2$ millimeters or $1-2$ degrees,
- 4 for right but unlabeled answers,
- 5 for completely right but incomplete answers (more than 50% correctly done),
- 7 for partly right and partly wrong answers (more than 50% correctly done),
- 9 for missing answers.

In this study, it was advantageous to deal with binary data, as we often compared the observed proportion of 'successes' in two groups, such as the observed proportion of 'successes' from the test group to those from the control group. This was done in order to verify relationships and check for any statistically significant difference between the proportions.

5.3.2 Binary data from the pre- and post-test

The categorical data of our pre- and post-test were transformed into binary data in accordance with the criteria laid down in Table 5.2: *"Binary data have been occupying a special place in the domain of data analysis." (Li, 2006, pg. 199)*, i.e. *"Binary data sets are interesting and useful for a variety reasons. They are the simplest form of data available in a computer and they can be used to represent categorical data" (Ordonez, 2003, pg. 12).*

In transforming this data, it was necessary that we consider that the children were only 10-13 years old. It would have been unfair and senseless to transform all criteria for evaluating the children's answers that was different from criterion $'1'$ into criterion $'0'$.

As outlined, some of the children from the test group were not able to complete the entire activity sequence. Accordingly, these children might demonstrate zero, or less, progress on some of the pre- and post-test items. To account for this, we created a file containing post-test_restricted data, explained in the next subsection. (see also subsection 3.2.5)

Table 5.2: Transformation of the detailed categorical data of the pre- and post-test into binary data

criterion categorical data		criterion binary data	explanation
$'0'$	$\longrightarrow$	$'0'$	evident
$'1'$	$\longrightarrow$	$'1'$	evident
$'2'$	$\longrightarrow$	$'0'$	if confusion in every item of a given exercise was noted, a comment expressing the number of children concerned as a percentage is added to the respective place(s) in our statistical analysis
$'3'$	$\longrightarrow$	$'0'$	if the given item was basic (e.g. with grid drawn)
$'3'$	$\longrightarrow$	$'1'$	if the given item was more complex (e.g. no grid drawn, axis of symmetry neither horizontal nor vertical)
$'4'$	$\longrightarrow$	$'1'$	no serious error for a child of that age, the principle was understood, the figure was clear for the data coders
$'5'$	$\longrightarrow$	$'0'$	
$'7'$	$\longrightarrow$	$'0'$	evident, since there was at least one mistake in the item
$'9'$	$\longrightarrow$	$'0'$	evident

5.3.3 Restriction of the post-test data set

To restrict the post-test binary data set, we considered only the items related to topics that the child had explored within the GEOGEBRATAO phase. The number of these items (both correct and incorrect answers) were counted per domain, which allowed us to later calculate the children's restricted scores as a percentage. All other items (non-treated matter) were encoded $'0'$.

Based on each child's last executed GEOGEBRATAO learning assignment, we were able to identify all the geometric concepts they had successfully explored by using our tool. Each user can only progress to the next concept after successfully completing the previous one. So, for the last activity performed only, we had to check if it had been successfully completed or if the child was redirected into a loop and to a repetition of prior learning assignments.

In the following subsection, we will compare the complete binary data set of the post-test with the restricted data set (test classes); more specifically, we will check if there is any significant difference between the complete binary data set and the restricted one.

5.3.4 Comparison of the complete binary data set of the post-test and the restricted data set

To compare the complete binary data set of the post-test to the restricted data, the failure rates of post-test_all were compared to those of post-test_restricted by fitting the following

Generalized Mixed-Effects Model:

$$\text{glmer\_Comp\_Post} \leftarrow glmer(\underbrace{\text{Pct\_refl}}_{\text{response variable}} \sim \underbrace{\text{Restrict * Group * Domain}}_{\text{fixed effects}} + \underbrace{(1 \mid \text{ClassNbr})}_{\text{random effect}}, \quad \text{family} = \text{Gamma(link = "log")}, \ \text{control} = \cdots, \ \text{data} = \text{datPost}\ ) \tag{5.1}$$

where

glmer_Comp_Post	: **G**eneralized **L**inear **M**ixed-**E**ffects Model with $\underline{R}$ for the **Comp**arison of the complete binary data set of the **Post**-test to the restricted one;
datPost	: post-test data set;
Pct_refl	: variable representing the failure rates in the post-test in percent (*'refl'* stands for reflection of the *success* scores) (cf. appendix E);
Restrict	: factor representing whether the post-test data are restricted or not through its levels TRUE and FALSE;
Group	: factor having the levels *treatment* (*test group*) and *control* (*control group*);
Domain	: factor having the $five$ domains (Co, DS, Fi, LS, RS [1]) plus the full post-test (*All*) as levels;
ClassNbr	: variable representing the class numbers.

The contrasts that are of interest to us indicate whether the use of restricted post-test data or non-restricted data shows a difference

1. between the test and control group;
2. between the different levels of factor *Domain*,

by using predictions of expected responses [2], i.e. of *Pct_refl*. Briefly, we will analyse how both levels of the factor *Restrict* affect the response *Pct_refl* and how the response *Pct_refl* differs between restricted and non-restricted post-test data.

Figure 5.1 shows that using restricted post-test data or non-restricted data only affects the scores of the experimental children (explained in subsection 3.2.5), in addition to the domain that is investigated. In the post-test, the test-classes' predictions of expected responses, i.e. of the failure rates, are lower for the restricted data than for the complete binary data set, except for *Domain* Co, in which there is equality. This is due to the fact that domain Co was completely explored by all the children within the GEOGEBRATAO phase.

In section 2.3, we mentioned that every control class managed to teach all the required subjects. Figure 5.1 reflects this evidence. The horizontal line plots represent equality between the restricted and non-restricted post-test data in all control-group results at all *Domain* level.

For each level of factor *Domain* (Co, DS, Fi, LS, RS, *All*), the intercept of the class lines differs across the $twelve$ test classes (and $five$ control classes), but their slope is the same within each group.

[1] Co: coordinates, DS: drawing of symmetry, Fi: shapes, LS: lines and segments, RS: recognizing of symmetry, *All*: full post-test

[2] predictions of expected responses = fitted response values:
"Predictions of expected responses, or response probabilities, are also often required. These are useful for interpreting and visualizing estimates for multilevel models using graphs." (Skrondal and Rabe-Hesketh, 2009, pg. 660)

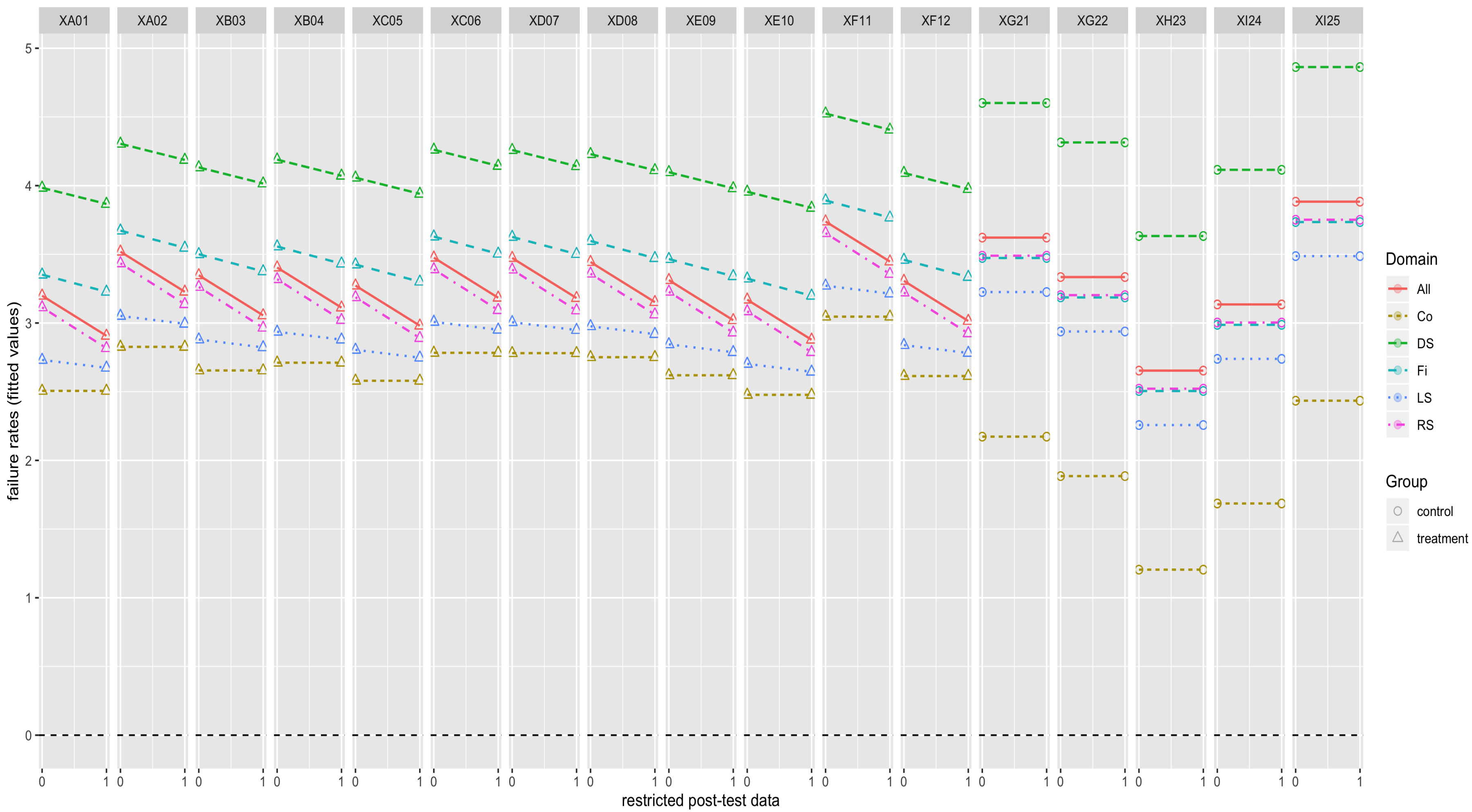


Figure 5.1: Comparison between the complete binary data set of the post-test ($Restrict = 0$) and the restricted data set ($Restrict = 1$)

In the summary of *glmer_Comp_Post* (cf. F.2 : appendix F), the base levels of our analysis are [3]:

1. FALSE for factor *Restrict*,
2. *control* for factor *Group*,
3. *All* for factor *Domain*.

The *Intercept*, the mean of the dependent variable in the $three$ base levels, has an estimate equal to $3,325$ and a standard error equal to $0,1907$.

The *directed* difference between non-restricted and restricted post-test data (failure rates) is lower for the test group than for the control group (estimate $= -0,2917$, std. error $= 0,2554$) where the *Domain* level is equal to All (its base level). We use the effect $RestrictTRUE : Grouptreatment$ (two-factor interaction) as our reference effect.

This reference effect is more or less different for:

- *Domain* Co compared to *Domain* All (estimate= 0,2917, std. error $= 0,3615$),
- *Domain* DS compared to *Domain* All (estimate= 0,1743, std. error $= 0,3616$),
- *Domain* Fi compared to *Domain* All (estimate= 0,1669, std. error $= 0,3617$),
- *Domain* LS compared to *Domain* All (estimate= 0,2347, std. error $= 0,3615$),
- *Domain* RS compared to *Domain* All (estimate= -0,0053, std. error $= 0,3617$).

The estimate of our reference effect (where *Domain* $= All$) and the estimate of our reference effect for *Domain* Co compared to *Domain* All (noted $RestrictTRUE : Grouptreatment : DomainCo$) are opposite numbers, implying that in Figure 5.1, the line plots representing the test-group results of *Domain* Co are horizontal (slope equal to $zero$).

The sum of the estimate of our reference effect (where *Domain* $= All$) and the estimate of our reference effect for *Domain* LS compared to *Domain* All [4] is equal to $-0,057$ ($\simeq 0$); domain LS was fully explored by most of the children within the GEOGEBRATAO phase. In Figure 5.1 line plots representing the test-group results of *Domain* LS are nearly horizontal.

Thus, the effects $RestrictTRUE : Grouptreatment : DomainCo$ and $RestrictTRUE : Grouptreatment : DomainLS$ are negligible; the use of restricted post-test data is not of interest for *Domain* levels Co and LS. Nevertheless, their use is of interest for *Domain* levels RS, All, Fi and DS, listed in descending order of importance. The effects $RestrictTRUE : Grouptreatment : DomainRS$ and $RestrictTRUE : Grouptreatment : DomainAll$ are stronger than the other two effects (absolute estimates difference $\approx 0,17$).

According to the Analysis of Deviance Table *(Type II Wald chisquare tests)* (cf. F.1 : appendix F), which is based on an ANOVA function doing a *Wald chisquare test*, the p-value of factor *Restrict* is equal to $0,03302$ $'*'$ [5]. This tells us that factor *Restrict* has a significant effect on *Pct_refl*, and hence a statistical significance in model *glmer_Comp_Post*. Removing factor *Restrict* from the model would harm the fit of our model.

[3]Note that the *glmer()* function *"took whatever comes first in the alphabet to be the reference level" (Winter, 2013, pg. 29).* In the case of factor *Restrict*, FALSE (0) comes before TRUE (1), so the slope represents the change from FALSE (0) to TRUE (1). That's why we refer to *directed* difference between two levels, e.g. between non-restricted and restricted post-test data.

[4]noted $RestrictTRUE : Grouptreatment : DomainLS$

[5]Bodo Winter writes, *"Unfortunately, p-values for mixed models aren't as straightforward as they are for the linear model. There are multiple approaches, and there's a discussion surrounding these, with sometimes wildly differing opinions about which approach is the best." (Winter, 2013, pg. 31)*

Random effect *ClassNbr* has much less variability (standard deviation $= 0,2315$) than *Residual*, the variability that is not due to *ClassNbr* [6] (standard dev. $= 0,8791$) (cf. F.2 : appendix F). The variation that is due to *Group* is confounded with the variation that is due to *ClassNbr*. Only one test class ($XF11$) slightly differs from all the classes (estimated deviation $= 0,3509$). In contrast, two of the $five$ control classes deviate considerably from all the classes (estimated dev. are $-0,6718$ ($XH23$), $0,5581$ ($XI25$)) and the control class $XG21$ does so slightly (estimated dev. $= 0,2964$).

In Chapter 6 '*Results of the analysis*' we provide an even more thorough analysis about our Generalized Mixed-Effects Model. This subsection primarily focussed on factor *Restrict*, its function and importance.

Concerning the data collected from the auxiliary sources (i.e. the teacher's questionnaire and field notes from the classrooms), they are less robust and so are only used to describe and contextualize outcomes.

After having described the process of information gathering from the GEOGEBRATAO tool logs and from the pre- and post-test (collecting and measuring the data), we will examine the molded data for interpretation to propose conclusions regarding our hypotheses in the next chapter. This chapter will be introduced by an item analysis and by a summary of quantitative data.

[6] *Pct_refl* is affected by some factors that are outside of the purview of our experiment (Winter, 2013).

Chapter 6

Results of the analysis

In this chapter we show to what extent the GeoGebraTAO tool fulfilled our fixed objectives. All results from the pre- and post-test and all experimental results from GeoGebraTAO's use in class are presented and examined in detail. On a number of occasions, the children from the test and control group are filtered, fully analysed and compared.

Our analysis starts with an item analysis of the pre- and post-test to examine child responses to individual items, followed by a summary of quantitative data at the class level. Thereafter we provide insights into how the children performed in the course of the GeoGebraTAO project via Generalized Mixed-Effects Models, which allow us to discuss the implications of our predicted hypotheses. The final section is illustrated by multiple effect graphs representing the results of our fitted models with different grouping levels.

6.1 Item analysis

6.1.1 Pre-test

Our pre-test data set consists of 121 items having varying difficulty, from very easy ones to more challenging ones. To measure its item interrelatedness (Eunseong and Seonghoon, 2015) and reliability, we use the Cronbach's Alpha [1], a modification of the **Kuder-Richardson Formula** 20 (Equation 3.1) (Ebel, 1967) applied to the data set which is encoded to binary data (cf. subsection 5.3.2). We obtain the following reliability analysis values, when using the `psych` package (Revelle, 2019): (Tables 6.1-6.6)

Table 6.1: List containing the main reliability analysis values for the complete pre-test

raw_alpha	std.alpha	average_r	mean	sd	median_r
0,93	0,93	0,10	0,61	0,14	0,08

One item (item Ex16*b*) was dropped from the pre-test during the reliability analysis; this item had no variance because all the children failed it (numerical value of 0 assigned to all the children). So, the Cronbach's Alpha is calculated for 120 pre-test items. In the *'Handbook of Psychological Testing'*, Paul Kline writes,

[1] Kuder-Richardson Formula 20 is equivalent to Cronbach's Alpha (1951) in the case of binary items (Cliff, 1984; Goforth, 2015): *"The elements in Cronbach's formula are identical to the elements in the K-R 20 formula except for $\sum$ variance" (Gravetter and Forzano, 2016, pg. 488).*

> *"It has been shown that the internal consistency reliability of a test must be as high as possible, $\cdots$ A reliability of $0,7$ is a minimum for a good test. This is simply because the standard error of measurement of a score increases as the reliability decreases" (Kline, 2000, pg. 15).*

From this point of view, the reliability of the complete pre-test scores is excellent, since the *raw_alpha* is equal to $0,93$. However, if α is too high *"it may suggest that some items are redundant as they are testing the same question but in a different guise" (Tavakol and Dennick, 2011, pg. 54).* The α coefficient is both a function of the covariances among items (item interrelatedness) and the number of items in the analysis. So a high α coefficient *"isn't in and of itself the mark of a 'good' or reliable set of items" (Goforth, 2015)*; *"a satisfactory level of $alpha$ can be obtained if there are a sufficient number of items" (Eunseong and Seonghoon, 2015, pg. 19)* in the set. A maximum α value of *"$0,90$ has been recommended" (Tavakol and Dennick, 2011, pg. 54).*

Since the 121 items of the pre-test cover 5 domains, *"it may not make sense to report $alpha$ for the test as a whole as the larger number of questions will inevitable inflate the value of $alpha$".* Therefore, *"$alpha$ should be calculated for each of the concepts rather than for the entire test or scale".* (Tavakol and Dennick, 2011, pg. 54)

Table 6.2: List containing the main reliability analysis values for the 'coordinates' items on the pre-test

raw_alpha	std.alpha	average_r	mean	sd	median_r
0,93	0,93	0,35	0,76	0,24	0,24

Table 6.3: List containing the main reliability analysis values for the 'lines and segments' items on the pre-test

raw_alpha	std.alpha	average_r	mean	sd	median_r
0,91	0,91	0,29	0,65	0,26	0,24

Table 6.4: List containing the main reliability analysis values for the 'recognizing symmetry' items on the pre-test

raw_alpha	std.alpha	average_r	mean	sd	median_r
0,84	0,84	0,11	0,68	0,15	0,10

Table 6.5: List containing the main reliability analysis values for the 'drawing symmetry' items on the pre-test

raw_alpha	std.alpha	average_r	mean	sd	median_r
0,74	0,73	0,13	0,23	0,15	0,11

Table 6.6: List containing the main reliability analysis values for the 'shapes' items on the pre-test

raw_alpha	std.alpha	average_r	mean	sd	median_r
0,76	0,76	0,29	0,49	0,27	0,24

For the 'coordinates' items and for the 'lines and segments' items, the *raw_alpha* is still $0,93$ and $0,91$, respectively. However, the *raw_alpha* dropped to $0,84$ (good reliability) for the 'recognizing symmetry' items and to below $0,80$ (still higher than $0,70$, acceptable reliability) for the 'drawing symmetry' items and 'shapes' items. For the set of 'shapes' items, there are only $eight$ items and *"$\cdots$ a satisfactory level of alpha can be obtained if there are a sufficient number of items" (Eunseong and Seonghoon, 2015, pg. 19).*

We analysed the pre-test reliability if an item is dropped from a domain, but we detected no great changes in its level nor more than a minute suggestion of worse reliability; therefore, there is no reason to delete any item(s) from any domain in the pre-test. In this context, we should also mention that *"researchers should be well versed in the substance of what they are studying and use that knowledge in conjunction with statistical indices to make judgments about the makeup of a measure" (Eunseong and Seonghoon, 2015, pg. 21).*

The following items do not correlate well with the overall score from the scale corresponding to their domain, as in each case, the item-total correlation [2] without the item itself, i.e. the *r.drop* value, is less than $0,30$: Ex$2b$, Ex$2f$, Ex$6a$, Ex$6b$, Ex$7a$, Ex$7b$, Ex$7c$, Ex$7d$, Ex$8h$, Ex$9d2$, Ex$9e2$, Ex$9h2$, Ex10.2., Ex$11a$, Ex$11c$, Ex$11d$, Ex$12a$, Ex$12b$, Ex$12c$, Ex$12d$, Ex$13c$, Ex$13e$, Ex$13f$, Ex$14f$, Ex$14h$, Ex$15a$, Ex$15b$, Ex$18b1$, Ex$18d1$, Ex$18d2$. Some of these odd items are extremely easy to assess because they are basic items covered during previous school years (cf. items Ex7.., Ex11..); this ease is designed to motivate lower performers. Others are difficult, advanced items (cf. items Ex$14f$, Ex$14h$, Ex15..), that should challenge higher performers and are hardly possible to answer correctly in the pre-test. It is noticeable that, among other things, parallelogram-related items do not correlate well with the overall score from the scale (cf. items Ex$9e2$, Ex$12c$, Ex$13c$).

6.1.2 Post-test

We also calculate the CRONBACH'S ALPHA to measure the item interrelatedness or reliability of our post-test data set which is encoded to binary data; the post-test is completely identical to the pre-test. From its reliability analysis, when using the `psych` package (Revelle, 2019), we obtain the following output (Tables 6.7-6.12):

Table 6.7: List containing the main reliability analysis values for the complete post-test

raw_alpha	std.alpha	average_r	mean	sd	median_r
0,93	0,93	0,10	0,70	0,13	0,08

No item was dropped from the post-test during the reliability analysis; 33 children successfully passed item Ex$16b$ in the post-test, which was dropped from the pre-test because all the children failed it. So, Cronbach's Alpha is calculated for the entire 121 items. As in the pre-test item analysis, the reliability of the complete post-test scores is excellent, demonstrated by a *raw_alpha* equal to $0,93$.

The post-test data set also includes the $five$ domains; we are analysing the reliability and, in particular, calculating a separate Cronbach's Alpha for each domain.

Table 6.8: List containing the main reliability analysis values for the 'coordinates' items on the post-test

raw_alpha	std.alpha	average_r	mean	sd	median_r
0,91	0,90	0,28	0,87	0,18	0,18

[2] In the book 'Psychometrics: An Introduction', R. Michael Furr and Verne R. Bacharach write: *"There are various ways of operationalizing an item's discrimination, one of which is the item-total correlation. We can compute the total score on a test $\cdots$ and then compute the correlation between an item with this total test score. The resulting correlation is called an item-total correlation, and it represents the degree to which differences among persons' responses to the item are consistent differences in their total test scores. A high item-total correlation indicates that the item is consistent with the test as a whole (which of course is a function of all of the items within the test), which is a desirable characteristic. In contrast, a low item-total correlation indicates that the item is inconsistent with the test as a whole, which would be an undesirable characteristic from the perspective of reliability." (Furr and Bacharach, 2013, pg. 189)*

Table 6.9: List containing the main reliability analysis values for the 'lines and segments' items on the post-test

raw_alpha	std.alpha	average_r	mean	sd	median_r
0,92	0,93	0,33	0,80	0,22	0,22

Table 6.10: List containing the main reliability analysis values for the 'recognizing symmetry' items on the post-test

raw_alpha	std.alpha	average_r	mean	sd	median_r
0,82	0,81	0,09	0,74	0,13	0,07

Table 6.11: List containing the main reliability analysis values for the 'drawing symmetry' items on the post-test

raw_alpha	std.alpha	average_r	mean	sd	median_r
0,81	0,81	0,18	0,35	0,19	0,15

Table 6.12: List containing the main reliability analysis values for the 'shapes' items on the post-test

raw_alpha	std.alpha	average_r	mean	sd	median_r
0,75	0,74	0,26	0,67	0,26	0,26

For the 'coordinates' items and 'lines and segments' items, the *raw_alpha* is still $0,91$ and $0,92$, respectively. However, the *raw_alpha* dropped to $0,82$ and $0,81$ (good reliability) for the 'recognizing symmetry' items and the 'drawing symmetry' items and even to $0,75$ (still higher than $0,70$, acceptable reliability) for the 'shapes' items. As in the pre-test, the number of 'shapes' items is not sufficient to obtain a satisfactory level of $alpha$.

Here, we also analysed the post-test reliability if an item is dropped from a domain. Again, we detected no great changes in its level nor more than a minute suggestion of worse reliability; therefore, there is no reason to delete any item(s) from any domain in the post-test.

The following items do not correlate well with the overall score from the scale corresponding to their domain, as in each case, the item-total correlation without the item itself, i.e. the *r.drop* value, is less than $0,30$: Ex$1b$, Ex$1d$, Ex$1f$, Ex$2b$, Ex$2f$, Ex$6a$, Ex$6b$, Ex$7a$, Ex$7b$, Ex$7c$, Ex$7d$, Ex$8a$, Ex$8f$, Ex$8g$, Ex$8i$, Ex$8k$, Ex$9c1$, Ex$9c2$, Ex$9d1$, Ex$9d2$, Ex$9e2$, Ex$9f2$, Ex$9h2$, Ex10.1., Ex10.2., Ex10.3., Ex$11a$, Ex$11b$, Ex$11c$, Ex$11d$, Ex$12a$, Ex$12d$, Ex$13f$, Ex$14a$, Ex$14c$, Ex$15a$, Ex$17a$, Ex$18d1$. In the next subsection, we will compare post-test proportion scores to pre-test ones for items that do not correlate well with the overall score from the scale corresponding to their domain either in the pre-test or post-test.

6.1.3 Comparison between pre- and post-test proportion scores of items having a low *r.drop* value within their domain

To compare pre- and post-test outcomes of items having a low *r.drop* value within their domain, i.e. items inconsistent with their domain as a whole in either the pre- or post-test, we calculate and use the proportion of children that got the *'correct'* answer on each of these items and call this measure an item difficulty score, which is *"a method of determining how 'easy' an item is (i.e., 'is the question answered correctly by a high proportion of individuals?') or how 'difficult' an item is (i.e., 'is the question answered correctly by a low proportion of individuals?')" (Smyth and Johnson, 2015, pg. 2).* In the pre-test, we also tried to use these scores to determine whether or not the children had possessed some prior knowledge about the subject matter assessed from previous school years.

Items that are too difficult, or too easy, are not particularly interesting for distinguishing between children within our sample. But items that were correctly answered by a **lower** proportion of children in the **pre-test** are most interesting tool with which to assess children's progress from pre-test to post-test.

Table 6.13 shows the item difficulty scores computed for items inconsistent with their domain as a whole in either the pre- or post-test.

Table 6.13: Comparisons made between item difficulty scores (success) computed for items that do not correlate well with the overall score from the scale corresponding to their domain in either the pre- or post-test (marked in red)

Items	Pre-test success	Post-test success	Difficulty level	Possibly pre-known subject matter
Ex1b	93%	97%	very low	yes
Ex1d	93%	97%	very low	yes
Ex1f	94%	97%	very low	yes
Ex2b	97%	99%	very low	yes
Ex2f	96%	100%	very low	yes
Ex6a	72%	90%	low	yes
Ex6b	27%	38%	high	?
Ex7a	98%	100%	very low	yes
Ex7b	91%	96%	very low	yes
Ex7c	90%	89%	low	yes
Ex7d	80%	85%	low	yes
Ex8a	87%	94%	low	yes
Ex8f	88%	96%	very low	yes
Ex8g	85%	94%	low	yes
Ex8h	68%	69%	medium	yes
Ex8i	89%	93%	low	yes
Ex8k	87%	96%	very low	yes
Ex9c1	83%	90%	low	yes
Ex9c2	88%	97%	very low	yes
Ex9d1	81%	91%	low	yes
Ex9d2	81%	93%	low	yes
Ex9e2	35%	46%	high	?
Ex9f2	65%	83%	low	?
Ex9h2	25%	33%	high	?

Items	Pre-test success	Post-test success	Difficulty level	Possibly pre-known subject matter
Ex10.1.	67%	88%	low	?
Ex10.2.	69%	89%	low	?
Ex10.3.	59%	83%	low	?
Ex11a	70%	89%	low	yes
Ex11b	75%	94%	low	yes
Ex11c	83%	96%	very low	yes
Ex11d	34%	28%	high	yes
Ex12a	94%	97%	very low	yes
Ex12b	35%	37%	high	yes
Ex12c	56%	56%	medium	?
Ex12d	78%	86%	low	yes
Ex13c	28%	36%	high	?
Ex13e	13%	19%	very high	no
Ex13f	33%	41%	high	?
Ex14a	66%	85%	low	yes
Ex14c	61%	74%	medium	yes
Ex14f	3%	14%	very high	no
Ex14h	1%	12%	very high	no
Ex15a	3%	5%	very high	no
Ex15b	2%	22%	high	no
Ex17a	75%	86%	low	yes
Ex18b1	36%	44%	high	no
Ex18d1	29%	34%	high	no
Ex18d2	2%	6%	very high	no

All items have been kept in the pre- and post-test results, even those that were either missed or answered correctly by a very large number of children, i.e. those with a very high difficulty level (item difficulty scores lower than 20%) or a very low difficulty level (average of pre- and post-test item difficulty scores higher than 90%).

> *"Extremely difficult or easy items will have low ability to discriminate but such items are often needed to adequately sample ⋯ objectives" (Office of Educational Assessment, n.d.).*

A **very high** difficulty level indicates that the children probably had no prior knowledge of the content in question from previous school years. A small change in item difficulty from pre-test to post-test (less than 12%) indicates that the item addresses one of the **most** advanced concepts in the test, for example Ex14h or Ex15a. In contrast, Ex15b was also likely brand-new content, but its change in item difficulty is equal to 20%; therefore, Ex15b only has a **high** difficulty level.

Items with a **very low** difficulty level were included to improve children's motivation; for these items, (with item difficulty scores over 85% in the pre-test, with one exception) the students had some good prior knowledge of the content assessed from previous school years. Even so, all these items' success rates increase in percentage from pre-test to post-test (between 2% and 10%). The exception is Ex11c, whose change in item difficulty is equal to 13% knowing that its pre-test item difficulty score was equal to 83%.

Some item difficulty scores, especially for true/false questions, may have been influenced by guessing. To ensure that the true/false questions were effective, we always included more than *one* item on the same subject matter.

In Table 6.13, there are mainly items that do not correlate well with the overall score from the scale corresponding to their domain in either the pre- or post-test (20 items in the pre-test **and** post-test); that's why items with a **medium** difficulty level are very scarce and **extremely difficult** ones are numerous.

6.2 Summary of quantitative data at the class level

6.2.1 Overview table

Table 6.14: Summary of quantitative data at the class level representing the **(success)** scores (% of correct answers) on the pre- and post-test (complete and restricted data) and the sequence lengths of correctly answered activities from the 74 main exploratory learning assignments (Tool). The *lowest* percentage in each row is written *in italics*, the **highest in bold**.

		XA01	XA02	XB03	XB04	XC05	XC06	XD07	XD08	XE09	XE10	XF11	XF12	XG21	XG22	XH23	XI24	XI25
min	Pre	23,97	35,54	42,98	47,11	33,06	29,75	38,02	46,28	35,54	43,80	38,84	*22,31*	27,27	34,71	49,59	**54,55**	26,45
	Post	45,45	*36,36*	55,37	*36,36*	57,85	42,15	54,55	52,89	58,68	47,11	*36,36*	38,02	48,76	52,07	**70,25**	56,20	44,63
	Rest	50,70	47,95	67,61	42,25	67,12	*35,42*	55,37	57,85	59,04	58,90	47,89	46,48	48,76	52,07	**70,25**	56,20	44,63
	Tool	55,41	67,57	54,05	54,05	74,32	*33,78*	**75,68**	**75,68**	37,84	56,76	68,92	55,41					
q1	Pre	51,66	51,86	47,52	53,93	53,72	45,25	48,76	54,96	54,96	56,61	*42,98*	45,45	43,39	44,63	63,64	**64,67**	54,75
	Post	64,88	55,58	58,47	65,08	72,11	65,29	58,27	63,84	67,15	65,70	*52,89*	66,12	59,50	61,98	**76,86**	72,94	57,85
	Rest	78,93	64,86	72,73	**79,94**	76,66	78,39	69,84	65,36	76,26	72,41	*57,53*	66,12	59,50	61,98	76,86	72,94	57,85
	Tool	75,68	77,37	75,34	*62,16*	77,70	71,96	78,38	**81,08**	78,38	77,70	78,38	79,06					
med	Pre	57,85	65,29	51,66	61,57	59,09	51,65	58,68	57,44	62,40	66,94	*47,93*	61,16	50,42	56,20	69,42	**71,48**	57,85
	Post	71,90	75,20	70,25	75,62	74,79	69,84	69,42	69,42	68,60	73,55	*57,02*	79,34	63,23	68,60	**85,95**	79,34	59,91
	Rest	83,72	78,10	77,72	84,51	79,84	82,50	75,34	76,19	78,30	78,51	63,64	79,80	63,23	68,60	**85,95**	79,34	*59,91*
	Tool	78,38	81,75	79,73	*75,67*	91,22	76,36	79,73	**100,00**	90,54	81,08	79,73	**100,00**					
mean	Pre	57,85	63,04	53,31	62,60	58,54	52,74	57,25	60,86	60,74	67,22	50,23	55,59	*50,21*	55,37	69,59	**72,22**	57,56
	Post	70,66	68,83	69,35	72,45	74,72	67,67	67,09	72,49	71,96	74,32	*61,80*	73,39	65,70	70,38	**82,48**	76,81	62,28
	Rest	**82,62**	73,62	79,18	80,58	80,29	79,69	75,87	76,21	78,17	81,08	67,58	77,04	65,70	70,38	82,48	76,81	*62,28*
	Tool	80,49	86,56	82,55	73,87	88,85	*73,14*	87,96	**91,12**	85,72	85,68	82,73	90,00					
q3	Pre	68,80	74,17	56,20	71,48	66,73	60,95	62,81	63,84	69,42	78,10	57,85	64,88	*55,58*	63,64	78,10	**82,23**	63,23
	Post	77,69	80,17	72,94	83,26	78,10	75,21	*71,90*	83,68	77,48	84,29	77,69	81,81	74,17	72,73	**86,78**	84,30	72,11
	Rest	90,86	82,59	81,04	87,44	85,27	86,98	83,88	86,00	84,01	**91,48**	81,13	86,78	74,17	72,73	86,78	84,30	*72,11*
	Tool	*82,43*	**100,00**	**100,00**	86,49	**100,00**	84,12	**100,00**	**100,00**	**100,00**	**100,00**	86,49	**100,00**					
max	Pre	75,21	84,30	79,34	79,34	77,69	66,94	79,34	85,12	81,82	**90,08**	*66,12*	79,34	78,51	78,51	83,47	85,12	77,69
	Post	81,82	89,26	95,87	89,26	90,08	81,82	85,95	93,39	85,95	**96,69**	85,95	92,56	81,82	89,26	89,26	88,43	*80,99*
	Rest	95,77	94,52	95,89	94,52	90,08	91,55	91,78	93,39	89,04	**100,00**	88,51	93,51	81,82	89,26	89,26	88,43	*80,99*
	Tool	**100,00**	**100,00**	**100,00**	**100,00**	**100,00**	*90,54*	**100,00**	**100,00**	**100,00**	**100,00**	**100,00**	**100,00**					

6.2.2 Summary of the pre-test data set

In the test classes, the minimum scores in the pre-test lie between $22,31\%$ (Class $XF12$, ISCED 1.6) and $47,11\%$ (Class $XB04$, ISCED 1.6). In the control classes, the minimum scores fall between $26,45\%$ (Class $XI25$, ISCED 1.5) and $54,55\%$ (Class $XI24$, ISCED 1.6). When we refer to the scores in this section, we mean the 'success' scores.

Regarding the maximum scores in the pre-test, they lie between $66,12\%$ (Class $XF11$, ISCED 1.5) and $90,08\%$ (Class $XE10$, ISCED 1.5) for the test classes, and between $77,69\%$ (Class $XI25$, ISCED 1.5) and $85,12\%$ (Class $XI24$, ISCED 1.6) for the control classes.

Table 6.15 shows the number of classes whose mean pre-test score falls in each *ten* percent interval, same for each quartile score.

Table 6.15: Number of test classes (T) out of 12 and control classes (C) out of 5 whose mean pre-test score falls in each ten percent interval, same for each quartile score

	[40%; 50%[		[50%; 60%[		[60%; 70%[		[70%; 80%[		[80%; 90%[	
1^{st} quartile	5 T,	2 C	7 T,	1 C	0 T,	2 C	0 T,	0 C	0 T,	0 C
median	1 T,	0 C	6 T,	3 C	5 T,	1 C	0 T,	1 C	0 T,	0 C
3^{rd} quartile	0 T,	0 C	2 T,	1 C	7 T,	2 C	3 T,	1 C	0 T,	1 C
mean	0 T,	0 C	7 T,	3 C	5 T,	1 C	0 T,	1 C	0 T,	0 C

The strongest test classes (3^{rd} quartile: $> 70\%$, mean $\in [60\%; 70\%[$) are $XE10$ (ISCED 1.5), $XA02$ (ISCED 1.5) and $XB04$ (ISCED 1.6). The weakest test class (median: $< 50\%$) is $XF11$ (ISCED 1.5).

The strongest control classes (3^{rd} quartile: $> 70\%$, mean $\in [65\%; 75\%[$) are $XI24$ (ISCED 1.6) and $XH23$ (ISCED 1.5).

Thus, both ISCED 1.5 and ISCED 1.6 classes belong to the strongest classes in both the test and control group *(research question $RQ1(b)$ (pg. 6))*.

6.2.3 Summary of the post-test_all data set

For the test classes, the minimum scores in the post-test_all range between $36,36\%$ (Classes $XA02$ (ISCED 1.5), $XB04$ (ISCED 1.6), $XF11$ (ISCED 1.5)) and $58,68\%$ (Class $XE09$, ISCED 1.5). For the control classes, the minimum scores lie between $44,63\%$ (Class $XI25$, ISCED 1.5) and $70,25\%$ (Class $XH23$, ISCED 1.5).

Regarding the maximum scores in the post-test_all, they are between $81,82\%$ (Classes $XA01$ (ISCED 1.5), $XC06$ (ISCED 1.5)) and $96,69\%$ (Class $XE10$, ISCED 1.5) for the test classes, and between $80,99\%$ (Class $XI25$, ISCED 1.5) and $89,26\%$ (Classes $XG22$ (ISCED 1.5), $XH23$ (ISCED 1.5)) for the control classes.

Table 6.16 shows the number of classes whose mean post-test_all score falls in each ten percent interval, same for each quartile score.

Table 6.16: Number of test classes (T) out of 12 and control classes (C) out of 5 whose mean post-test_all score falls in each ten percent interval, same for each quartile score

	[50%; 60%[		[60%; 70%[		[70%; 80%[		[80%; 90%[	
1^{st} quartile	4 T,	2 C	7 T,	1 C	1 T,	2 C	0 T,	0 C
median	1 T,	1 C	4 T,	2 C	7 T,	1 C	0 T,	1 C
3^{rd} quartile	0 T,	0 C	0 T,	0 C	7 T,	3 C	5 T,	2 C
mean	0 T,	0 C	5 T,	2 C	7 T,	2 C	0 T,	1 C

The strongest test classes (3^{rd} quartile: $> 80\%$, mean $\in [70\%; 80\%[$) are $XE10$ (ISCED 1.5), $XD08$ (ISCED 1.5), $XB04$ (ISCED 1.6) and $XF12$ (ISCED 1.6).

The strongest control classes (3^{rd} quartile: $> 80\%$, mean $\in [75\%; 85\%[$) are $XH23$ (ISCED 1.5) and $XI24$ (ISCED 1.6).

As in the pre-test, both grades (ISCED 1.5 and ISCED 1.6) appear in the strongest class lists of post-test_all *(research question $RQ1(b)$ (pg. 6))*. Broadly speaking, the data summaries of the pre-test data set and post-test_all data set show an improvement of at least 10% in the children's scores for both the test and control group.

6.2.4 Summary of the post-test_restricted data set

As already mentioned, the teachers from the control classes managed to teach all the required subjects during the specified time. Hence, the post-test_restricted data are different from the post-test_all data only for the test classes.

The test classes' minimum scores from the post-test_restricted data set vary from $35,42\%$ (Class $XC06$, ISCED 1.5) to $67,61\%$ (Class $XB03$, ISCED 1.5); their maximum scores range from $88,51\%$ (Class $XF11$, ISCED 1.5) to $100,00\%$ (Class $XE10$, ISCED 1.5). These scores are higher than those calculated for the post-test_all data set, except the smallest minimum score.

Table 6.17 shows the number of classes whose mean post-test_restricted score falls in each ten percent interval, same for each quartile score. We added the results of the control classes to enable comparison.

Table 6.17: Number of test classes (T) out of 12 and control classes (C) out of 5 whose mean post-test_restricted score falls in each ten percent interval, same for each quartile score

	[50%; 60%[		[60%; 70%[		[70%; 80%[		[80%; 90%[		[90%; 100%[	
1^{st} quartile	1 T,	2 C	4 T,	1 C	7 T,	2 C	0 T,	0 C	0 T,	0 C
median	0 T,	1 C	1 T,	2 C	8 T,	1 C	3 T,	1 C	0 T,	0 C
3^{rd} quartile	0 T,	0 C	0 T,	0 C	0 T,	3 C	10 T,	2 C	2 T,	0 C
mean	0 T,	0 C	1 T,	2 C	7 T,	2 C	4 T,	1 C	0 T,	0 C

The strongest test classes (3^{rd} quartile: $> 90\%$, mean $\in [80\%; 85\%[$) are $XE10$ (ISCED 1.5) and $XA01$ (ISCED 1.5).

By considering only the items covered with the GEOGEBRATAO tool on the post-test, we observe an improvement of about 20% from the pre-test scores. That's a difference of about 10% compared to the post-test_all data set. Furthermore, one child (class $XE10$) received 100% in post-test_restricted.

6.2.5 Summary of the GEOGEBRATAO tool data set

The GEOGEBRATAO tool provides a sequence of 90 exploratory learning assignments (activities), 74 without the additional assignments called *'external prim'* (cf. subsection 2.2.4). The minimum sequence lengths of students' completed activities (out of 74) vary from $33,78\%$ (Class $XC06$, ISCED 1.5) to $75,68\%$ (Classes $XD07$ (ISCED 1.5), $XD08$ (ISCED 1.5)), and the maximum sequence lengths of succeeded activities are 100.00% for all classes except for $XC06$ (ISCED 1.5), which has a maximum sequence length of $90,54\%$.

Table 6.18 shows the number of test classes whose mean sequence length of completed activities (out of 74) falls in each ten percent interval, same for each quartile score.

Table 6.18: Number of test classes (T) out of 12 whose mean sequence length of completed exploratory learning assignments (out of 74) falls in each ten percent interval, same for each quartile score

	[60%; 70%[	[70%; 80%[	[80%; 90%[	[90%; 100%[	100%
1^{st} quartile	1 T	10 T	1 T	0 T	0 T
median	0 T	6 T	2 T	2 T	2 T
3^{rd} quartile	0 T	0 T	4 T	0 T	8 T
mean	0 T	2 T	8 T	2 T	0 T

The strongest test classes (median: $= 100\%$, mean $\in [90\%; 92\%[$) are $XD08$ (ISCED 1.5) and $XF12$ (ISCED 1.6); at least 50% of their children finished the entire programmed geometry learning assignment sequence. Eight other classes have a 3^{rd} quartile equal to 100%, which means that at least 25% of their children got to the end of the sequence. $XD08$ (ISCED 1.5) has the best results for each quartile, the best minimum and maximum value, and the best mean.

The weakest test classes (mean $\in$ $[70\%; 75\%[$) are $XC06$ (ISCED 1.5) and $XB04$ (ISCED 1.6). $XC06$ (which used iPads) is also the only class in which no child completed the sequence (maximum sequence length equal to $90, 54\%$).

As we can see, both the strongest and weakest classes, with regard to the sequence lengths of succeeded exploratory learning assignments, consist of one ISCED 1.5 and one ISCED 1.6 class *(research question $RQ1(b)$ (pg. 6))*.

6.2.6 Observations

It is important to note that control class $XH23$ (ISCED 1.5) is strong compared to all other participating classes; all its values listed in Table 6.14 are comparatively high; in fact, most of its post-test values are bolded, indicating that they are the highest post-test scores among all classes. The lowest score out of this high-performing class was $70, 25\%$, which is a quite high score overall.

The test class $XE10$ (ISCED 1.5) has the highest maximum scores of all classes. Furthermore, its 3^{rd} quartile value of post-test_restricted is larger than that of control class $XH23$.

In the following section, we will further investigate the differences between classes with particular interest in any score differences observed between the test and control groups at the conclusion of the study *(research questions $RQ1$ and $RQ2$ (pg. 6))*.

6.3 Test of predicted hypothesis *one* related to the pre- and post-test

In this section we prove hypothesis *one* through **Generalized Mixed-Effects Models**; it postulated that, after an initial testing phase (pre-test) and after exposure to the exact same content topics in elementary geometry as in the pre-test, our control group would not evidence statistically significant higher post-test learning outcomes in comparison to our test group working with the GEOGEBRATAO tool.

6.3.1 Comparisons between the pre- and post-test calculated at the children level

6.3.1.1 Fitted Generalized Mixed-Effects Model

To compare the failure rates of the $three$ data sets - 0) the pre-test binary data set, 1) the complete binary data set of the post-test and 2) the restricted data set - we used the following **Generalized Mixed-Effects Model**:

$$\text{glmer\_Comp\_PrePost} \leftarrow glmer(\underbrace{\text{Pct\_refl}}_{\text{response variable}} \sim \underbrace{\text{Test\_3 * Group * Domain}}_{\text{fixed effects}} + \underbrace{(1 \,|\, \text{ClassNbr/ChildInClass})}_{\text{random effects}}, \\ \text{family} = \text{Gamma(link = "log")},\ \text{control} = \cdots,\ \text{data} = \text{datPrePost}\ ) \tag{6.1}$$

where

glmer_Comp_PrePost	: **G**eneralized **L**inear **M**ixed-**E**ffects Model with R for the **Comp**arison of the **Pre**-test binary data set to the complete binary data set of the **Post**-test and to the restricted one;
datPrePost	: pre- and post-test data set, including restricted data;
Pct_refl	: variable representing the failure rates in the tests in percent (*'refl'* stands for reflection of the *success* scores) (similar to explanation given in appendix E);
Test_3	: factor representing the data set in question through its levels *Pre* (*pre-test*), *Post* (*post-test*) and *Rest* (*restricted data of the post-test*);
Group	: factor having the levels *treatment* (*test group*) and *control* (*control group*);
Domain	: factor having the $five$ domains (Co, DS, Fi, LS, RS [3]) plus the full pre- and post-test (All) as levels;
ClassNbr	: variable representing the class numbers;
ChildInClass	: factor representing the child numbers of each class (children nested within separate classes).

This model is similar to Model 5.1, except that we compare $three$ data sets, not just the restricted and non-restricted post-test data. In addition, this model is calculated at the student level, considering that the children are nested within separate classes.

The contrasts that are of interest to us correspond to whether there is a difference in the assessment of a child's progress from pre-test to post-test

1. between the test and control group;

2. between the different levels of factor *Domain*,

[3] Co: coordinates, DS: drawing of symmetry, Fi: shapes, LS: lines and segments, RS: recognizing of symmetry, All: full post-test

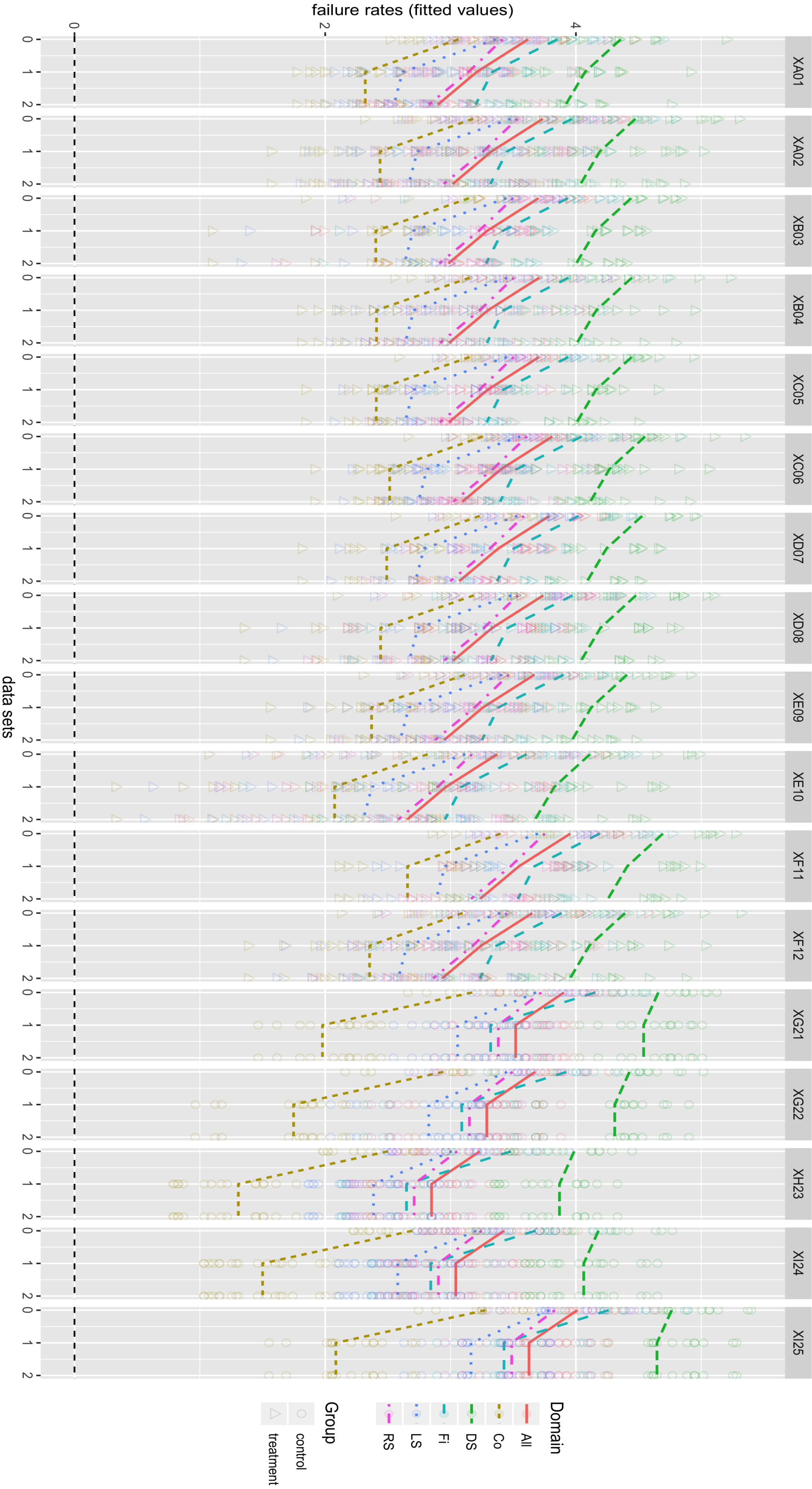


Figure 6.1: Classroom-based comparisons between the *three* data sets: the pre-test binary data set ($DataSet = 0$), the complete binary post-test data set ($DataSet = 1$) and the restricted binary post-test data set ($DataSet = 2$)

by using predictions of expected responses [4], i.e. of *Pct_refl*. Briefly, we will investigate the effect of factor *Test_3* at the fitted *Pct_refl* values between the *Domain* levels. We will analyse how much the response variable *Pct_refl* changes through the $three$ levels of factor *Test_3*, i.e. from pre-test to post-test, for each *Domain* level. Furthermore, we will study the effect of factor *Domain* at the fitted *Pct_refl* values between the levels of factor *Test_3*.

6.3.1.2 Related figure comments and interpretation

Figure 6.1 obviously shows a progression from pre-testing ($DataSet = 0$) to complete post-testing ($DataSet = 1$) for both groups and at all *Domain* levels; the classes' predictions of expected responses, i.e. of the failure rates, are lower for the complete data set of the post-test than for the pre-test data set. The test-group failure rates of *Domain* levels RS, All, Fi and DS show even greater decreases in the restricted post-test data set ($DataSet = 2$) than they do in the non-restricted data ($DataSet = 1$). This has been clearly explained in subsection 5.3.4.

For each level of factor *Domain* (Co, DS, Fi, LS, RS, All), the intercept of the class broken-lines differs across the $twelve$ test classes (and $five$ control classes), but their profile is the same within each group. The broken-lines clearly represent the change from using the complete post-test binary data set to using the restricted one.

Comparing the *Domain* levels Co and DS, most and least thoroughly addressed subject matter, we see that their difference (in absolute value) in failure rates is higher in the control group than the test group. Furthermore, the failure rates themselves at all *Domain* levels vary more in the control group than the test group. This may result from the fact that the control classes obviously had different teachers who might teach a certain topic more or less effectively to their children; the test classes, in comparison, had all the same *twin* teacher: the GEOGEBRATAO tool.

6.3.1.3 Essential information provided by the Deviance Table

The results of the Analysis of Deviance Table *(Type II Wald chisquare tests)* (cf. G.1 : appendix G) suggest a highly significant effect on *Pct_refl* ('***')

- of the factors: *Test_3* (p-value $< 2{,}2e^{-16}$), *Domain* (p-value $< 2{,}2e^{-16}$);
- of the two-factor interactions:
  '*Test_3* by *Domain*' (p-value $= 1{,}43e-09$), '*Group* by *Domain*' (p-value $= 3{,}806e-12$).

Furthermore, there appear to be significant interactions on *Pct_refl* ('*') :

- the two-factor interaction: '*Test_3* by *Group*' (p-value $= 4{,}692e-02$);
- the $three$-factor interaction: latter two-factor interaction by *Domain* (p-value $= 1{,}668e-02$).

So, the factor *Group* only seems to have a significant effect on *Pct_refl* in interaction with one or both remaining factors. We will further describe the two-factor interactions '*Test_3* by *Group*' and '*Group* by *Domain*', and the $three$-factor interaction in the next subsubsection.

[4] predictions of expected responses = fitted response values

6.3.1.4 Fitted model summary statistics and interpretation

In the summary of *glmer_Comp_PrePost* (cf. G.2 : appendix G) we see that the base levels of our analysis are:

1. *Pre* for factor *Test_3*,
2. *control* for factor *Group*,
3. *All* for factor *Domain*.

The *Intercept*, the mean of the dependent variable in the $three$ base levels, has an estimate equal to $3{,}6497$ and a standard error equal to $0{,}1476$.

In the pre-test, the *directed* difference between the control group and test group has an estimate equal to $0{,}0556$ and a standard error equal to $0{,}1764$, estimated with respect to the base level *Domain* All. This means that the test group has slightly higher failure rates on the full pre-test than the control group. By separately considering the $five$ domains (Co, DS, Fi, LS, RS), only the *Domain* level Co shows a higher *directed* difference between the control group and the test group in contrast to the base level *Domain* All (estimate $= 0{,}1765$; std. error $= 0{,}1644$). For the other four domains, a smaller difference exists between both groups (where *Domain* $= All$). In brief, there are no significant differences between the pre-test scores of the test group and those of the control group.

The *directed* difference between pre-test and post-test data (failure rates) is lower for the test group than for the control group. Considering

- the non-restricted post-test data (estimate $= -0{,}0242$; std. error $= 0{,}1629$);
- the restricted post-test data (estimate $= -0{,}3347$; std. error $= 0{,}163$; $'*'$ significant).

where the *Domain* level is equal to All (its base level). We use the effects $Test\_3Post : Grouptreatment$ and $Test\_3Rest : Grouptreatment$ (two-factor interactions) as our reference effects.

For each of these two reference effects and for each *Domain* level, the sum of the estimate of our reference effect (where *Domain* $= All$) and the estimate of our reference effect for the respective *Domain* level compared to the base level *Domain* All is equal to:

Table 6.19: For each of the two reference effects and for each *Domain* level, the sum of the estimate of our reference effect at the base level *Domain* All and the estimate of our reference effect for the respective *Domain* level compared to *Domain* All

reference effect of	**··· compared to *Domain* All**	**sig.**	**sum of estimates**	**in favour of**
Post	*Domain* Co	$'*'$	$-0{,}0242 + 0{,}4753 = 0{,}4511$	control group (since $sum > 0$)
Rest	*Domain* Co	$'***'$	$-0{,}3347 + 0{,}7858 = 0{,}4511$	control group (since $sum > 0$)
Post	*Domain* DS		$-0{,}0242 + (-0{,}1427) = -0{,}1669$	test group (since $sum < 0$)
Rest	*Domain* DS		$-0{,}3347 + 0{,}0139 = -0{,}3208$	test group (since $sum < 0$)
Post	*Domain* Fi		$-0{,}0242 + 0{,}3389 = 0{,}3147$	control group (since $sum > 0$)
Rest	*Domain* Fi	$'*'$	$-0{,}3347 + 0{,}5171 = 0{,}1824$	control group (since $sum > 0$)
Post	*Domain* LS		$-0{,}0242 + (-0{,}0947) = -0{,}1189$	test group (since $sum < 0$)
Rest	*Domain* LS		$-0{,}3347 + 0{,}1469 = -0{,}1878$	test group (since $sum < 0$)
Post	*Domain* RS		$-0{,}0242 + 0{,}0945 = 0{,}0703$	control group (since $sum > 0$)
Rest	*Domain* RS		$-0{,}3347 + 0{,}0886 = -0{,}2461$	test group (since $sum < 0$)

Thus, the *Domain* Co and Fi are in favour of the control group, while the *Domain* DS, LS and RS are more in favour of the test group.

However, we should reflect that on EX.2, a basic gap fill exercise concerning coordinates,

- in the pre-test, 26,22% of the test group and 19,44% of the control group confused x-coordinates with y-coordinates in every item of the exercise (no other error(s));
- in the post-test, 16,46% of the test group and 2,78% of the control group who still confused x-coordinates with y-coordinates in every item of the exercise (no other error(s)),

which partly explains the significantly better outcomes of the control group compared to those of the test group. In EX.3 (drawing points on a grid), we noted that confusions between x- and y-coordinates had also occurred, along with other types of errors, i.e.:

- in the pre-test, 14,02% of the test group and 1,39% of the control group confused x-coordinates with y-coordinates in 5 of 6 possible situations;
- in the post-test, 3,66% of the test group and 1,39% of the control group who still confused x-coordinates with y-coordinates in 5 of 6 possible situations.

This specific coordinate-related error (a routine) should not adversely affect the outcomes of our statistical analysis.

Regarding *Domain* Fi, we should keep in mind that only 8 of the 121 pre- and post-test items cover the domain shapes (figures). As already mentioned in subsection 6.1.1, *"Being more homogeneous in content, subscales may have higher reliabilities than the total scale. ⋯ It has ⋯ been recommended that subscale scores be based on a sufficient number of items to demonstrate reasonable reliabilities (Sinharay, Puhan, & Haberman,* 2010*)" (Little, 2013, pg. 728).* Therefore, *Domain* Fi is not really a conclusive indicator; it has only been inserted to check the children's retention of basic shape names: *"Memory was improved when items were presented both aurally and visually, in comparison with the pure-modality condition" (Lehnert and Zimmer, 2006, pg. 1082).* This is one possible advantage of a paper-and-pencil geometry course with traditional didactic teaching methods in comparison to the silent video clips included in the GEOGEBRATAO tool as visual explanations (cf. subsection 2.2.7).

6.3.1.5 Variability of the random effects of *glmer_Comp_PrePost*

In our model, we allow only the intercept to vary between our classes (1 | ClassNbr/ChildInClass). A side effect of the nested random effect is that we can quickly quantify the variation between classes *(ClassNbr)*, as well as children *(ChildInClass)* within classes.

G.2 of appendix G demonstrates that random effect *ClassNbr* has much less variability (standard deviation = 0,1624) than random effect 'children within the classes' (standard deviation = 0,3285). This effect in turn has much less variability than *Residual*, the amount of variation caused by some extraneous factors outside of the purview of our experiment, and therefore cannot be attributed to either the classes or the children within the classes (Winter, 2013) (standard deviation = 0,7371).

The prediction intervals of the random effect *ClassNbr* (Figure 6.2) show that the test classes have less variability than the control classes. Most of the classes are distributed symmetrically around $1,00$ (vertical line), except two of the $twelve$ test classes (XE10, XF11) and $four$ of the $five$ control classes (XG21, XH23, XI24, XI25), which deviate considerably from $1,00$.

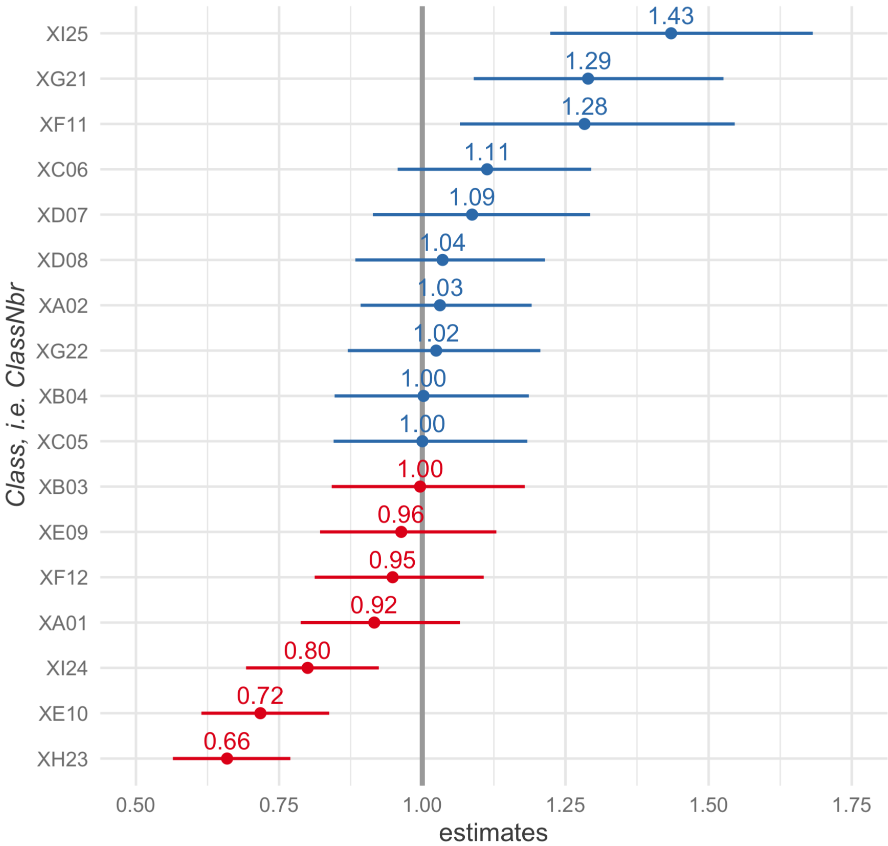


Figure 6.2: Visualization of the random effect *ClassNbr* of the fitted model *glmer_Comp_PrePost* with class estimates sorted in descending order, with highest estimate at the top; $XA - XF$: test classes, $XG - XI$: control classes

Figure 6.2 shows that:

- the control classes $XH23$ and $XI24$ and the test class $XE10$ have lower failure rates estimates,
- the control classes $XI25$ and $XG21$ and the test class $XF11$ have higher failure rates estimates

than most classes.

However, the conditional distribution of the random effect *'ClassNbr'* in a particular class, say class $XA01$, has less variability than that in class $XC06$. The conditional distribution of the random effect *'ClassNbr'* has the least variability in class $XH23$ and the most in class $XI25$.

6.3.1.6 Summary graph based on factor *Domain*

Let us fit one more **Generalized Mixed-Effects Model**, slightly different from Model 6.1, to compare the failure rates of the $three$ data sets - 0) the pre-test binary data set, 1) the complete binary post-test data set and 2) the restricted binary post-test data set. More precisely, we take into account the fact that the children are nested within separate classes **and** the classes within separate groups (test group, control group), in order to make comparisons between both groups in relation to the different levels of factor *Domain*.

$$\text{glmer\_Comp\_PrePost\_Dom} \leftarrow glmer(\underbrace{\text{Pct\_refl}}_{\text{response variable}} \sim \underbrace{\text{Test\_3 * Group * Domain}}_{\text{fixed effects}} + \underbrace{(1 \mid \text{Group/ClassNbr/ChildInClass})}_{\text{random effects}}, \quad \text{family} = \text{Gamma(link = "log")},\ \text{control} = \cdots,\ \text{data} = \text{datPrePost\_Dom}\ ) \tag{6.2}$$

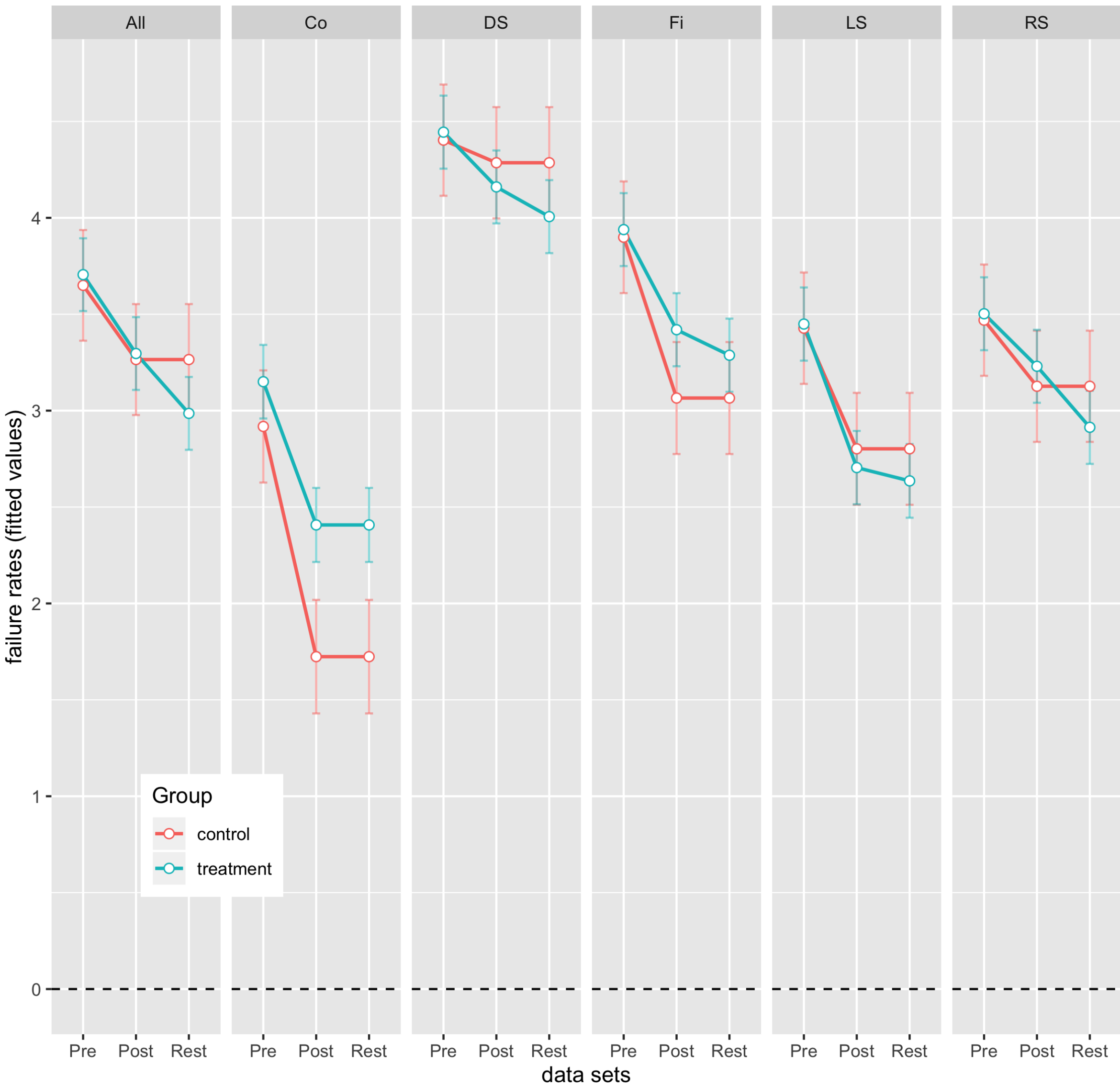


Figure 6.3: Pre- and post-test failure rates comparisons made between the test group and control group in relation to the different levels of factor *Domain*

Figure 6.3 shows conclusions that are consistent with our prior analysis.
(in answer to research questions $RQ1$ and $RQ2$ (pg. 6))

1. The test group and control group started with almost the **same basic knowledge in all the domains** except ‘coordinates’. However, the difference between the test and control groups for *Domain* level Co is not significant.

2. The **control group shows a greater decrease in** both the ‘coordinates’ and the ‘shapes’ (Fi) **failure rates** than the test group. However, we should keep in mind the comments concerning these domains, specifically the issue of confusing x- and y-coordinates and the low number of shape-related questions (comments made in subsubsection 6.3.1.4).

3. The **test group shows a greater decrease in** both the ‘drawing of symmetry’ and the ‘lines and segments’ **failure rates** than the control group.
4. By considering the restricted binary data set of the post-test, the **test group also shows a greater decrease in** the ‘recognizing of symmetry’ **failure rates** than the control group.
5. ***Domain*** Co **is obviously the easiest domain for the children and *Domain*** DS **appears to be the most challenging one.** However, we must take into account that the ‘drawing of symmetry’ items were individually much more complex than the ‘coordinates’ items. This complexity may be related to an increase in the number of objects to be drawn. For example, children had to draw several points and/or several lines to create a ‘drawing of symmetry’ item. In contrast, in a ‘coordinates’ item it sufficed to draw one point correctly on a grid based on its ‘coordinates’ to succeed in this item. As a reminder, the whole pre- and post-test is appended to this manuscript (cf. appendix O).

In the next subsection, we will include the factor *Gender* in our statistical analysis to further compare the differences in achievement between the test and control groups on the pre- and post-test, as well as any *‘gender gaps’*. Do learning outcomes vary between boys using the GEOGEBRATAO tool and boys following a traditional paper-and-pencil course only? How about for girls? Are there any significant differences between the boys’ and girls’ learning outcomes? These questions are relevant to either the acceptance or rejection of hypothesis one.

6.3.2 Comparisons between *‘gender gaps’* in the pre- and post-test *research question* $RQ1(a)$ *(pg. 6)*

6.3.2.1 Fitted Generalized Mixed-Effects Model

In this subsection, we fit a **Generalized Mixed-Effects Model** similar to Model 6.1; in this case, the fixed effect factor *Group* is replaced by the two-factor interaction ‘*Group* by *Gender*’. This is done in order to identify and analyse potential *‘gender gaps’* existing at the different levels of factor *Domain* within and between the test and control groups. Factor *Gender* obviously includes the levels B *(boy)* and G *(girl)*.

$$\text{glmer\_Comp\_PrePost\_Gen} \leftarrow glmer(\underbrace{\text{Pct\_refl}}_{\text{response variable}} \sim \underbrace{\text{Test\_3 * Domain * Group:Gender}}_{\text{fixed effects}} + \underbrace{(1\,|\,\text{ClassNbr/ChildInClass})}_{\text{random effects}}, \quad \text{family} = \text{Gamma(link = "log")},\ \text{control} = \cdots,\ \text{data} = \text{datPrePost\_Gen}\ ) \tag{6.3}$$

6.3.2.2 General comments and interpretation

According to the results of the Analysis of Deviance Table *(Type* II *Wald chisquare tests)* (cf. H.1 : appendix H), the two-factor interaction ‘*Group* by *Gender*’ only seems to have a significant effect, albeit a highly significant one, on *Pct_refl* ($'***\,'$) in interaction with the factor *Domain*, building the $three$-factor interaction $Domain : Group : Gender$.

In the summary of *glmer_Comp_PrePost_Gen* (cf. H.2 : appendix H) we see that the base levels of our fixed effects factors are:

1. *Pre* for factor *Test_3*,
2. *All* for factor *Domain*,
3. *control:B* for the two-factor interaction ‘*Group* by *Gender*’.

Going forward, the following terms are used: *‘control boys’* is used for the boys in the control group, *‘control girls’* for the girls in the control group, *‘test boys’* for the boys in the test group (i.e. treatment group), *‘test girls’* for the girls in the test group.

6.3.2.3 Interpretation related to the pre-test

In the full pre-test, i.e. with effects estimated with respect to the base level *Domain* All, the *directed* difference between the control boys and:

Table 6.20: *Directed* failure rates differences between the control boys and each of the components of the two-factor interaction '*Group* by *Gender*' on the full pre-test; i.e. effects estimated with respect to the base level *Domain All*

the $\cdots$ has	**an estimate equal to**	**and a standard error equal to**
control girls	0,0296	0,2174
test boys	0,058	0,2077
test girls	0,0759	0,2089

This means that the test girls, the test boys and the control girls have slightly higher failure rates on the full pre-test than the control boys (base level).

By separately considering the $five$ domains (Co, DS, Fi, LS, RS), the previously described differences between the *Gender* subgroups are visibly larger for *Domain* level Co than for the base level *Domain* All; this can also be observed in Figure 6.4. The *directed* difference in failure rates between *Domain* levels All and Co for each *Gender* subgroup is equal to:

Table 6.21: Failure rates *directed* difference between *Domain* levels All and Co on the pre-test for each *Gender* subgroup

***Gender* subgroup**	**difference between *Domain* levels All and Co**
control boys	$-\ 0{,}8121$
control girls	$-\ 0{,}8121 + 0{,}1793 = -\ 0{,}6328$
test boys	$-\ 0{,}8121 + 0{,}2129 = -\ 0{,}5992$
test girls	$-\ 0{,}8121 + 0{,}3025 = -\ 0{,}5096$

Each *Gender* subgroup performed better on the 'coordinates' part than on the full pre-test.

The four other domains have minor differences in failure rates between the four *Gender* subgroups in the pre-test, indicating that all groups had almost the **same initial level of knowledge**.

6.3.2.4 Interpretation related to the progress from pre- to post-test

Let us expand our comparisons further to the pre- **and** post-test. Comparing the failure rates of the pre-test to those of the post-test, their *directed* difference for *Domain* level RS and for each individual *Gender* subgroup is equal to (detailed calculations for this *Domain* level):

Table 6.22: Failure rate comparisons between the pre- and post-test at *Domain* level RS for each individual *Gender* subgroup (*directed* differences)

***Gender* subgroup**	***Post / Rest***	***directed* difference *Pre-Post*, resp. *Pre-Rest***
control boys	*Post*	$-0{,}3766 + 0{,}0983 = -0{,}2783$
	Rest	$-0{,}3766 + 0{,}0983 = -0{,}2783$
control girls	*Post*	$-0{,}3766 + 0{,}0983 + (-0{,}021) + (-0{,}1331) = -0{,}4324$ [a]
	Rest	$-0{,}3766 + 0{,}0983 + (-0{,}0211) + (-0{,}1331) = -0{,}4325$
	Post	$\lvert\Delta\rvert$ [b] $\simeq 0{,}15$
	Rest	$\lvert\Delta\rvert \simeq 0{,}15$
test boys	*Post*	$-0{,}3766 + 0{,}0983 + (-0{,}0522) + 0{,}0609 = -0{,}2696$
	Rest	$-0{,}3766 + 0{,}0983 + (-0{,}317) + 0{,}036 = -0{,}5593$
test girls	*Post*	$-0{,}3766 + 0{,}0983 + (-0{,}0118) + 0{,}0156 = -0{,}2745$
	Rest	$-0{,}3766 + 0{,}0983 + (-0{,}3715) + 0{,}0289 = -0{,}6209$
	Post	$\lvert\Delta\rvert \simeq 0$
	Rest	$\lvert\Delta\rvert \simeq 0{,}06$

[a] Sum of the estimates of the effects:
$Test\_3Post$,
$Test\_3Post : DomainRS$,
$Test\_3Post : Groupcontrol : GenderG$,
$Test\_3Post : DomainRS : Groupcontrol : GenderG$

[b] $\lvert\Delta\rvert$ is the progress difference between the girls and the boys in absolute value

We have chosen the *Domain* level RS for these detailed calculations because of a special feature distinguishing it from the other levels. It slightly favoured the control group when using the complete binary post-test data set, but favoured the test group when using the restricted data. Comparing the progress of the girls and boys from pre-test to post-test, we see that the difference (in absolute value) is larger in the control group ($\lvert\Delta\rvert \simeq 0,15\%$) than in the test group ($\lvert\Delta\rvert \simeq 0\%$, resp. $0,06\%$); this holds true for both post-test data sets. The most progress was made by the control girls according to the complete binary post-test data set, but by the test girls when using the restricted data.

Regarding the progress from pre-test to post-test at **each** level of factor *Domain*, the *directed* difference for each individual *Gender* subgroup is equal to (summary table):

Table 6.23: Failure rate comparisons between the pre- and post-test at all level of factor *Domain* for each individual *Gender* subgroup (*directed* differences)

***Gender* subgroup**	***Post / Rest***	***directed* difference *Pre-Post*, resp. *Pre-Rest***					
		All	*Co*	*DS*	*Fi*	*LS*	*RS*
control boys	*Post*	– 0,3766	– 1,4114	– 0,163	– 0,7462	– 0,5878	– 0,2783
	Rest	– 0,3766	– 1,4115	– 0,163	– 0,7462	– 0,5878	– 0,2783
control girls	*Post*	– 0,3976	– 0,9944	– 0,0521	– 0,9778	– 0,6944	– 0,4324
	Rest	– 0,3977	– 0,9946	– 0,0521	– 0,9779	– 0,6945	– 0,4325
	Post	$\|\Delta\|$ [a] $\simeq$ 0,02	$\|\Delta\| \simeq$ 0,42	$\|\Delta\| \simeq$ 0,11	$\|\Delta\| \simeq$ 0,23	$\|\Delta\| \simeq$ 0,11	$\|\Delta\| \simeq$ 0,15
	Rest	$\|\Delta\| \simeq$ 0,02	$\|\Delta\| \simeq$ 0,42	$\|\Delta\| \simeq$ 0,11	$\|\Delta\| \simeq$ 0,23	$\|\Delta\| \simeq$ 0,11	$\|\Delta\| \simeq$ 0,15
test boys	*Post*	– 0,4288	– 0,7037	– 0,2915	– 0,4936	– 0,9469	– 0,2696
	Rest	– 0,6936	– 0,7037	– 0,4382	– 0,6199	– 1,0339	– 0,5593
test girls	*Post*	– 0,3884	– 0,7846	– 0,2775	– 0,5503	– 0,5508	– 0,2745
	Rest	– 0,7481	– 0,7846	– 0,4413	– 0,6898	– 0,6032	– 0,6209
	Post	$\|\Delta\| \simeq$ 0,04	$\|\Delta\| \simeq$ 0,08	$\|\Delta\| \simeq$ 0,01	$\|\Delta\| \simeq$ 0,06	$\|\Delta\| \simeq$ 0,4	$\|\Delta\| \simeq$ 0
	Rest	$\|\Delta\| \simeq$ 0,05	$\|\Delta\| \simeq$ 0,08	$\|\Delta\| \simeq$ 0	$\|\Delta\| \simeq$ 0,07	$\|\Delta\| \simeq$ 0,43	$\|\Delta\| \simeq$ 0,06

[a] $|\Delta|$ is the progress difference between the girls and the boys in absolute value

In analysing the data of Table 6.23 and looking at Figure 6.4, we maintain the following:

1. Comparing the progress of the girls and boys from pre-test to post-test, we see that the difference (in absolute value) (cf. $|\Delta|$ values) is larger in the control group than in the test group for most *Domain* levels. To put it briefly, **the difference between girls' and boys' progress is larger in the control group than the test group for most *Domain* levels**. This is true for the *Domain* levels Co, DS, Fi and RS, when using both the complete binary post-test data set and the restricted data. In the test group, these gender differences are lower than $0,09\%$ apart from *Domain* level LS, which has a progress difference of about $0,40\%$.
2. For the **full pre- and post-test**, these **differences between girls and boys are minimal** in both the test and control group (lower than $0,06\%$). The most progress was made by the test boys according to the complete binary post-test data set, but by the test girls when using the restricted data.
3. For the *Domain* level Co, the most progress was largely made by the control boys. For the *Domain* level Fi, the control girls made the most progress. Evidently, both *Gender* subgroups in the control group seem to have different strong points.
4. For the *Domain* level LS, the test boys largely made the most progress when using both the complete binary post-test data set and the restricted data.

To summarize, there are minor differences in failure rates between the *Gender* subgroups in the pre-test, except on the topic of 'coordinates', which possesses visibly larger differences. There is some progress difference between the control girls and control boys in the different domains, but not in the full pre- and post-test. As for the test girls and test boys, there seems to be hardly any progress difference except for *Domain* level LS, in which the test boys made more significant progress.

In the next subsection, we focus on the progress differences between children who 'like' mathematics in general and those who don't. This may enable us to draw conclusions about hypothesis one.

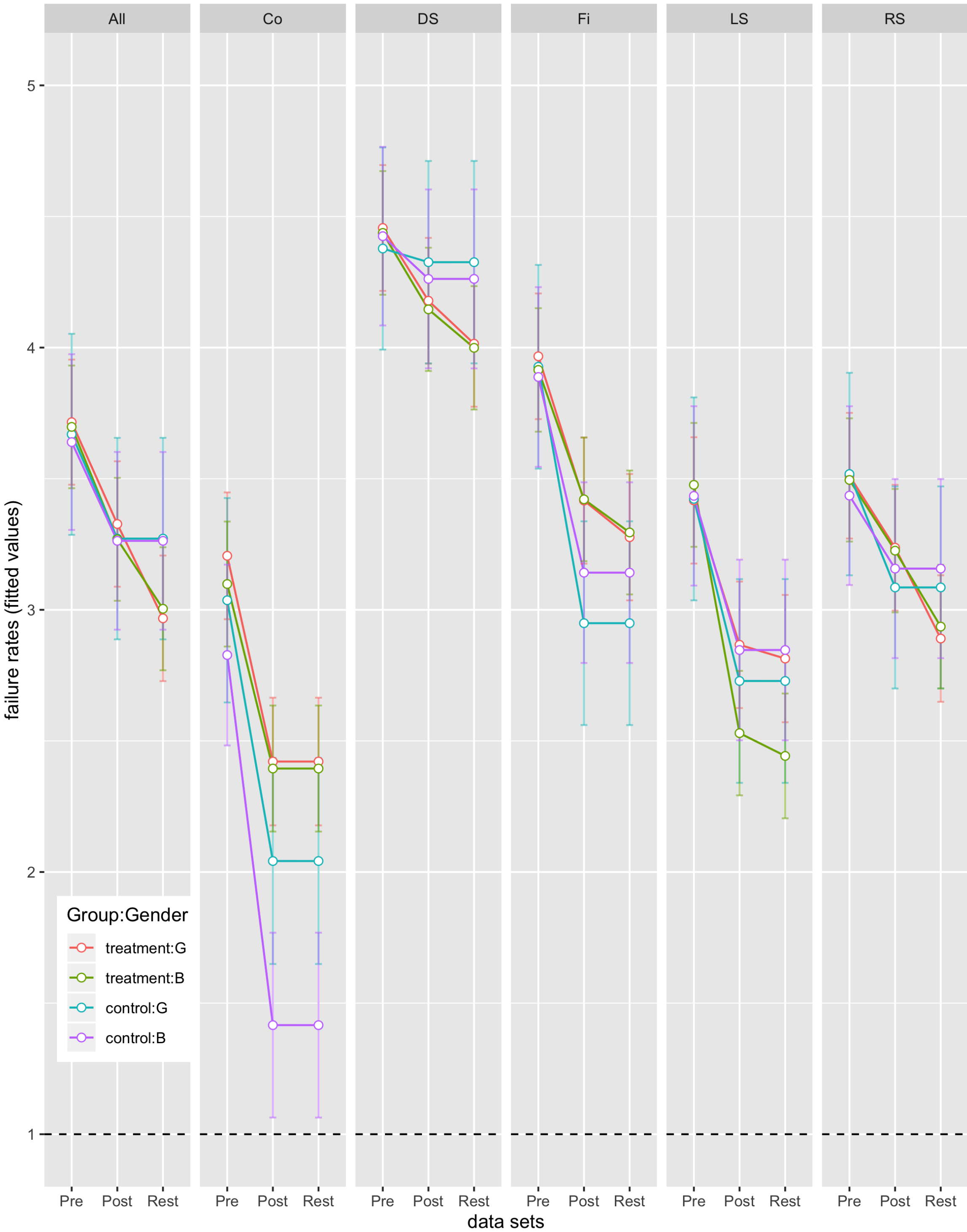


Figure 6.4: Pre- and post-test failure rate comparisons made between the test group and control group by *Gender* in relation to the different levels of factor *Domain*

6.3.3 Comparisons between children who *'like'* maths and those who don't *research question* $RQ1(c)$ *(pg. 6)*

6.3.3.1 Fitted Generalized Mixed-Effects Model

In this subsection, we proceed as in the previous one. We fit a **Generalized Mixed-Effects Model** similar to Model 6.3; the fixed effect factor *Gender* is replaced by the factor *Like*, which has the levels $LikeM$ *(loving mathematics)* and $DislikeM$ *(not really loving mathematics)*. This in order to identify and analyse potential progress differences within and between the test and control group among the children who *'like'* doing mathematics and those who do not particularly like them.

$$\text{glmer\_Comp\_PrePost\_LikeM} \leftarrow glmer(\underbrace{\text{Pct\_refl}}_{\text{response variable}} \sim \underbrace{\text{Test\_3 * Domain * Group:Like}}_{\text{fixed effects}} + \underbrace{(1 \mid \text{ClassNbr/ChildInClass})}_{\text{random effects}}, \\ \text{family} = \text{Gamma(link = "log")},\ \text{control} = \cdots,\ \text{data} = \text{datPrePost\_LikeM}\ ) \tag{6.4}$$

6.3.3.2 General comments and interpretation

According to the results of the Analysis of Deviance Table *(Type II Wald chisquare tests)* (cf. I.1 : appendix I), the two-factor interaction '*Group* by *Like*' seems to have a significant effect on *Pct_refl* ('**'). This effect even appears to be highly significant ('*** ') in interaction with factor *Domain*, building the $three$-factor interaction $Domain : Group : Like$.

In the summary of *glmer_Comp_PrePost_LikeM* (cf. I.2 : appendix I), we see that the base levels of our fixed effects factors are:

1. *Pre* for factor *Test_3*,
2. *All* for factor *Domain*,
3. *control:LikeM* for the two-factor interaction '*Group* by *Like*'.

Going forward, the following terms are used to represent:

- *'control lovers(+)'* the children in the control group who love mathematics;
- *'control lovers(−)'* the children in the control group who do not really love mathematics;
- *'test lovers(+)'* the children in the test group (i.e. treatment group) who love maths;
- *'test lovers(−)'* the children in the test group who do not really love mathematics.

6.3.3.3 Interpretation related to the pre-test

In the full pre-test, i.e. with effects estimated with respect to the base level *Domain All*, the *directed* difference between the control lovers(+) and:

Table 6.24: *Directed* failure rate differences between the control lovers(+) and each of the components of the two-factor interaction '*Group* by *Like*' in the full pre-test; i.e. effects estimated with respect to the base level *Domain All*

the ··· has	an estimate equal to	and a standard error equal to
control lovers(−)	0,1513	0,3588
test lovers(+)	0,0129	0,2252
test lovers(−)	0,2482	0,2601

This means that the test lovers(+) have approximately the same failure rates in the full pre-test as the control lovers(+) (base level). However, the test lovers(−) and the control lovers(−) have slightly higher failure rates in the full pre-test than the lovers(+) subgroups. Furthermore, the test lovers(−) performed somewhat lower than the control lovers(−).

By separately considering the $five$ domains (Co, DS, Fi, LS, RS), the previously described differences are visibly larger for *Domain* level Co than for the base level *Domain* All, similar to the observation made in the *Gender* analysis (cf. subsubsection 6.3.2.3); this can also be seen in Figure 6.5. The *directed* difference in failure rates between the *Domain* levels All and Co for each *'lovers(+/−)'* subgroup is equal to:

Table 6.25: Failure rates *directed* difference between the *Domain* levels All and Co in the pre-test for each *'lovers(+/−)'* subgroup

***'lovers(+/−)'* subgroup**	**difference between the *Domain* levels All and Co**
control lovers(+)	− 0,7637
control lovers(−)	− 0,7637 + 0,2968 = − 0,4669
test lovers(+)	− 0,7637 + 0,075 = − 0,6887
test lovers(−)	− 0,7637 + 0,5142 = − 0,2495

Seeing these outcomes, it is not surprising that each *'lovers(+/−)'* subgroup performed better on the 'coordinates' section than in the full pre-test. The difference (in absolute value) between the control lovers(+) (the highest performers at *Domain* level Co) and the test lovers(−) (the lowest performers at the same level) is equal to $\simeq 0,76\%$.

For the other four domains, the test lovers(+) and control lovers(+) demonstrate almost the same basic knowledge level in the pre-test and hence the same initial knowledge in our project. The test lovers(−) and control lovers(−), on the other hand, performed slightly lower than both lovers(+) subgroups.

6.3.3.4 Interpretation related to the progress from pre- to post-test

We now analyse the progress from the pre- to post-test at each level of factor *Domain* for each individual *'lovers(+/−)'* subgroup (cf. Table 6.26). While doing so, we keep in mind Figure 6.5 and maintain the following:

1. Comparing the progress from pre-test to post-test for both the *'lovers(+)'* and *'lovers(−)'*, we see that their difference (in absolute value) (cf. $|\Delta|$ values) is lower than $0,10\%$ at *Domain* levels All, DS, Fi and LS (with one exception hardly worth mentioning). This is true for both the test group and control group, which means that **the *'lovers(+)'* and the *'lovers(−)'*** from both the test group and control group made **the same amount of progress** at *Domain* levels All, DS, Fi and LS.

2. Regarding *Domain* level Co, this difference between the *'lovers(+)'* and the *'lovers(−)'* is larger in the test group than the control group ($|\Delta| \simeq 0,16\%$ vs $0,09\%$). The *'test lovers(−)'* made slightly more progress than the *'test lovers(+)'*.

3. At the *Domain* level RS, the difference between the *'lovers(+)'* and the *'lovers(−)'* is somewhat larger in the test and control group compared to the other *Domain* levels when using the restricted binary data set of the post-test ($|\Delta| \simeq 0,17\%$, resp. $0,15\%$). In this case, the most progress was made by the *'test lovers(+)'*.

4. **The *'lovers(+)'* and the *'lovers(−)'*** from both the test group and control group made almost **the same amount of progress in the full pre- and post-test**. However, the most progress was made by the *'test lovers(−)'* when using the restricted binary data set of the post-test.

Table 6.26: Failure rate comparisons between the pre- and post-test at all levels of factor *Domain* for each individual *'lovers(+/−)'* subgroup (*directed* differences)

'lovers(+/−)' **subgroup**	***Post / Rest***	***directed* difference *Pre-Post*, resp. *Pre-Rest***					
		All	Co	DS	Fi	LS	RS
control lovers(+)	*Post*	− 0,3769	− 1,1437	− 0,1134	− 0,8016	− 0,591	− 0,3612
	Rest	− 0,377	− 1,1438	− 0,1135	− 0,8017	− 0,5911	− 0,3613
control lovers(−)	*Post*	− 0,3413	− 1,2289	− 0,087	− 0,7302	− 0,5603	− 0,1941
	Rest	− 0,3414	− 1,2291	− 0,0872	− 0,7303	− 0,5605	− 0,1943
	Post	$\|\Delta\|$ [a] $\simeq 0{,}04$	$\|\Delta\| \simeq 0{,}09$	$\|\Delta\| \simeq 0{,}03$	$\|\Delta\| \simeq 0{,}07$	$\|\Delta\| \simeq 0{,}03$	$\|\Delta\| \simeq 0{,}17$
	Rest	$\|\Delta\| \simeq 0{,}04$	$\|\Delta\| \simeq 0{,}09$	$\|\Delta\| \simeq 0{,}03$	$\|\Delta\| \simeq 0{,}07$	$\|\Delta\| \simeq 0{,}03$	$\|\Delta\| \simeq 0{,}17$
test lovers(+)	*Post*	− 0,3891	− 0,6217	− 0,282	− 0,4922	− 0,7359	− 0,268
	Rest	− 0,6753	− 0,6219	− 0,4162	− 0,6049	− 0,8172	− 0,6246
test lovers(−)	*Post*	− 0,3862	− 0,7799	− 0,1996	− 0,4821	− 0,6728	− 0,2524
	Rest	− 0,7324	− 0,78	− 0,3516	− 0,6608	− 0,706	− 0,473
	Post	$\|\Delta\| \simeq 0$	$\|\Delta\| \simeq 0{,}16$	$\|\Delta\| \simeq 0{,}08$	$\|\Delta\| \simeq 0{,}01$	$\|\Delta\| \simeq 0{,}06$	$\|\Delta\| \simeq 0{,}02$
	Rest	$\|\Delta\| \simeq 0{,}06$	$\|\Delta\| \simeq 0{,}16$	$\|\Delta\| \simeq 0{,}06$	$\|\Delta\| \simeq 0{,}06$	$\|\Delta\| \simeq 0{,}11$	$\|\Delta\| \simeq 0{,}15$

[a] $|\Delta|$ is the progress difference between the lovers(+) and the lovers(−) in absolute value

5. For *Domain* level Co, the most progress was largely made by the *'control lovers(−)'*. For *Domain* level Fi, the *'control lovers(+)' made the most progress*.
6. For *Domain* levels DS and LS, the *'test lovers(+)'* made the most progress.

To summarize, at *Domain* levels All, DS, Fi and LS there are either no or very minor progress differences between the *'lovers(+)'* and *'lovers(−)'* subgroups, both for the test group and control group; this is also true at *Domain* level Co, but exclusively for the control group. At *Domain* levels RS (test and control group) and Co (test group), some slightly larger progress differences are visible compared to the other four *Domain* levels.

However, unlike in the *Gender* analysis, the two lovers(−) subgroups performed visibly worse **on the pre-test** than both lovers(+) subgroups; both lovers(+) subgroups despite demonstrated almost the same level of initial knowledge.

So, for most *Domain* levels, the two lovers(−) subgroups progressed more or less as much as their lovers(+) counterpart, but with somewhat less initial knowledge. The same obviously applies regarding the full pre- and post-test.

We will now proceed to the overall conclusion regarding hypothesis one.

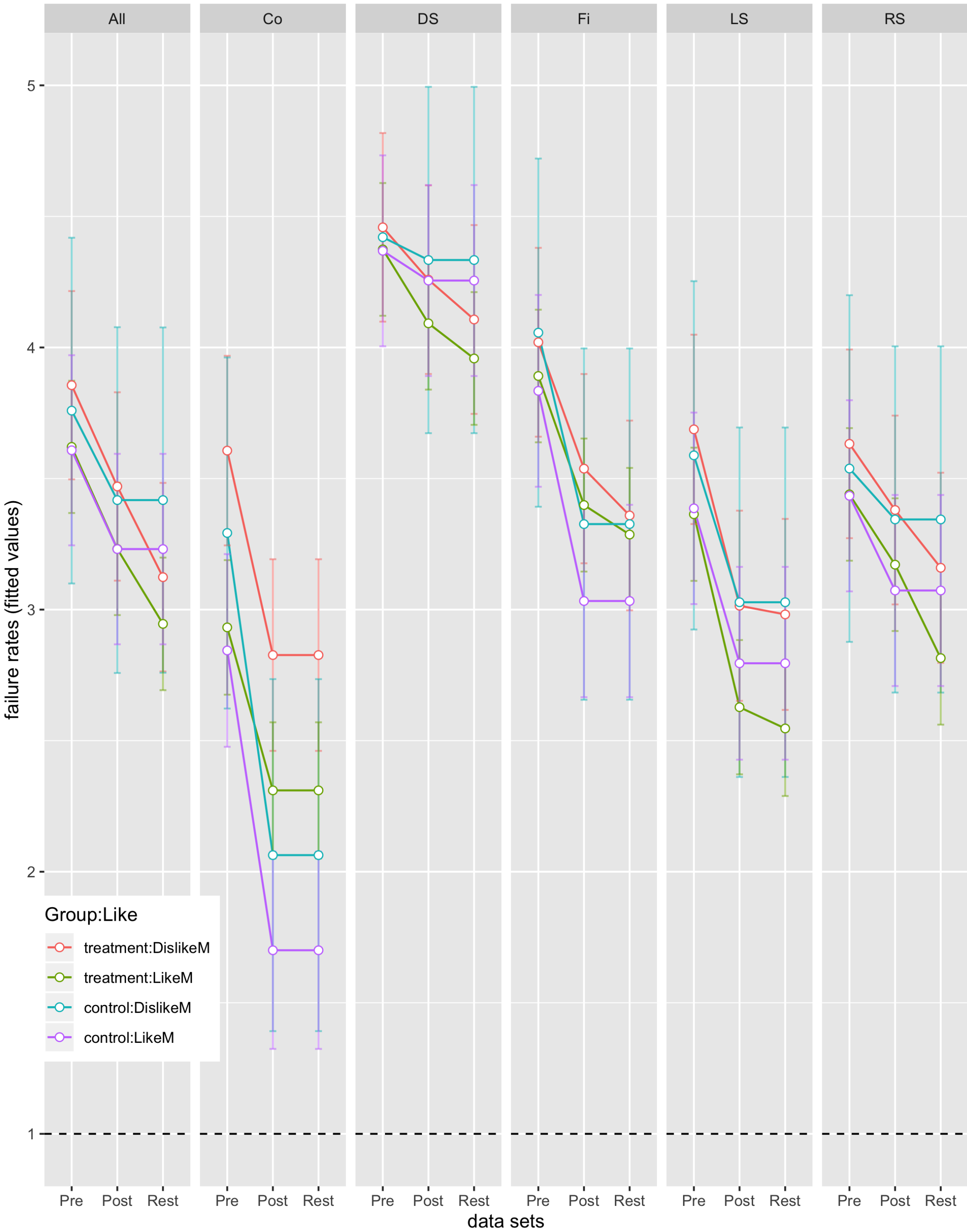


Figure 6.5: Pre- and post-test failure rate comparisons made between the test group and control group according to children's attitudes toward mathematics and in relation to the different levels of factor *Domain*

6.3.4 Conclusion regarding hypothesis *one* *(in answer to research questions $RQ1$ and $RQ2$ (pg. 6))*

Considering the **full pre- and post-test**, we have just proven that, **after an initial testing phase (pre-test) and after exposure to the exact same content topics in elementary geometry, our control group does not evidence statistically significant higher post-test learning outcomes compared with our test group working with the GEOGEBRATAO tool**.

However, at the domain level there are some content topics in favour of the control group ('coordinates', 'shapes'), whilst others favour the test group ('drawing of symmetry', 'lines and segments'). For those in favour of the control group, a more in-depth analysis reveals that, in particular, a systematic confusion error in the domain 'coordinates' was particuarly prevalent in the test group and less so in the control group. This exclusive confusion error in 'coordinate' exercises should not adversely affect the outcomes of our statistical analysis; the children concerned acquired the basic principle and main knowledge on this subject matter. Regarding the domain 'shapes', the number of items is not large enough to be a conclusive indicator; besides, its items were not related to dynamic geometry.

For the domains 'recognizing of symmetry', 'drawing of symmetry' and 'shapes', it is important to differentiate between use of the complete binary post-test data set or the restricted one. This is particularly crucial for 'recognizing of symmetry' because the test group better performed than the control group when considering only the items related to topics that the child had explored within the GEOGEBRATAO activity sequence.

The effect of using the dynamic interactive geometry tool seems to be most apparent in the outcomes of the domain 'drawing of symmetry'. This domain appears to be the most challenging for all the children, but was somewhat easier for the test group than the control group.

Considering any *'gender gaps' (research question $RQ1(a)$ (pg. 6))*, they are larger in the control group than the test group for most subject matter from the pre-test to the post-test. However, for the full pre- and post-test, *'gender gaps'* are minimal in both groups.

For the full pre- and post-test, we should also note that the most progress was made by the test boys when using the complete binary post-test data set and by the test girls when using the restricted data set. This strengthens our established hypothesis *one* even further.

Regarding children who generally 'like' doing mathematics in class and those who do not *(research question $RQ1(c)$ (pg. 6))*, we detected a fundamental difference from our previous analyses. In both the test and control group, these *two* categories began with different levels of subject matter knowledge (pre-tested). From the pre-test to the post-test, for most domains, the children who do not 'like' mathematics progressed more or less as much as those who do, but began with somewhat less initial knowledge. This result has also been shown for the full pre- and post-test.

For the full pre- and post-test, and considering only the items related to studied topics within the GEOGEBRATAO phase, the most progress was made by the children in the test group who do not 'like' mathematics; this also supports our hypothesis.

Let us now proceed to the analysis of the third *research question $RQ3$ (pg. 6)* that is related exclusively to the progress of the test-group.

6.4 Test of predicted hypothesis *two* related to the progress of the test-group

In this section we prove hypothesis *two* through **clustering methods** in combination with **Generalized Mixed-Effects Models**; it postulated that, student-centered software would allow teachers to create a blended learning environment in which children learn geometric concepts autonomously, thereby reducing teacher interventions for the children in need of (individualized learning) support.

6.4.1 Cluster analysis using the 'complete' binary post-test data set

6.4.1.1 Procedure of PAM cluster analysis

Before determining the optimal number of clusters, we test the data set for the similarity-based clusterability of nodes using 'the Hopskins Statistic'. The value of 'the Hopskins Statistic' for our data set is approximately equal to '$0,24$', significantly below '$0,5$', which indicates that the data is highly clusterable based on the similarity of 1) the children's failure rates on the pre-test, 2) those on the **complete** binary post-test, and 3) the number of main exploratory learning assignments completed by the children (all data in %).

> *"A lower value for the Hopskins Statistic is definitely an indication that the nodes do not exhibit similar values for the centrality metrics considered." (Latifi, 2019, pg. 242-243)*

To estimate the optimal number of clusters, we use the `NbClust` package. This package *"$\cdots$ provides* 30 *indices which determine the number of clusters in a data set"* and offers *"the best clustering scheme from different results to the user" (Charrad et al., 2014, pg. 31)*. According to the majority rule, the best number of clusters is two (proposed by ten indices), followed by $five$ (proposed by $seven$ indices) (cf. Table 6.27). To test our hypothesis, it seems appropriate to use the higher (optimal) number of clusters, i.e. $five$ clusters, in order to differentiate the average performers in addition to low and high performers.

Table 6.27: Optimal number of clusters less than or equal to six

optimal number of clusters	0	2	3	4	5	6
proposed by $\cdots$ indices	2	10	2	2	7	3

All the test-group children are important for our statistical analysis and for testing hypothesis two, even those who deviate greatly from the other children, called *outliers*: *"Many data mining algorithms try to minimize the influence of outliers or eliminate them all together. However, it may result in the loss of important hidden information" (Lei et al., 2012, pg. 1045)*.

For hypothesis two, we use `PAM` (Partition Around Medoids) Clustering, which is *"better $\cdots$ because of its robustness to noisy data and outliers" (Kashef and Kamel, 2008, pg. 423)*. This method searches for 5 (in the present study) representative objects or *cluster medoids* among the children of the data set. A *cluster medoid "is defined as the most centrally located"* child *"in a cluster, that is, the"* child *"in the cluster whose average dissimilarity to all other"* children *"in the same cluster is minimal" (in (Handl and Knowles, 2005, pg. 633) by (Kaufman and Rousseeuw, 1990))*. The 5 clusters are constructed by assigning each child to the nearest medoid. The goal is to find 5 representative objects (the medoids) which minimize the sum of the children's dissimilarities to their closest representative object. (Maechler et al., 2019)

In our study, we go further in visualizing the clustering results by choosing a $three$-dimensional representation. A concentration ellipsoid is drawn around each cluster (see Figures 6.7a and 6.7b). For this we use the `scatter3d` function from the `car` package (Fox and Weisberg, 2018) which uses the `rgl` package (Adler et al., 2020) to produce **interactive** $three$-dimensional plots: *"Three-dimensional images can be difficult to depict in two dimensions but the `rgl` package provides functionality that offers three-dimensional, real-time visualization of points, shapes, and surfaces, and the package allows the user to generate interactive 3D graphics to help visualize the plot" (Ekstrom, 2017, pg. 244).*

Additionally, we provide a two-dimensional representation of the clusters (see Figure 6.6) in the same colors as the $three$-dimensional one to increase interpretability. In this two-dimensional representation, the dimensionality of our data set is obviously reduced, though we **minimize information loss** by preserving as much statistical information as possible using the principal components technique (Jolliffe and Cadima, 2016). The `fviz_cluster` function from the `factoextra` package (Kassambara and Mundt, 2020) enables us to provide this elegant two-dimensional visualization of the clustering results, i.e. of the 5 performance groups nested within our test group.

6.4.1.2 Summary of the clusters obtained through PAM cluster analysis

In the following table (Table 6.28) we briefly summarize some clusters information obtained when using the PAM (Partition Around Medoids) Clustering. The table colors are consistent with those in the two- and $three$-dimensional representations of the performance groups (cf. Figures 6.6-6.7b).

Table 6.28: Table containing 1) the description name of each cluster, 2) the number of children per cluster and 3) cluster medoid information using PAM (Partition Around Medoids) Clustering

cluster name	description name performance group	number of children	cluster medoid child	Tool [a]	Pre [b]	Post [c]
*Cluster*1	Average-Performers-Strength_Pre&Post	43	$XD0711$	78, 38	38, 03	27, 28
*Cluster*2	Second-Lower-Performers-Strength_Tool	42	$XD0704$	77, 03	51, 25	42, 16
*Cluster*3	Higher-Performers	31	$XA0206$	100, 00	23, 15	14, 89
*Cluster*4	Average-Performers-Strength_Tool	35	$XE1005$	100, 00	42, 99	26, 46
*Cluster*5	Lower-Performers	13	$XE1013$	56, 76	56, 21	48, 77

[a] *Tool* represents the number of main exploratory learning assignments the child has completed (in %)
[b] = the child's failure rate on the pre-test (in %)
[c] = the child's failure rate on the **'complete'** post-test (in %)

- *Cluster*3 contains the children who generally had the **lowest failure rates** on the pre- and post-test and who came **closest to completing** our sequence of dynamic learning assignments.

- *Cluster*5 contains the children who generally had the **highest failure rates** on the pre- and post-test and who made **least progress** in the sequence of assignments offered by the GEOGEBRATAO tool.

- *Cluster*4 contains the children who progressed **almost as far as** the children in *Cluster*3 in the sequence of dynamic learning assignments, but generally had **higher failure rates** on the pre- and post-test.

- *Cluster*1 contains the children who **progressed significantly less far** in the GEOGEBRATAO tool sequence than the children in *Cluster*4, but generally had **slightly better** results on the pre- and post-test.

- *Cluster*2 contains the children who generally performed **worse** on the pre- and post-test than the children in *Clusters*3,4&1, but **better** than those in *Cluster*5. Regarding their progress in the sequence of assignments offered by our tool, they progressed **as far as** the children in *Cluster*1.

6.4.1.3 Visualizations of the performance groups

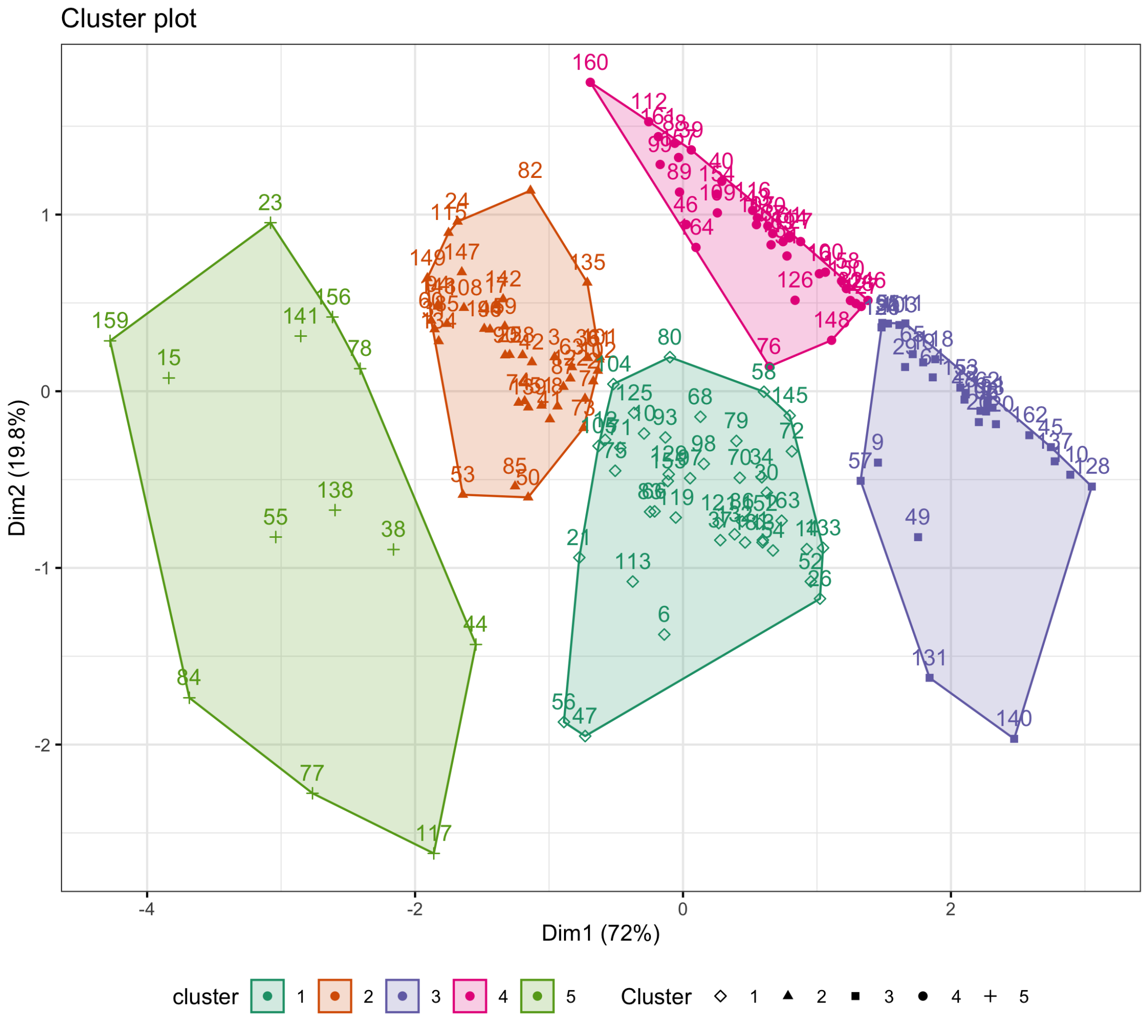
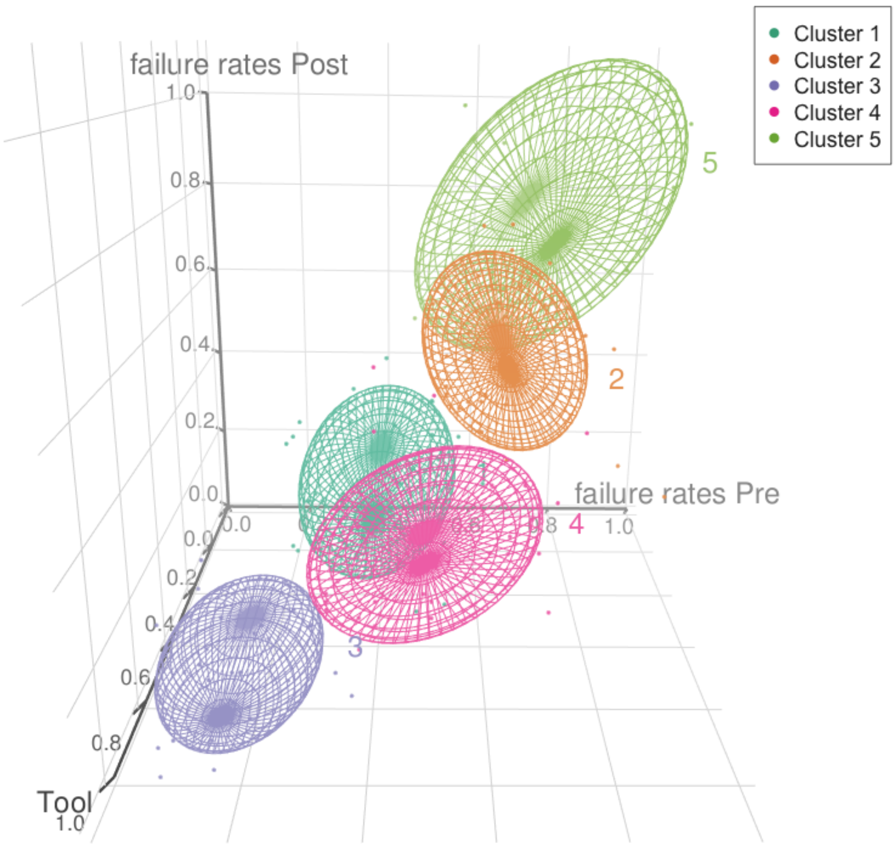
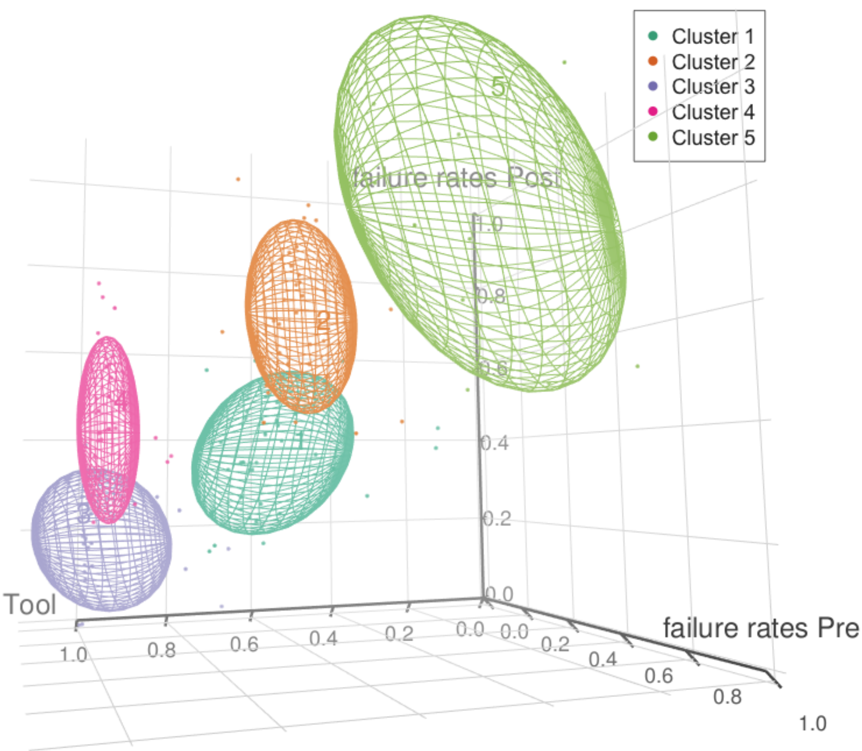


Figure 6.6: *Two*-dimensional visualization of the performance groups nested within the test group; children are represented by points in the plot, using principal components. The **complete** binary post-test data set is used.

*Cluster*1:	Average-Performers-Strength_Pre&Post
*Cluster*2:	Second-Lower-Performers-Strength_Tool
*Cluster*3:	Higher-Performers
*Cluster*4:	Average-Performers-Strength_Tool
*Cluster*5:	Lower-Performers



(View 1)



(View 2)

Figure 6.7: $Three$-dimensional visualizations of the performance groups nested within the test group; the **complete** binary post-test data set is used

6.4.2 Cluster analysis using the ‘restricted’ binary post-test data set

6.4.2.1 Procedure of `PAM` cluster analysis

It seems preferable to proceed in the same manner with the **restricted** data as for the data set containing the **complete** binary post-test data (cf. subsubsection 6.4.1.1), i.e. calculating the clustering tendency of our data set, determining the optimal number of clusters according to our needs, using `PAM` (Partition Around Medoids) Clustering to extract hidden patterns from our data set and find relationships between these patterns, and most importantly, providing clear two- and $three$-dimensional visualizations of the determined clusters.

The value of ‘the Hopskins Statistic’ calculated for the data set containing the **restricted** binary post-test data is also approximately equal to ‘$0,24$’ and therefore significantly below ‘$0,5$’. This allows us to conclude that this data set is highly clusterable as well, based on the similarity of 1) the children’s failure rates on the pre-test, 2) those on the **restricted** binary post-test data, and 3) the number of main exploratory learning assignments completed by the children (all data in %).

Following the `NbClust` package and according to the majority rule, the optimal number of clusters is again two (proposed by $seven$ indices), followed by equally optimal options of $three$ clusters and $five$ clusters (both proposed by $five$ indices) (cf. Table 6.29). To test our hypothesis, it seems appropriate to use the same number of clusters used for the data set containing the **complete** binary post-test data, which has already enabled us to clearly differentiate the average performers in addition to low and high performers (cf. subsubsection 6.4.1.2).

Table 6.29: Optimal number of clusters less than or equal to six

optimal number of clusters	0	1	2	3	4	5	6
proposed by $\cdots$ **indices**	2	1	7	5	2	5	4

All the children in the test group remain important for our statistical analysis and for testing hypothesis two, i.e. including the outliers.

Again, we perform `PAM` (Partition Around Medoids) Clustering to assign each child to one of the $five$ clusters corresponding to the nearest medoid (Maechler et al., 2019). The medoids are not necessarily the same as those determined for the data set containing the **complete** binary post-test data. In fact, *"partitioning around medoids (`PAM`) (Kaufman and Rousseeuw, 1990) starts from an initial set of medoids, and iteratively replaces one of the medoids by one of the non-medoids if it improves the total distance of the resulting clustering" (Maimon and Rokach, 2010, pg. 480).* Consequently the composition of the clusters is not the same as that of the **complete** binary post-test data set.

Finally, we create two- and $three$-dimensional visualization plots of the clustering results comparable to those of the data set containing the **complete** binary post-test data (cf. subsubsection 6.4.1.3); the same colors as in the previous subsection (6.4.1) are used.

6.4.2.2 Summary of the clusters obtained through `PAM` cluster analysis

Let us summarize once again some information obtained when using `PAM` (Partition Around Medoids) Clustering. The table colors are still consistent with the rest of this section.

Table 6.30: Table containing 1) the description name of each cluster, 2) the number of children per cluster and 3) cluster medoid information using PAM (Partition Around Medoids) Clustering

cluster name	description name performance group	number of children	cluster medoid child	Tool [a]	Pre [b]	Post [c]
*Cluster*1	Best-Implementer_of_Tool_Topics	35	$XC0617$	$78,38$	$33,07$	$10,97$
*Cluster*2	Average-Performers-Best-Progress	41	$XB0308$	$78,38$	$47,94$	$19,19$
*Cluster*3	Higher-Performers	28	$XF1204$	$100,00$	$23,98$	$14,89$
*Cluster*4	Average-Performers-Strength_Tool	33	$XE1005$	$100,00$	$42,99$	$26,46$
*Cluster*5	Lower-Performers	27	$XD0811$	$78,38$	$53,73$	$38,37$

[a] *Tool* represents the number of main exploratory learning assignments the child has completed (in %)
[b] = the child's failure rate on the pre-test (in %)
[c] = the child's failure rate on the **'restricted'** post-test (in %)

- *Cluster*3 contains the children who generally had the **lowest failure rates** on the pre-test, and who came **closest to completing** our sequence of dynamic learning assignments. Regarding the failure rates on the post-test, the children in *Cluster*1 generally received slightly better results. However we have to consider that for the children in *Cluster*3, some **highly complex items** are included in their post-test results which were omitted for children who did not explore the related topic within the GeoGebraTAO phase.
- *Cluster*5 contains the children who generally had the **highest failure rates** on the pre- and post-test.
- *Cluster*4 contains the children who progressed **almost as far as** the children in *Cluster*3 in the sequence of dynamic learning assignments, but had **higher failure rates** on the pre- and post-test. In general they had **higher failure rates** on the pre- and post-test than *Cluster*1, and **higher failure rates** on the post-test than *Cluster*2.
- *Cluster*1 contains the children who **progressed significantly less** in the GeoGebraTAO tool sequence than the children in *Clusters*3&4. However, this cluster is **quite interesting** because its children **performed best** in terms of converting topics they explored themselves within the GeoGebraTAO phase into a paper-and-pencil assessment.
- *Cluster*2 contains the children who generally performed **worse** on the pre-test compared to the children in *Clusters*3&1, but **better** than those in *Cluster*5. Regarding their progress in the sequence of assignments offered by our tool, they got **as far as** the children in *Cluster*1. Nevertheless, **most interesting** is these children's *good / great* **progress** from the pre- to post-test.

We should note that the children in *Cluster*4 probably moved too quickly through the sequence of dynamic learning activities, and thus did not achieve a deepened understanding of some topics. In particular, in the next subsection we will analyse through **Generalized Mixed-Effects Models** whether these children watched significantly fewer *short* video clips aimed at improving understanding when needed than the children in other clusters.

Regarding the children in *Cluster*1 and *Cluster*2, who made *good* progress from the pre- to post-test and successfully translated their acquired knowledge from our GeoGebraTAO tool into similar paper-and-pencil exercises, we will further explore their way of working with our tool and how they progressed along the sequence of dynamic learning assignments. Among other things, we will analyse whether these children

- watched more *short* video clips compared to the children in other clusters;
- were redirected to prior learning assignments and additional activities more often than the children in other clusters.

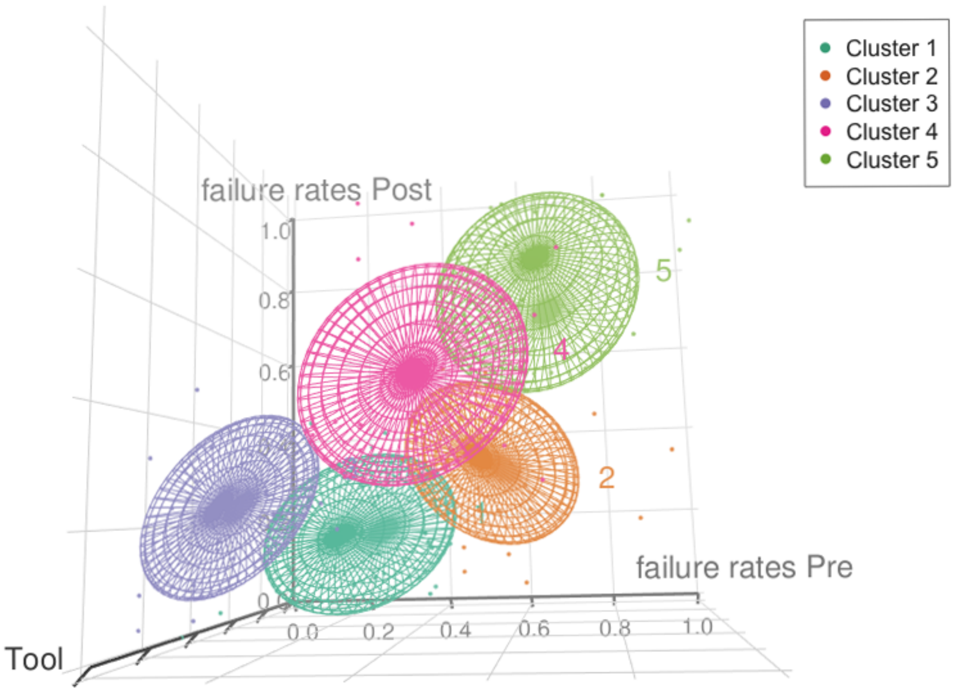
By comparing the children in the clusters using the **complete** binary post-test data set with those in the clusters using the **restricted** binary data set, we observed that

- 125 children of the 164 ($\simeq 76,22\%$) are in equivalent clusters, i.e. in clusters having the same cluster name;
- 17 children of the 164 ($\simeq 10,37\%$) 'move' from *Cluster*2 (**complete** data set) into *Cluster*5 (**restricted** data set);
- 13 children of the 164 ($\simeq 7,93\%$) 'move' from *Cluster*1 (**complete** data set) into *Cluster*2 (**restricted** data set).

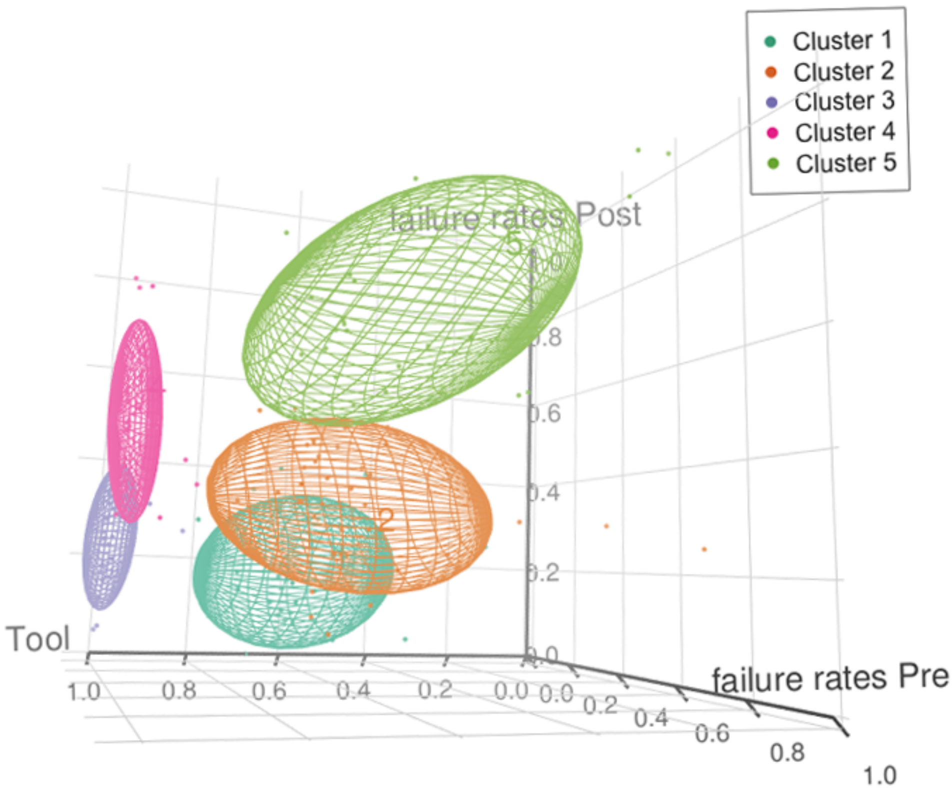
6.4.2.3 Visualizations of the performance groups

Figure 6.8: Two-dimensional visualization of the performance groups nested within the test group; children are represented by points in the plot, using principal components. The data set containing the **restricted** binary post-test data is used.

*Cluster*1:	Best-Implementer_of_Tool_Topics	*Cluster*4:	Average-Performers-Strength_Tool
*Cluster*2:	Average-Performers-Best-Progress	*Cluster*5:	Lower-Performers
*Cluster*3:	Higher-Performers		



(View 1)



(View 2)

Figure 6.9: $Three$-dimensional visualizations of the performance groups nested within the test group; the **restricted** binary post-test data set is used

6.4.3 Cluster-based comparisons in relation to the different help modi offered by the GEOGEBRATAO tool

6.4.3.1 Introduction

In this part of the analysis, we limit ourselves to clusters based on the **'restricted'** binary post-test data set. This clustering provides the same information as that based on the **'complete'** binary post-test data set, but is clearer and more conclusive. Furthermore, it seems logical to perform this analysis on the topics that the children actually explored within the GEOGEBRATAO phase.

To introduce this analysis, we provide below an overview of the number of children in the different classes spread across the $five$ clusters based on the **'restricted'** binary post-test data set, as mentioned above. In addition, we specify the number of girls and boys to get a clearer idea of how the children are distributed within each cluster.

Table 6.31: Table containing the number of children in the different classes spread across the $five$ clusters, additionally subdivided by gender

	XA01	*XA02*	*XB03*	*XB04*	*XC05*	*XC06*	*XD07*	*XD08*	*XE09*	*XE10*	*XF11*	*XF12*	*Total*
*Cluster*1 [a]	4 ♀: 2 ♂: 2	4 ♀: 2 ♂: 2	1 ♀: 1 ♂: 0	6 ♀: 2 ♂: 4	2 ♀: 1 ♂: 1	5 ♀: 2 ♂: 3	2 ♀: 2 ♂: 0	0 ♀: 0 ♂: 0	3 ♀: 2 ♂: 1	5 ♀: 3 ♂: 2	1 ♀: 1 ♂: 0	2 ♀: 1 ♂: 1	35 ♀: 19 ♂: 16
*Cluster*2 [b]	7 ♀: 4 ♂: 3	2 ♀: 1 ♂: 1	5 ♀: 4 ♂: 1	3 ♀: 2 ♂: 1	3 ♀: 1 ♂: 2	9 ♀: 5 ♂: 4	3 ♀: 1 ♂: 2	4 ♀: 2 ♂: 2	3 ♀: 3 ♂: 0	1 ♀: 1 ♂: 0	0 ♀: 0 ♂: 0	1 ♀: 1 ♂: 0	41 ♀: 25 ♂: 16
*Cluster*3 [c]	2 ♀: 1 ♂: 1	5 ♀: 4 ♂: 1	1 ♀: 0 ♂: 1	1 ♀: 0 ♂: 1	3 ♀: 0 ♂: 3	0 ♀: 0 ♂: 0	2 ♀: 1 ♂: 1	4 ♀: 2 ♂: 2	3 ♀: 1 ♂: 2	3 ♀: 2 ♂: 1	1 ♀: 0 ♂: 1	3 ♀: 2 ♂: 1	28 ♀: 13 ♂: 15
*Cluster*4 [d]	2 ♀: 0 ♂: 2	2 ♀: 0 ♂: 2	4 ♀: 1 ♂: 3	0 ♀: 0 ♂: 0	3 ♀: 0 ♂: 3	0 ♀: 0 ♂: 0	3 ♀: 1 ♂: 2	4 ♀: 2 ♂: 2	4 ♀: 1 ♂: 3	3 ♀: 2 ♂: 1	1 ♀: 1 ♂: 0	7 ♀: 3 ♂: 4	33 ♀: 11 ♂: 22
*Cluster*5 [e]	1 ♀: 0 ♂: 1	5 ♀: 3 ♂: 2	1 ♀: 0 ♂: 1	2 ♀: 1 ♂: 1	1 ♀: 1 ♂: 0	2 ♀: 2 ♂: 0	1 ♀: 0 ♂: 1	2 ♀: 1 ♂: 1	1 ♀: 0 ♂: 1	3 ♀: 1 ♂: 2	6 ♀: 2 ♂: 4	2 ♀: 1 ♂: 1	27 ♀: 12 ♂: 15
Total	16 ♀: 7 ♂: 9	18 ♀: 10 ♂: 8	12 ♀: 6 ♂: 6	12 ♀: 5 ♂: 7	12 ♀: 3 ♂: 9	16 ♀: 9 ♂: 7	11 ♀: 5 ♂: 6	14 ♀: 7 ♂: 7	14 ♀: 7 ♂: 7	15 ♀: 9 ♂: 6	9 ♀: 4 ♂: 5	15 ♀: 8 ♂: 7	164 ♀: 80 ♂: 84

[a] Cluster1 = Best-Implementer_of_Tool_Topics
[b] Cluster2 = Average-Performers-Best-Progress
[c] Cluster3 = Higher-Performers
[d] Cluster4 = Average-Performers-Strength_Tool
[e] Cluster5 = Lower-Performers

The following classes are most prevalent in each respective cluster:

- $50,00\%$ of the children of class *XB04* belong to *Cluster*1; they represent $17,14\%$ of this cluster;
- $56,25\%$ of the children of class *XC06* belong to *Cluster*2; they represent $21,95\%$ of this cluster;
- $27,78\%$ of the children of class *XA02* belong to *Cluster*3; they represent $17,86\%$ of this cluster;
- $46,67\%$ of the children of class *XF12* belong to *Cluster*4; they represent $21,21\%$ of this cluster;
- $66,67\%$ of the children of class *XF11* belong to *Cluster*5; they represent $22,22\%$ of this cluster.

The children in the rest of the classes are more or less spread evenly across the $five$ clusters. Furthermore, we note that in *Cluster*4, only one-third of the children are girls; two-thirds are boys.

In the following subsubsections we will, among other things, more precisely analyse some common ways of working with our dynamic geometry tool within the clusters, leading to overall improvement at an individualized level.

6.4.3.2 Fitted Generalized Mixed-Effects Models for the *short* video clips

To determine the children's common way of working within each cluster and to compare the $five$ clusters, we first focus on an affordance available on a **voluntary basis**, i.e. the *short* video clips that children could consult. These video clips also include some static pictures and short texts explaining specific concepts.

We calculate a percentage based on

1. the total number of **distinct** *short* video clips consulted compared to the total number available to the individual child;
2. the number of clicks on the *short* video clips (including **multiple clicks** on the same video clip) compared to the total number of *short* video clips available to the individual child,

and fit the following **Generalized Mixed-Effects Model** to determine some of the characteristics of each of the clusters:

$$\text{glmmPQL\_shVid} \leftarrow glmmPQL(\underbrace{\text{WatchedPct}}_{\text{response variable}} \sim \underbrace{\text{WhichPct * Cluster * Gender}}_{\text{fixed effects}},\ \underbrace{\text{random} = \text{list(}\ \text{Children} =\sim 1\ )}_{\text{random effects}},$$
$$\text{family} = \text{Gamma(link} = \text{`identity'}),\ \text{data} = \text{datCluster\_shVid}\,) \tag{6.5}$$

where

glmmPQL_shVid	:	Generalized Linear Mixed Model (**GLMM** model) with multivariate normal random effects, using Penalized Quasi-Likelihood (**PQL**) for cluster-based comparisons based on the percentages of ***sh***ort **vid**eo clips consulted by each child;
datCluster_shVid	:	data set containing the clusters built and information about the *short* video clips;
WatchedPct	:	variable representing the number of *short* video clips watched, or rather clicked on by the child, compared to the total number of *short* video clips available to the child (expressed as a percentage);
WhichPct	:	factor having the levels *Different_Videos_Consulted* (*different short video clips consulted*) and *Multiple_Clicks_Considered* (*including multiple clicks on the same short video clip*);
Cluster	:	factor representing the cluster numbers;
Gender	:	factor having the levels B (*boy*) and G (*girl*);
Children	:	variable obviously representing the children.

A similar **Generalized Mixed-Effects Model** is fitted to compare the $twelve$ test classes in order to detect any classes who might have worked differently than others due to possible technical issues (such as slow internet connection) or excessive help from the classroom teacher. The former issue could lead to impatience with the clips' loading speed and the latter issue might cause children not to see any benefit to consulting this additional help offered by the GEOGEBRATAO tool.

6.4.3.3 Analysis and interpretation of the *short* video clips at *Cluster* level

The results provided by the Analysis of Deviance Table *(Type II tests)* (cf. J.1 : appendix J) suggest a highly significant effect on the variable *WatchedPct*

- of the factor *Cluster* ('***');
- of the $three$-factor interaction, i.e. of the two-factor interaction '*WhichPct* by *Cluster*' varying across levels of the factor *Gender* ('***');
- of the two-factor interaction '*WhichPct* by *Cluster*' (p-value $\simeq 0,005$) ('**').

Clearly, the factor *Cluster* created through PAM cluster analysis has a significant effect on the number of *short* video clips watched, or at least clicked on, by the child, allowing us to provide initial evidence of some common ways of working with the GEOGEBRATAO tool within clusters.

To clearly present the outcome of the fitted Generalized Mixed-Effects Model *glmmPQL_shVid* 6.5 (cf. J.2 : appendix J), we summarize the results in two tables.

The first table (see Table 6.32) shows the *directed* differences between *Cluster*1 (base level of the factor *Cluster* in our analysis) and every other cluster for both levels of factor *WhichPct*, i.e. for both

1. the percentage of **distinct** *short* video clips consulted and
2. the percentage of clicks on the *short* video clips (including **multiple clicks** on the same video clip).

In the second table (see Table 6.33) we provide the *directed* differences between both levels of factor *WhichPct* for each of the $five$ clusters at the *Gender* level. This enables us to analyse whether any cluster(s) or gender-based subgroup(s) tend(s) to watch the *short* video clips more or less often than the other clusters or gender-based subgroups.

Note that throughout the remainder of this section, *Test of predicted hypothesis* two, we follow the procedure of summarizing the results in two tables similar to those described above.

Table 6.32: *Directed* differences between *Cluster*1 (base level of the factor *Cluster* of the analysis) and every other cluster both
- for the percentage of **distinct** *short* video clips consulted and
- for the percentage of clicks on the *short* video clips (including **multiple clicks** on the same video clip)

between the clusters	***directed* difference for** **Different_Videos_Consulted**	**Multiple_Clicks_Considered**
1 *and* 2 *(Girls)*	$-0{,}23$	$(-0{,}23) + (-3{,}82) = -4{,}05$ [a]
1 *and* 2 *(Boys)*	$(-0{,}23) + (-3{,}85) = -4{,}08$	$(-0{,}23) + (-3{,}85) + (-3{,}82) + 6{,}52 = -1{,}38$ [b]
1 *and* 3 *(Girls)*	$5{,}82$	$5{,}82 + 0{,}03 = 5{,}85$
1 *and* 3 *(Boys)*	$5{,}82 + 2{,}09 = 7{,}91$	$5{,}82 + 2{,}09 + 0{,}03 + 1{,}73 = 9{,}67$
1 *and* 4 *(Girls)*	$7{,}54$	$7{,}54 + 6{,}86 = 14{,}4$
1 *and* 4 *(Boys)*	$7{,}54 + 3{,}36 = 10{,}9$	$7{,}54 + 3{,}36 + 6{,}86 + (-2{,}08) = 15{,}68$
1 *and* 5 *(Girls)*	$-0{,}58$	$(-0{,}58) + 13{,}78 = 13{,}2$
1 *and* 5 *(Boys)*	$(-0{,}58) + 4{,}67 = 4{,}09$	$(-0{,}58) + 4{,}67 + 13{,}78 + (-15{,}62) = 2{,}25$

[a] Sum of the values of the effects:
$Cluster2$,
$WhichPctMultiple\_Clicks\_Considered : Cluster2$

[b] Sum of the values of the effects:
$Cluster2$,
$Cluster2 : GenderB$,
$WhichPctMultiple\_Clicks\_Considered : Cluster2$,
$WhichPctMultiple\_Clicks\_Considered : Cluster2 : GenderB$

*Cluster*1 = Best-Implementer_of_Tool_Topics
*Cluster*2 = Average-Performers-Best-Progress
*Cluster*3 = Higher-Performers
*Cluster*4 = Average-Performers-Strength_Tool
*Cluster*5 = Lower-Performers

By considering the **separate** *short* video clips, Table 6.32 shows that both genders in *Cluster*4 watched **the most** video clips, followed by *Cluster*3. It should be noted that both clusters came closest to finishing our sequence of dynamic learning assignments, which does not affect the results since the percentage of video clips consulted was calculated with respect to the total number of *short* video clips available to each individual child.

Regarding the number of clicks on *short* video clips and including **multiple clicks** on the same video clip, the boys and girls in *Cluster*4 once again used this assistance (affordance), offered by the GEOGEBRATAO tool, the **most often**. In addition, the girls in *Cluster*5 voluntarily clicked on the *short* video clips **almost as many times** as the pupils in *Cluster*4. The children in *Cluster*3 consulted this help modus **more frequently** than the children in *Clusters*1,2&5(♂), but **less frequently** than those in *Clusters*4&5(♀).

Returning to the analysis of the **different** *short* video clips, the girls in *Clusters*1,2&5 all tended to click on roughly the **same number** of video clips; for the boys in these clusters, this rate **varied slightly**.

Furthermore, we observe in Figure 6.10 that the error bars, representing confidence intervals (CIs)[5], of *Clusters*1,2&5 are **narrower** than those of *Clusters*3&4, except the error bar related to the girls in *Cluster*5 for *Multiple_Clicks_Considered*. The narrower the error bars, the more concentrated and precise the values representing the estimates of the true number of *short* video clips watched for the respective cluster. This means that the respective *'sample'* mean is more likely and reliable. Confidence intervals (CIs) directly related to standard errors can be considered measures *"of the precision of the sample mean. The standard error of the sample mean depends on both the standard deviation and the sample size" (Altman and Bland, 2005, pg. 903).*

The error bars related to level *Multiple_Clicks_Considered* overlap less than those of level *Different_Videos_Consulted*; therefore, the differences between clusters may be more significant: *"The smaller the overlap of bars, or the larger the gap between bars, the smaller the P value and the stronger the evidence for a true difference" (Cumming et al., 2007, pg. 11).*

Table 6.33: *Directed* difference between
- the percentage of **distinct** *short* video clips consulted and
- the percentage of clicks on the *short* video clips (including **multiple clicks** on the same video clip)
for each of the $five$ clusters at the *Gender* level

Cluster	***directed* difference between Different_Videos_Consulted and Multiple_Clicks_Considered (in %)** Girls	Boys
*Cluster*1 (♀: 19, ♂: 16)	$3{,}82$	$3{,}82 + (-1{,}97) = 1{,}85$
*Cluster*2 (♀: 25, ♂: 16)	$3{,}82 + (-3{,}82) = 0$ [a]	$3{,}82 + (-1{,}97) + (-3{,}82) + 6{,}52 = 4{,}55$ [b]
*Cluster*3 (♀: 13, ♂: 15)	$3{,}82 + 0{,}03 = 3{,}85$	$3{,}82 + (-1{,}97) + 0{,}03 + 1{,}73 = 3{,}61$
*Cluster*4 (♀: 11, ♂: 22)	$3{,}82 + 6{,}86 = 10{,}68$	$3{,}82 + (-1{,}97) + 6{,}86 + (-2{,}08) = 6{,}63$
*Cluster*5 (♀: 12, ♂: 15)	$3{,}82 + 13{,}78 = 17{,}6$	$3{,}82 + (-1{,}97) + 13{,}78 + (-15{,}62) = 0{,}01$

[a] Sum of the values of the effects:
$WhichPctMultiple\_Clicks\_Considered$,
$WhichPctMultiple\_Clicks\_Considered : Cluster2$

[b] Sum of the values of the effects:
$WhichPctMultiple\_Clicks\_Considered$,
$WhichPctMultiple\_Clicks\_Considered : GenderB$,
$WhichPctMultiple\_Clicks\_Considered : Cluster2$,
$WhichPctMultiple\_Clicks\_Considered : Cluster2 : GenderB$

From Table 6.33 we learn that in general, the girls in *Cluster*5, followed by the girls in *Cluster*4, tended to consult some *short* video clips **more often** than the remaining children. The boys in *Cluster*4 performed in a similar, but less pronounced, way than both groups of girls.

[5] According to Cumming et al., *"for people reading $\cdots$ CIs make things easier to understand" "$\cdots$ you can easily swap in your mind's eye between SE bars and 95% CIs. If a figure shows SE bars you can mentally double them in width, to get approximate 95% CIs, as long as n is 10 or more." "$M \pm 2 \times SE$ intervals are quite good approximations to 95% CIs when n is 10 or more $\cdots$" "CIs can be thought of as SE bars that have been adjusted by a factor (t) so they can be interpreted the same way, regardless of n" (Cumming et al., 2007, pg. 10).*
(SE = standard errors, n = number of data points, M = mean of the data)

However we have no knowledge or proof of whether any children really watched the video clips or just clicked on them, nor do we know how long they focused on each video clip. Some children probably used this affordance offered by the GEOGEBRATAO tool almost automatically due to concentration difficulties. Some pupils might have difficulty recognizing the essential learning strategies of this *'new'* tool and therefore may not have used the clips as intended. (Naparstek, 2010)

This lack of knowledge could be studied in a further project focusing on the child's gestures, genuine constituents of thinking (Radford, 2009), and gazes while using the GEOGEBRATAO tool, perhaps in a usability laboratory. McNeill claims that gestures are a necessary component of thinking and speaking, and that gestures reflect an online dynamic process of the *"dialectic between imagery and language" (Kita and Özyürek, 2007)*. In addition, they could provide information on children's concentration levels during the GEOGEBRATAO learning phase, thus opening new doorways in our research field.

The girls in *Cluster*2 and the boys in *Cluster*5 tended to watch the video clips **only once**.

Knowing that the girls in *Cluster*5 tended to consult some video clips **more often** than the remaining children in the test group, and that the boys in the same cluster tended to watch the video clips **only once**, we can affirm that a *'gender gap'* seems to exist in *Cluster*5 regarding the *voluntary* use of the affordance provided. However, this *'gender gap'* obviously can't be explained because of the small cluster size.

The remaining subgroups not yet mentioned only replayed **a few** video clips and showed no other notable behaviors.

Figure 6.10 illustrates and clarifies the previous observations.

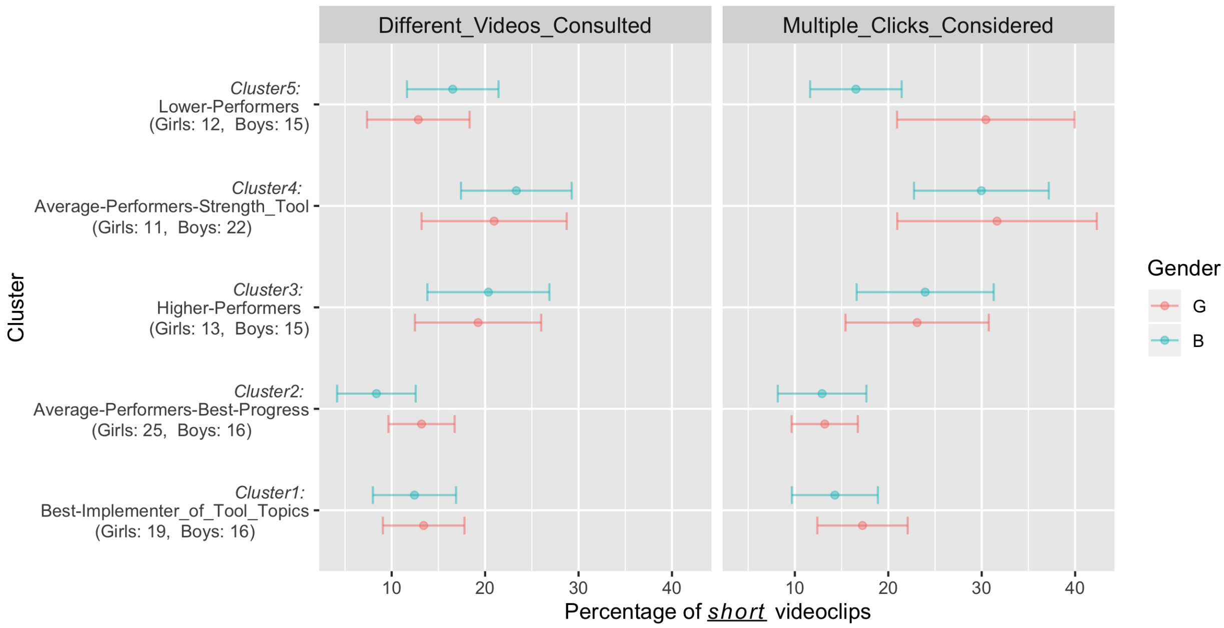


Figure 6.10: **Cluster-based** comparisons regarding the number of *short* video clips watched, or clicked on, *voluntarily* by each child in relation to the total number available to the child (expressed as a percentage).
These comparisons were made both
- for the total number of **distinct** *short* video clips consulted, and
- for the number of clicks on the *short* video clips, including **multiple clicks** on the same video clip.
This figure shows 95% CI (confidence interval) bars.
The *percentage* values have been fitted through the Generalized Mixed-Effects Model 6.5.

6.4.3.4 Summary table

In the following subsections, we always summarize our findings in a table, i.e. we rank the use of the affordances provided by the GEOGEBRATAO tool on a rating scale from one to $five$, with

- $\underline{one}$ representing an affordance that is **used much less**;
- $\underline{five}$ representing an affordance that is **used much more**

by the cluster in question than by the remaining clusters. The rating scale from one to $five$ is defined based on the difference of the lowest and the highest calculated *directed* differences (see Table 6.32, Table 6.33), subdivided into $five$ intervals (per gender). Then we calculate in which of these intervals the *directed* difference of each cluster falls.

For instance (in the case of Table 6.33 (♀)), a *directed* difference of

- $3,85$ falls in the $second$ interval;
- $10,68$ falls in the $fourth$ interval;

the $five$ intervals are $[0; 3,52]$, $]3,52; 7,04]$, $]7,04; 10,56]$, $]10,56; 14,08]$ and $]14,08; 17,60]$.

This and the following summary tables are extremely important for the conclusion of hypothesis two.

Table 6.34: The use of the affordance *'short video clips'* on a rating scale from 1 to 5, with 1 representing an affordance that is used much less by the cluster in question than by the remaining clusters and 5 an affordance that is used much more, always considering *Gender*

voluntary (*) use of available affordances	***Cluster*1 Best-Implementer_ of_Tool_Topics**		***Cluster*2 Average-Performers-Best-Progress**		***Cluster*3 Higher-Performers**		***Cluster*4 Average-Performers-Strength_Tool**		***Cluster*5 Lower-Performers**	
(*) different *short* video clips (Table 6.32)	♀: 1	♂: 2	♀: 1	♂: 1	♀: 4	♂: 4	♀: 5	♂: 5	♀: 1	♂: 3
(*) $re-$watchings of *short* video clips (Table 6.33)	♀: 2	♂: 2	♀: 1	♂: 4	♀: 2	♂: 3	♀: 4	♂: 5	♀: 5	♂: 1

After having accomplished our analysis and interpretation of the *short* video clips at ***Cluster* level**, we will examine, in a similar way, these *short* video clips data at ***Class* level**.

6.4.3.5 Analysis and interpretation of the *short* video clips at *Class* level

Let us take a look at the results provided by the Analysis of Deviance Table *(Type II tests)* (cf. K.1 : appendix K). The two-factor interaction '*WhichPct* by *Class*' has a highly significant effect on the variable *WatchedPct* ('*** '), followed by the $three$-factor interaction, i.e. by two-factor interaction '*WhichPct* by *Class*' varying across levels of the factor *Gender* ('** ').

The factor *Class* has a less significant effect on the number of *short* video clips watched ('* ') than the factor *Cluster*, which has a highly significant effect on this number (cf. subsubsection 6.4.3.3).

Table 6.35: *Directed* differences between the reference test class *ClassXA01* and every other test class both
- for the percentage of **distinct** *short* video clips consulted and
- for the percentage of clicks on the *short* video clips, including **multiple clicks** on the same video clip

between the classes	***directed* difference for** Different_Videos_Consulted	Multiple_Clicks_Considered
XA01 and XA02 (Girls)	$-1{,}68$	$(-1{,}68)+(-4{,}51)=-6{,}19$ [a]
XA01 and XA02 (Boys)	$(-1{,}68)+(-0{,}26)=-1{,}94$	$(-1{,}68)+(-0{,}26)+(-4{,}51)+(-1{,}68)=-8{,}13$ [b]
XA01 and XB03 (Girls)	$6{,}16$	$6{,}16+5{,}64=11{,}8$
XA01 and XB03 (Boys)	$6{,}16+2{,}11=8{,}27$	$6{,}16+2{,}11+5{,}64+(-3{,}03)=10{,}88$
XA01 and XB04 (Girls)	$3{,}2$	$3{,}2+(-0{,}01)=3{,}19$
XA01 and XB04 (Boys)	$3{,}2+(-0{,}06)=3{,}14$	$3{,}2+(-0{,}06)+(-0{,}01)+1{,}03=4{,}16$
XA01 and XC05 (Girls)	$-6{,}29$	$(-6{,}29)+14{,}47=8{,}18$
XA01 and XC05 (Boys)	$(-6{,}29)+6{,}4=0{,}11$	$(-6{,}29)+6{,}4+14{,}47+(-12{,}43)=2{,}15$
XA01 and XC06 (Girls)	$-3{,}7$	$(-3{,}7)+8{,}29=4{,}59$
XA01 and XC06 (Boys)	$(-3{,}7)+(-1{,}58)=-5{,}28$	$(-3{,}7)+(-1{,}58)+8{,}29+(-16{,}01)=-13$
XA01 and XD07 (Girls)	$6{,}25$	$6{,}25+0{,}79=7{,}04$
XA01 and XD07 (Boys)	$6{,}25+(-2{,}31)=3{,}94$	$6{,}25+(-2{,}31)+0{,}79+(-8{,}16)=-3{,}43$
XA01 and XD08 (Girls)	$1{,}08$	$1{,}08+(-3{,}78)=-2{,}7$
XA01 and XD08 (Boys)	$1{,}08+(-1{,}56)=-0{,}48$	$1{,}08+(-1{,}56)+(-3{,}78)+(-0{,}85)=-5{,}11$
XA01 and XE09 (Girls)	$-3{,}33$	$(-3{,}33)+(-5{,}02)=-8{,}35$
XA01 and XE09 (Boys)	$(-3{,}33)+3{,}83=0{,}5$	$(-3{,}33)+3{,}83+(-5{,}02)+(-0{,}83)=-5{,}35$
XA01 and XE10 (Girls)	$1{,}91$	$1{,}91+1{,}46=3{,}37$
XA01 and XE10 (Boys)	$1{,}91+(-0{,}51)=1{,}4$	$1{,}91+(-0{,}51)+1{,}46+13{,}17=16{,}03$
XA01 and XF11 (Girls)	$1{,}55$	$1{,}55+(-4{,}25)=-2{,}7$
XA01 and XF11 (Boys)	$1{,}55+0{,}95=2{,}5$	$1{,}55+0{,}95+(-4{,}25)+1{,}74=-0{,}01$
XA01 and XF12 (Girls)	$2{,}49$	$2{,}49+3{,}09=5{,}58$
XA01 and XF12 (Boys)	$2{,}49+(-1{,}93)=0{,}56$	$2{,}49+(-1{,}93)+3{,}09+(-10{,}82)=-7{,}17$

[a] Sum of the values of the effects:
ClassXA02,
WhichPctMultiple_Clicks_Considered:ClassXA02

[b] Sum of the values of the effects:
ClassXA02,
ClassXA02:GenderB,
WhichPctMultiple_Clicks_Considered:ClassXA02,
WhichPctMultiple_Clicks_Considered:ClassXA02:GenderB

Table 6.35 and Table 6.36 summarize the outcome of the fitted Generalized Mixed-Effects Model at *Class* level.

By considering the **different** *short* video clips, Table 6.35 shows that both genders in classes *XB03* and *XD07* watched **the most** video clips, while the children in class *XC06* and the girls in class *XC05* watched **the fewest**. Classes *XC05* and *XC06* were the iPads classes with a poor Internet connection, which might explain limited use of the video clips. As for the rest, the differences between the classes are not noteworthy.

Regarding the *short* video clips when including **multiple clicks** on the same video clip, the children in class *XB03*, the boys in class *XE10*, and the girls in classes *XC05* and *XD07* clicked **most frequently** on the video clips. In contrast, the children in classes *XA02* and *XE09* and the boys in classes *XC06* and *XF12* clicked **least** frequently on this help modus.

Table 6.36: *Directed* difference between
- the percentage of **distinct** *short* video clips consulted and
- the percentage of clicks on the *short* video clips (including **multiple clicks** on the same video clip)
for each of the *twelve* test classes at the *Gender* level

Class	*directed* difference between Different_Videos_Consulted and Multiple_Clicks_Considered (in %)	
	Girls	**Boys**
XA01 (♀: 7, ♂: 9)	5,02	$5{,}02 + 2{,}7 = 7{,}72$
XA02 (♀: 10, ♂: 8)	$5{,}02 + (-4{,}51) = 0{,}51$ [a]	$5{,}02 + 2{,}7 + (-4{,}51) + (-1{,}68) = 1{,}53$ [b]
XB03 (♀: 6, ♂: 6)	$5{,}02 + 5{,}64 = 10{,}66$	$5{,}02 + 2{,}7 + 5{,}64 + (-3{,}03) = 10{,}33$
XB04 (♀: 5, ♂: 7)	$5{,}02 + (-0{,}01) = 5{,}01$	$5{,}02 + 2{,}7 + (-0{,}01) + 1{,}03 = 8{,}74$
XC05 (♀: 3, ♂: 9)	$5{,}02 + 14{,}47 = 19{,}49$	$5{,}02 + 2{,}7 + 14{,}47 + (-12{,}43) = 9{,}76$
XC06 (♀: 9, ♂: 7)	$5{,}02 + 8{,}29 = 13{,}31$	$5{,}02 + 2{,}7 + 8{,}29 + (-16{,}01) \simeq 0$
XD07 (♀: 5, ♂: 6)	$5{,}02 + 0{,}79 = 5{,}81$	$5{,}02 + 2{,}7 + 0{,}79 + (-8{,}16) = 0{,}35$
XD08 (♀: 7, ♂: 7)	$5{,}02 + (-3{,}78) = 1{,}24$	$5{,}02 + 2{,}7 + (-3{,}78) + (-0{,}85) = 3{,}09$
XE09 (♀: 7, ♂: 7)	$5{,}02 + (-5{,}02) = 0$	$5{,}02 + 2{,}7 + (-5{,}02) + (-0{,}83) = 1{,}87$
XE10 (♀: 9, ♂: 6)	$5{,}02 + 1{,}46 = 6{,}48$	$5{,}02 + 2{,}7 + 1{,}46 + 13{,}17 = 22{,}35$
XF11 (♀: 4, ♂: 5)	$5{,}02 + (-4{,}25) = 0{,}77$	$5{,}02 + 2{,}7 + (-4{,}25) + 1{,}74 = 5{,}21$
XF12 (♀: 8, ♂: 7)	$5{,}02 + 3{,}09 = 8{,}11$	$5{,}02 + 2{,}7 + 3{,}09 + (-10{,}82) = -0{,}01$

[a] Sum of the values of the effects:
WhichPctMultiple_Clicks_Considered,
WhichPctMultiple_Clicks_Considered:ClassXA02

[b] Sum of the values of the effects:
WhichPctMultiple_Clicks_Considered,
WhichPctMultiple_Clicks_Considered:GenderB,
WhichPctMultiple_Clicks_Considered:ClassXA02,
WhichPctMultiple_Clicks_Considered:ClassXA02:GenderB

In Figure 6.11, we see that the error bars, representing confidence intervals (CIs), are narrower, are more homogeneous and overlap to a greater degree for the *Different_Videos_Consulted* than for the *Multiple_Clicks_Considered*, i.e.

- the respective *'sample'* means are more reliable for the *Different_Videos_Consulted* than for the *Multiple_Clicks_Considered*, and
- the differences between the classes are surely not significant at this level.

It should be noted that *"the smaller the overlap of bars, or the larger the gap between bars, the smaller the P value and the stronger the evidence for a true difference"* and *"when $n >= 10$, if CI error bars overlap by half the average arm length, $P \approx 0.05$. If the tips of the error bars just touch, $P \approx 0.01$" (Cumming et al., 2007, pg. 10-11)* ($P =$ the probability, $n =$ number of data points).

According to Table 6.36, the pupils in classes *XB03* and *XC05*, the boys in class *XE10*, and the girls in class *XC06* tended to repeat some *short* video clips **the most** compared to the remaining test group children. In contrast, the boys in classes *XC06*, *XF12* and *XD07* and the girls in classes *XE09*, *XA02* and *XF11* tended to consult each *voluntary* help video **only once**.

Two apparent *'gender gaps'* also emerge regarding the *voluntary* use of the affordance provided; one in class *XC06* and a second, less obvious one in class *XE10*. The girls' and boys' respective error bars in class *XC06* hardly overlap at the level *Multiple_Clicks_Considered*; in class *XE10*, this is not the case.

In Figure 6.11, we see the observations we have made and explained represented. Next, we will analyse the fitted Generalized Mixed-Effects Model for the *overlay* video clips that are no longer *voluntary*.

6.4.3.6 Fitted Generalized Mixed-Effects Model for the *overlay* videos

To continue our cluster characteristics analysis, we compare the percentages of *overlay* video clips displayed, i.e. **imposed** in some specific error scenarios.

As for the *short* video clips (cf. subsubsection 6.4.3.2), these percentages are based on

1. the total number of **different** *overlay* video clips displayed to each individual child compared to the total number of potential *overlay* video clips;
2. the number of *overlay* video clips displayed to each individual child, including **multiple displays** of the same video clip, compared to the total number of potential *overlay* video clips.

The fitted **Generalized Mixed-Effects Model** looks similar to the Model 6.5:

$$\text{glmmPQL\_ovlayVid} \leftarrow glmmPQL(\underbrace{\text{WatchedPct}}_{\text{response variable}} \sim \underbrace{\text{WhichPct * Cluster * Gender}}_{\text{fixed effects}}, \underbrace{\text{random} = \text{list(}\ \text{Children} =\sim 1\ )}_{\text{random effects}}, \\ \text{family} = \text{Gamma(link} = \text{`identity'}),\ \text{data} = \text{datCluster\_ovlayVid}\,) \tag{6.6}$$

where

glmmPQL_ovlayVid	: Generalized Linear Mixed Model (**GLMM** model) with multivariate normal random effects, using Penalized Quasi-Likelihood (**PQL**) for cluster-based comparisons based on the percentages of ***ov**er**lay*** **vid**eo clips displayed to each child;
datCluster_ovlayVid	: data set containing the clusters built and information about the *overlay* video clip displays;
WatchedPct	: variable representing the number of *overlay* video clips displayed to each individual child, compared to the total possible number of *overlay* video clips (expressed as a percentage);
WhichPct	: factor having the levels *Different_Videos_Displayed* (*different overlay video clips displayed to the child*) and *Multiple_Views_Considered* (*including multiple displays of the same overlay video clip*).

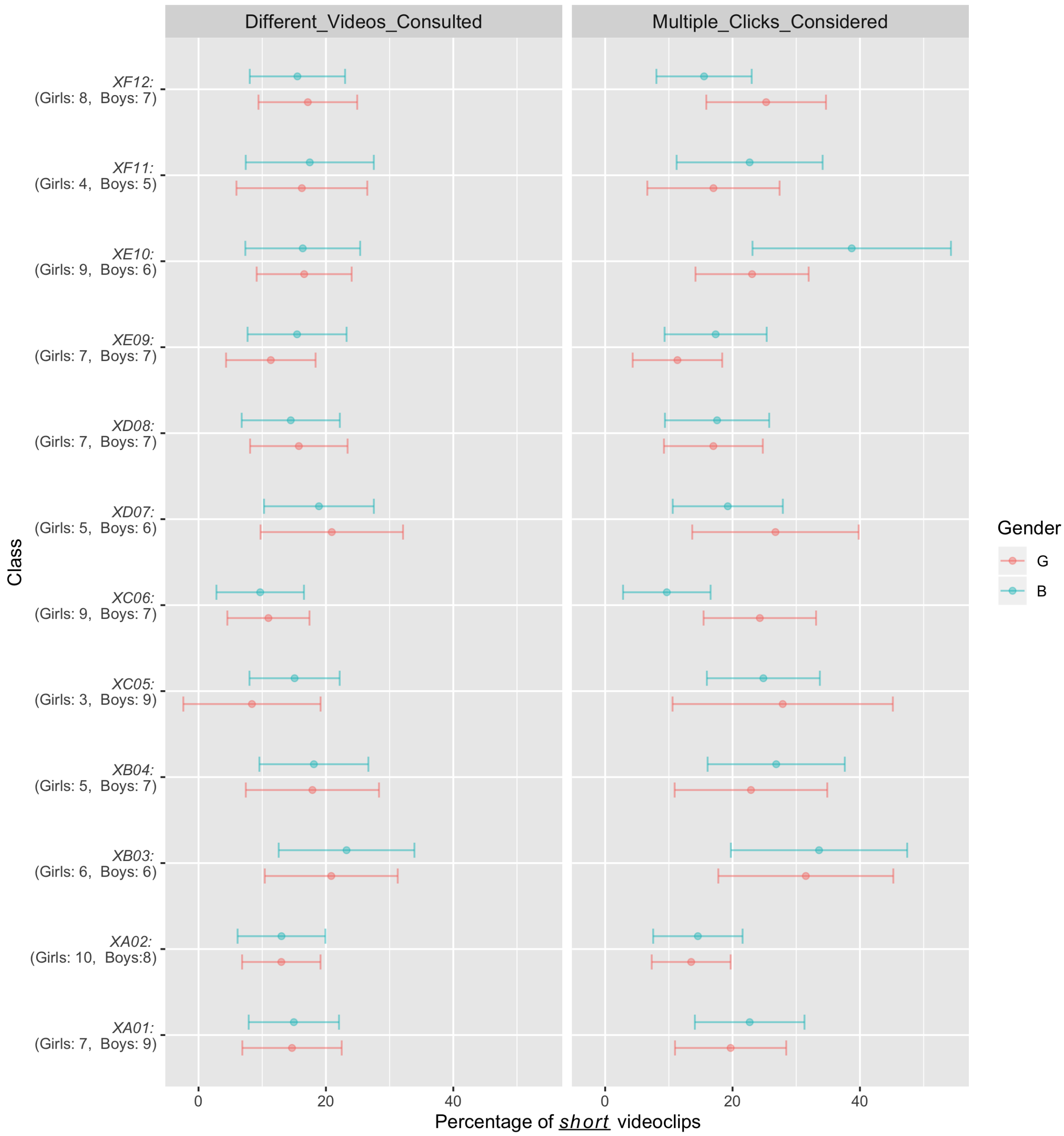


Figure 6.11: **Class-based** comparisons regarding the number of *short* video clips watched, or clicked on *voluntarily*, by each child in relation to the total number of *short* video clips available to the child (expressed as a percentage).
These comparisons were made both
- for the total number of **distinct** *short* video clips consulted, and
- for the number of clicks on the *short* video clips, including **multiple clicks** on the same video clip.
This figure shows 95% CI (confidence interval) bars.
The *percentage* values have been fitted through a Generalized Mixed-Effects Model similar to 6.5 except that factor *Cluster* has been replaced by factor *Class*.

6.4.3.7 Analysis and interpretation of the *overlay* videos at the *Cluster* level

According to the Analysis of Deviance Table *(Type II tests)* (cf. L.1 : appendix L),

- the factors *WhichPct* and *Cluster*,
- the two-factor interaction '*WhichPct* by *Cluster*'

seem to have a highly significant effect on the variable *WatchedPct* ('***'). This permits us to provide further evidence of some common ways of working with our dynamic geometry tool.

Again, two tables will summarize the results of the fitted Generalized Mixed-Effects Model *glmmPQL_ovlayVid* 6.6 related to the *overlay* video clips displayed at the cluster level (cf. L.2 : appendix L).

The first table (Table 6.37) includes the *directed* differences between *Cluster*1 (base level of factor *Cluster* in our analysis) and every other cluster for both levels of factor *WhichPct*.

Table 6.37: *Directed* differences between *Cluster*1 (base level of factor *Cluster*) and every other cluster both
- for the percentage of **different** *overlay* video clips displayed to the individual child and
- for the percentage of *overlay* video clips displayed when including **multiple displays** of the same video clip

between the clusters	***directed* difference for** Different_Videos_Displayed	Multiple_Views_Considered
1 *and* 2 (*Girls*)	$4{,}17$	$4{,}17 + 4{,}02 = 8{,}19$ [a]
1 *and* 2 (*Boys*)	$4{,}17 + 1{,}33 = 5{,}5$	$4{,}17 + 1{,}33 + 4{,}02 + 10{,}96 = 20{,}48$ [b]
1 *and* 3 (*Girls*)	$-1{,}81$	$(-1{,}81) + (-7{,}71) = -9{,}52$
1 *and* 3 (*Boys*)	$(-1{,}81) + (-0{,}99) = -2{,}8$	$(-1{,}81) + (-0{,}99) + (-7{,}71) + 4{,}31 = -6{,}2$
1 *and* 4 (*Girls*)	$4{,}5$	$4{,}5 + 5{,}3 = 9{,}8$
1 *and* 4 (*Boys*)	$4{,}5 + (-2{,}2) = 2{,}3$	$4{,}5 + (-2{,}2) + 5{,}3 + (-0{,}85) = 6{,}75$
1 *and* 5 (*Girls*)	$5{,}15$	$5{,}15 + 21{,}37 = 26{,}52$
1 *and* 5 (*Boys*)	$5{,}15 + 3{,}01 = 8{,}16$	$5{,}15 + 3{,}01 + 21{,}37 + (-4{,}97) = 24{,}56$

[a] Sum of the values of the effects:
$Cluster2$,
$WhichPctMultiple\_Views\_Considered : Cluster2$

[b] Sum of the values of the effects:
$Cluster2$,
$Cluster2 : GenderB$,
$WhichPctMultiple\_Views\_Considered : Cluster2$,
$WhichPctMultiple\_Views\_Considered : Cluster2 : GenderB$

*Cluster*1 = Best-Implementer_of_Tool_Topics
*Cluster*2 = Average-Performers-Best-Progress
*Cluster*3 = Higher-Performers
*Cluster*4 = Average-Performers-Strength_Tool
*Cluster*5 = Lower-Performers

First of all, we should recall that

- the *overlay* video clips are **enforced** in some specific error scenarios (cf. subsection 2.2.7, subsection 2.2.4). Interestingly, 10 *overlay* video clips were never imposed on any child because they were not needed;
- for all tasks, after a child's third incorrect attempt, which means after the sequence '1) errors - *overlay* video clip display, 2) errors - *overlay* video clip display, 3) errors -', the software locks automatically and displays a message to call the classroom teacher (cf. subsection 2.2.4). As a result, the number of times that the same *overlay* video clip can be watched is limited, unlike a *short* video clip. On a related note, the number of times a message to call the teacher is displayed will be studied in a later subsubsection.

By considering the **different** *overlay* video clips displayed, Table 6.37 shows that the pupils of both genders in *Cluster*3 and *Cluster*1 had the **fewest** video clips imposed, which also means that they failed to pass the same task **least frequently**. The boys in *Cluster*5 had the **most** video clips imposed, though the amount was **not significant**; they also tended to fail to pass the same task the **most frequently**.

Regarding the *overlay* video clips displayed, including **multiple displays** of the same video clip, the *Lower-Performers (Cluster5)* and the boys in *Cluster*2 watched the **most** video clips, while the *Higher-Performers (Cluster3)* watched the **fewest**. Slightly more video clips were imposed on the children in *Cluster*1 compared to *Cluster*3.

In analysing the error bars in Figure 6.12, representing confidence intervals (CIs), we observe that there seem to be **significant differences** between the clusters at level *Multiple_Views_Considered*; these error bars are more dispersed, i.e. they overlap less than those of level *Different_Videos_Displayed*. However, the *'sample'* means are **more reliable** for the **different** *overlay* video clips than for the **multiple views** of the same video clip (narrower error bars).

The second table (Table 6.38) includes the *directed* differences between both levels of factor *WhichPct* for each of the $five$ clusters at the *Gender* level.

Table 6.38: *Directed* difference between
- the percentage of **different** *overlay* video clips displayed to the individual child and
- the percentage of *overlay* video clips displayed when including **multiple displays** of the same video clip
for each of the $five$ clusters at the *Gender* level

Cluster	***directed* difference between Different_Videos_Displayed and Multiple_Views_Considered (in %)**	
	Girls	**Boys**
*Cluster*1 (♀: 19, ♂: 16)	18,01	$18{,}01 + (-2{,}71) = 15{,}3$
*Cluster*2 (♀: 25, ♂: 16)	$18{,}01 + 4{,}02 = 22{,}03$ [a]	$18{,}01 + (-2{,}71) + 4{,}02 + 10{,}96 = 30{,}28$ [b]
*Cluster*3 (♀: 13, ♂: 15)	$18{,}01 + (-7{,}71) = 10{,}3$	$18{,}01 + (-2{,}71) + (-7{,}71) + 4{,}31 = 11{,}9$
*Cluster*4 (♀: 11, ♂: 22)	$18{,}01 + 5{,}3 = 23{,}31$	$18{,}01 + (-2{,}71) + 5{,}3 + (-0{,}85) = 19{,}75$
*Cluster*5 (♀: 12, ♂: 15)	$18{,}01 + 21{,}37 = 39{,}38$	$18{,}01 + (-2{,}71) + 21{,}37 + (-4{,}97) = 31{,}7$

[a] Sum of the values of the effects:
$WhichPctMultiple\_Views\_Considered$,
$WhichPctMultiple\_Views\_Considered : Cluster2$

[b] Sum of the values of the effects:
$WhichPctMultiple\_Views\_Considered$,
$WhichPctMultiple\_Views\_Considered : GenderB$,
$WhichPctMultiple\_Views\_Considered : Cluster2$,
$WhichPctMultiple\_Views\_Considered : Cluster2 : GenderB$

From the data presented in Table 6.38 we conclude that globally, the pupils were enforced to $re-$watch between $10,30\%$ (Girls *Cluster*3) and $39,38\%$ (Girls *Cluster*5) of the *overlay* video clips. It should be noted that we kept the GEOGEBRATAO learning environment as close as possible to the everyday learning environment in which pupils do not remember everything that they studied in class. Thus holidays or weekends might, among other things, cause **multiple displays** of some video clips.

In general, the *Lower-Performers (Cluster*5*)* and the boys in *Cluster*2 were susceptible to watching the **most** *overlay* video clips twice, or even more than two times, compared to the children in the other clusters. The *Higher-Performers (Cluster*3*)*, by contrast, were obliged to $re-$watch the **fewest** video clips, followed by the children in *Cluster*1.

No *'gender gap'* seems to exist in the study of the *overlay* video clips; this can also be seen in Figure 6.12, next to the illustrations of the previous made observations.

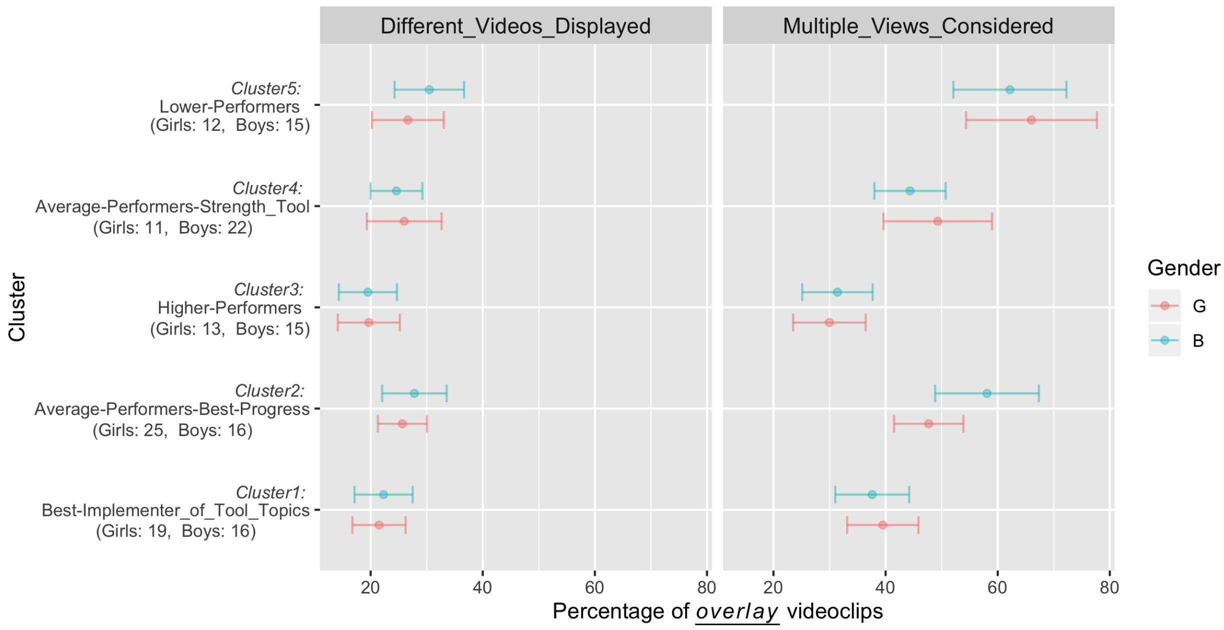


Figure 6.12: **Cluster-based** comparisons regarding the number of *overlay* video clips *imposed* on each individual child in relation to the total number of *overlay* video clips that could be displayed to the child (expressed as a percentage).
These comparisons were made both
- for the total number of **different** *overlay* video clips imposed, and
- for the number of *overlay* video clips watched when including **multiple displays** of the same video clip.
This figure shows 95% CI (confidence interval) bars.
The *percentage* values have been fitted through the Generalized Mixed-Effects Model 6.6.

6.4.3.8 Summary table

Table 6.39: The use of the affordance *'overlay video clips'* on a rating scale from 1 to 5, with 1 representing an affordance that is used much less by the cluster in question than by the remaining clusters and 5 an affordance that is used much more, always considering *Gender* (greater detail in subsubsection 6.4.3.4)

forced use of available affordances	*Cluster*1 **Best-Implementer_of_Tool_Topics**		*Cluster*2 **Average-Performers-Best-Progress**		*Cluster*3 **Higher-Performers**		*Cluster*4 **Average-Performers-Strength_Tool**		*Cluster*5 **Lower-Performers**	
different *overlay* video clips (Table 6.37)	♀: 2	♂: 2	♀: 5	♂: 4	♀: 1	♂: 1	♀: 5	♂: 3	♀: 5	♂: 5
$re-$displays of *overlay* video clips (Table 6.38)	♀: 2	♂: 1	♀: 3	♂: 5	♀: 1	♂: 1	♀: 3	♂: 2	♀: 5	♂: 5

Let us discuss potential cluster characteristics related to the repetition of prior learning assignments.

6.4.3.9 Fitted Generalized Mixed-Effects Model for the *repetition* of prior learning assignments

A further step in adequately reflecting the clusters' characteristics consists of comparing children's percentages of loops and repetitions of prior learning assignments, which are **imposed** in some specific error scenarios. The latter repeat in a manner similar to the *overlay* video clips.

We calculate these percentages as follows:

1. the total number of **different** loops completed compared to the total number of possible loops calculated for the individual child;
2. the number of loops imposed, including repetitions of the same loop, compared to the total number of possible loops calculated for the individual child.

On these calculations, we fit the following **Generalized Mixed-Effects Model** to embody further statistical assumptions concerning the common ways of working with the GEOGEBRATAO tool:

$$\text{glmmPQL\_repeat} \leftarrow glmmPQL(\underbrace{\text{LoopsPct}}_{\text{response variable}} \sim \underbrace{\text{WhichPct * Cluster * Gender}}_{\text{fixed effects}}, \underbrace{\text{random} = \text{list(Children} =\sim 1\text{)}}_{\text{random effects}}, \text{family} = \text{Gamma(link} = \text{‘identity’)}, \text{ data} = \text{datCluster\_repeat)} \quad (6.7)$$

where

glmmPQL_repeat	:	Generalized Linear Mixed Model (**GLMM** model) with multivariate normal random effects, using Penalized Quasi-Likelihood (**PQL**) for cluster-based comparisons based on the percentages of loops imposed / **rep**etition of prior learning assignments performed by each child;
datCluster_repeat	:	data set containing the clusters built and information about the loops and repetitions of prior learning assignments;
LoopsPct	:	variable representing the number of loops completed by each child, compared to the total number of possible loops calculated for the individual child (expressed as a percentage);
WhichPct	:	factor having the levels *Different_Loops_Achieved* (*different redirections into loops imposed to the child*) and *Repeat_Same_Loops_Considered* (*including repetitions of the same loop*).

6.4.3.10 Analysis and interpretation of loops and repetitions of prior learning assignments at *Cluster* level

Once more, the outcome provided by the Analysis of Deviance Table *(Type II tests)* (cf. M.1 : appendix M) suggests a highly significant effect on the variable *LoopsPct*

- of the factor *Cluster* ('***');
- of the two-factor interaction '*WhichPct* by *Cluster*' ('***');
- of the $three$-factor interaction, i.e. of two-factor interaction '*WhichPct* by *Cluster*' varying across levels of the factor *Gender* (p-value $\simeq 0,003$) ('**').

Hence, the factor *Cluster* has a significant effect on the number of loops imposed on the child. This enables us to determine further cluster-specific characteristics regarding the children's ways of progressing through the GEOGEBRATAO tool. Tables 6.40 and 6.41 and Figure 6.13, which are similar to our previous tables and figures, show these results.

Table 6.40: *Directed* differences between *Cluster*1 (base level of factor *Cluster*) and every other cluster both
- for the percentage of **different** loops completed by each individual child and
- for the percentage of loops imposed, including **repetitions** of the same loop

between the clusters	***directed* difference for** **Different_Loops_Achieved**	**Repeat_Same_Loops_Considered**
1 *and* 2 *(Girls)*	5,44	$5{,}44 + 11{,}79 = 17{,}23$ [a]
1 *and* 2 *(Boys)*	$5{,}44 + (-0{,}53) = 4{,}91$	$5{,}44 + (-0{,}53) + 11{,}79 + 1{,}3 = 18$ [b]
1 *and* 3 *(Girls)*	$-\,2{,}55$	$(-2{,}55) + (-3{,}3) = -5{,}85$
1 *and* 3 *(Boys)*	$(-2{,}55) + (-0{,}9) = -3{,}45$	$(-2{,}55) + (-0{,}9) + (-3{,}3) + 3{,}3 = -3{,}45$
1 *and* 4 *(Girls)*	1,39	$1{,}39 + (-3{,}3) = -1{,}91$
1 *and* 4 *(Boys)*	$1{,}39 + 1{,}29 = 2{,}68$	$1{,}39 + 1{,}29 + (-3{,}3) + 14{,}82 = 14{,}2$
1 *and* 5 *(Girls)*	13,33	$13{,}33 + 4{,}6 = 17{,}93$
1 *and* 5 *(Boys)*	$13{,}33 + (-11{,}86) = 1{,}47$	$13{,}33 + (-11{,}86) + 4{,}6 + 6{,}87 = 12{,}94$

[a] Sum of the values of the effects:
$Cluster2$,
$WhichPctRepeat\_Same\_Loops\_Considered : Cluster2$

[b] Sum of the values of the effects:
$Cluster2$,
$Cluster2 : GenderB$,
$WhichPctRepeat\_Same\_Loops\_Considered : Cluster2$,
$WhichPctRepeat\_Same\_Loops\_Considered : Cluster2 : GenderB$

*Cluster*1 = Best-Implementer_of_Tool_Topics
*Cluster*2 = Average-Performers-Best-Progress
*Cluster*3 = Higher-Performers
*Cluster*4 = Average-Performers-Strength_Tool
*Cluster*5 = Lower-Performers

By considering the **different** loops achieved, Table 6.40 shows that the *Higher-Performers (Cluster*3*)* of both genders had the **fewest** loops imposed; this is a logical result for a *Higher-Performer* cluster. This cluster is once again followed by the cluster containing the *Best-Implementer* of the GEOGEBRATAO tool topics *(Cluster*1*)*. The girls in *Cluster*5 were directed to complete the **most** loops; they tended to fail the same task **most repeatedly**. In contrast to the *overlay* video clips imposed, the boys in *Cluster*5 did not complete an unusually high number of loops.

Regarding the loops completed when including **repeated** loops,

- the pupils in *Cluster*3 and *Cluster*1 and the girls in *Cluster*4

  completed **significantly fewer** loops than

- the pupils in *Cluster*2 and *Cluster*5 and the boys in *Cluster*4.

The associated error bars, representing confidence intervals (CIs), **hardly overlap**. Hence, the evidence for true differences is relatively strong.

As in the previous analyses, Table 6.41 includes the *directed* differences between both levels of the factor *WhichPct* for each of the $five$ clusters at the *Gender* level.

Table 6.41: *Directed* difference between
- the percentage of **different** loops completed by each individual child and
- the percentage of loops imposed, including **repeated loops**,
for each of the $five$ clusters at the *Gender* level

Cluster	***directed* difference between Different_Loops_Achieved and Repeat_Same_Loops_Considered (in %)**	
	Girls	**Boys**
*Cluster*1 (♀: 19, ♂: 16)	3,3	$3{,}3 + (-3{,}3) = 0$
*Cluster*2 (♀: 25, ♂: 16)	$3{,}3 + 11{,}79 = 15{,}09$ [a]	$3{,}3 + (-3{,}3) + 11{,}79 + 1{,}3 = 13{,}09$ [b]
*Cluster*3 (♀: 13, ♂: 15)	$3{,}3 + (-3{,}3) = 0$	$3{,}3 + (-3{,}3) + (-3{,}3) + 3{,}3 = 0$
*Cluster*4 (♀: 11, ♂: 22)	$3{,}3 + (-3{,}3) = 0$	$3{,}3 + (-3{,}3) + (-3{,}3) + 14{,}82 = 11{,}52$
*Cluster*5 (♀: 12, ♂: 15)	$3{,}3 + 4{,}6 = 7{,}9$	$3{,}3 + (-3{,}3) + 4{,}6 + 6{,}87 = 11{,}47$

[a] Sum of the values of the effects:
$WhichPctRepeat\_Same\_Loops\_Considered$,
$WhichPctRepeat\_Same\_Loops\_Considered : Cluster2$

[b] Sum of the values of the effects:
$WhichPctRepeat\_Same\_Loops\_Considered$,
$WhichPctRepeat\_Same\_Loops\_Considered : GenderB$,
$WhichPctRepeat\_Same\_Loops\_Considered : Cluster2$,
$WhichPctRepeat\_Same\_Loops\_Considered : Cluster2 : GenderB$

According to the data presented in Table 6.41, the children in *Cluster*3, the boys in *Cluster*1, and the girls in *Cluster*4 tended to progress through the sequence of learning assignments **without** repeating loops.

On the other hand, the children in *Cluster*2 and the boys in *Cluster*4 and *Cluster*5 tended to $re-$execute between 11% and 16% of the loops. It should be recalled that holidays or weekends could cause the recurrence of some loops.

A *'gender gap'*, clearly visible in Figure 6.13, seems to exist in *Cluster*4.

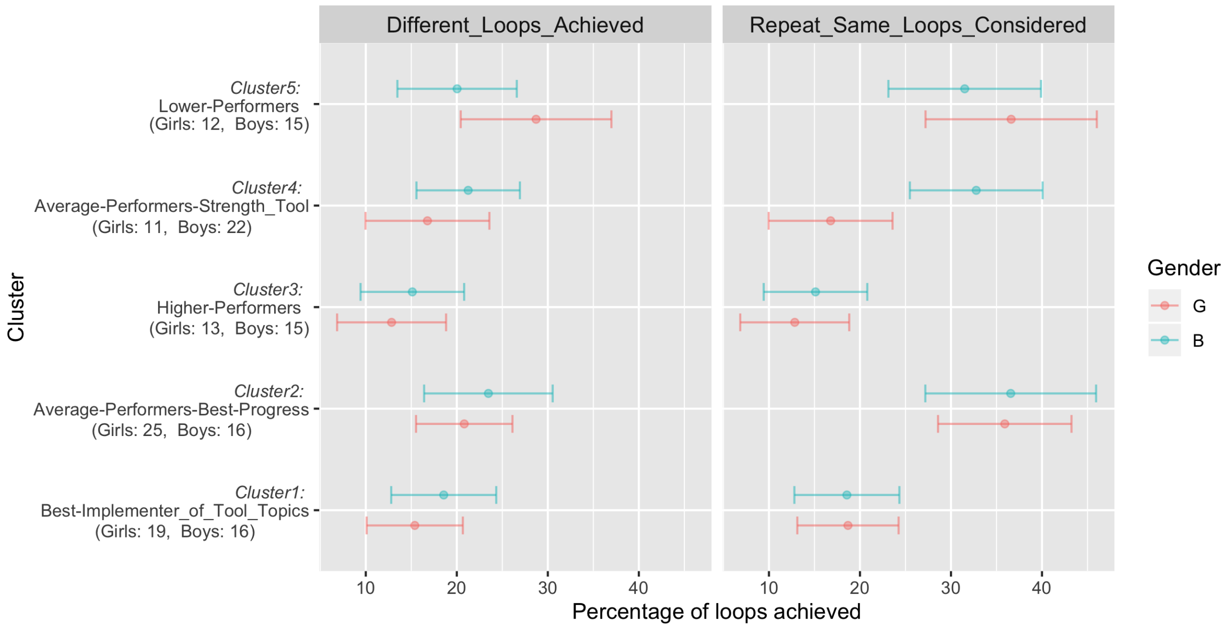


Figure 6.13: **Cluster-based** comparisons regarding the number of loops and repetitions of prior learning assignments performed on an *enforced* basis by each child in relation to the total number of possible loops calculated for the individual child (expressed as a percentage).
These comparisons were made both
- for the total number of **distinct** loops completed by each individual child, and
- for the number of loops imposed when we consider that the same loop can be executed **several times**.
This figure shows 95% CI (confidence interval) bars.
The *percentage* values have been fitted through the Generalized Mixed-Effects Model 6.7.

6.4.3.11 Summary table

Table 6.42: The use of the affordance *'repetition loops'* on a rating scale from 1 to 5, with 1 representing an affordance that is used much less by the cluster in question than by the remaining clusters and 5 an affordance that is used much more, always considering *Gender* (greater detail in subsubsection 6.4.3.4)

forced use of available affordances	***Cluster*1 Best-Implementer_of_Tool_Topics**	***Cluster*2 Average-Performers-Best-Progress**	***Cluster*3 Higher-Performers**	***Cluster*4 Average-Performers-Strength_Tool**	***Cluster*5 Lower-Performers**
different *'repetition loops'* (Table 6.40)	♀: 1 ♂: 3	♀: 3 ♂: 5	♀: 1 ♂: 1	♀: 2 ♂: 4	♀: 5 ♂: 3
$re-$executings of *'repetition loops'* (Table 6.41)	♀: 2 ♂: 1	♀: 5 ♂: 5	♀: 1 ♂: 1	♀: 1 ♂: 5	♀: 3 ♂: 5

Finally, an analysis to study the frequency of call-the-teacher messages will be performed which particularly concerns weaker pupils. With this information, the clusters' characteristics will be complemented.

6.4.3.12 Fitted Generalized Mixed-Effects Model related to the message to call the teacher

Let us briefly compare the children's percentages of **distinct** learning assignments for which the teacher had to be called, which is a feature **imposed** by the GEOGEBRATAO tool. As explained in subsection 2.2.4, after a child's third incorrect attempt on any task the software locks automatically and displays a message to call the teacher, later called the *'Call-Teacher-Message'*. Only the teacher is able to unlock the tool by entering a password.

To calculate these percentages for each individual child, we compare the total number of **distinct** learning assignments for which the *Call-Teacher-Message* was displayed to the total number of learning assignments performed. In this case, including the same learning assignment **several times** did not produce any reliable results, since this concerned only a few pupils and technical issues could not be excluded.

We once again fit a **Generalized Mixed-Effects Model**:

$$\text{glmmPQL\_mes} \leftarrow glmmPQL(\underbrace{\text{Call\_Teacher}}_{\text{response variable}} \sim \underbrace{\text{Cluster * Gender}}_{\text{fixed effects}}, \underbrace{\text{random} = \text{list}(\ \text{Children} =\sim 1\ )}_{\text{random effects}}, \\ \text{family} = \text{gaussian}(\text{link} = \text{'identity'}),\ \text{data} = \text{datCluster\_mes}\,) \tag{6.8}$$

where

glmmPQL_mes : Generalized Linear Mixed Model (**GLMM** model) with multivariate normal random effects, using Penalized Quasi-Likelihood (**PQL**) for cluster-based comparisons based on the percentages of *Call-Teacher-**Mes**sages* displayed to each child;
datCluster_mes : data set containing the clusters built and information about the *Call-Teacher-Messages*;
Call_Teacher : variable representing the number of time the *Call-Teacher-Message* was displayed to each child compared to the total number of learning assignments performed by the individual child (expressed as a percentage).

6.4.3.13 Analysis and interpretation of the *Call-Teacher-Messages* at *Cluster* level

According to the Analysis of Deviance Table *(Type II tests)*, there is only one highly significant effect on the variable *Call_Teacher*, namely the effect of the factor *Cluster* ($'***'$) (cf. N.1 : appendix N). There is no *Gender* related effect. Further cluster characteristics regarding how children progress through the GEOGEBRATAO tool can therefore be identified using the following table.

Table 6.43: *Directed* differences between *Cluster*1 (base level of the factor *Cluster*) and every other cluster regarding the percentage of **distinct** learning assignments for which the teacher had to be called by the individual child

between the clusters	*directed* difference for Call_Teacher_Message
1 *and* 2 *(Girls)*	1,43
1 *and* 2 *(Boys)*	$1{,}43 + 1{,}33 = 2{,}76$ [a]
1 *and* 3 *(Girls)*	$-\,2{,}01$
1 *and* 3 *(Boys)*	$(-2{,}01) + 1{,}59 = -0{,}42$
1 *and* 4 *(Girls)*	0,93
1 *and* 4 *(Boys)*	$0{,}93 + 0{,}93 = 1{,}86$
1 *and* 5 *(Girls)*	4,77
1 *and* 5 *(Boys)*	$4{,}77 + (-1{,}43) = 3{,}34$

*Cluster*1 = Best-Implementer_of_Tool_Topics
*Cluster*2 = Average-Performers-Best-Progress
*Cluster*3 = Higher-Performers
*Cluster*4 = Average-Performers-Strength_Tool
*Cluster*5 = Lower-Performers

[a] Sum of the values of the effects:
$Cluster2$,
$Cluster2 : GenderB$

Table 6.43 shows that the *Higher-Performers (Cluster*3*)* of both genders received the *Call-Teacher-Message* **least frequently**, followed by the *Best-Implementer* of the GEOGEBRATAO tool topics *(Cluster*1*)*. Obviously, the *Lower-Performers (Cluster*5*)* had to ask their teacher for help on the **most** learning assignments compared to their peers; they tended to fail the same task **most frequently**, i.e. at least

- 15 times for accomplishments through actions in the GeoGebra frame;
- 6 times for accomplishments through choices on Likert scales or in drop-down menus.

The pupils in *Cluster*4 needed to call their teacher somewhat less often than those in *Cluster*2. No *'gender gap'* seems to exist regarding the *Call-Teacher-Messages* criterion in any cluster.

Once more, it is especially useful to examine the degree of overlap between confidence intervals (CIs) in Figure 6.14. We observe, among other things, that the error bars associated with *Cluster*3 and *Cluster*5 do not overlap; those associated with *Cluster*1 and *Cluster*5 hardly overlap, just like those associated with *Cluster*3 and *Cluster*2. This demonstrates strong evidence for true differences between the respective clusters.

6.4.3.14 Summary table

Table 6.44: The use of the affordance *'Call-Teacher-Messages'* on a rating scale from 1 to 5, with 1 representing an affordance that is used much less by the cluster in question than by the remaining clusters and 5 an affordance that is used much more, always considering *Gender* (greater detail in subsubsection 6.4.3.4)

forced use of available affordance	***Cluster*1 Best-Implementer_ of_Tool_Topics**		***Cluster*2 Average-Performers- Best-Progress**		***Cluster*3 Higher- Performers**		***Cluster*4 Average-Performers- Strength_Tool**		***Cluster*5 Lower- Performers**	
displays of *Call-Teacher-Messages* (Table 6.43)	♀: 2	♂: 1	♀: 3	♂: 5	♀: 1	♂: 1	♀: 3	♂: 4	♀: 5	♂: 5

To summarize our findings, the number of

1. *short* video clips clicked on voluntarily,
2. *overlay* video clips displayed involuntarily,
3. *'repetition loops'* imposed,
4. *Call-Teacher-Messages* received

will be used to draw conclusions about each cluster's overall characteristics and methods of interacting with the GEOGEBRATAO tool.

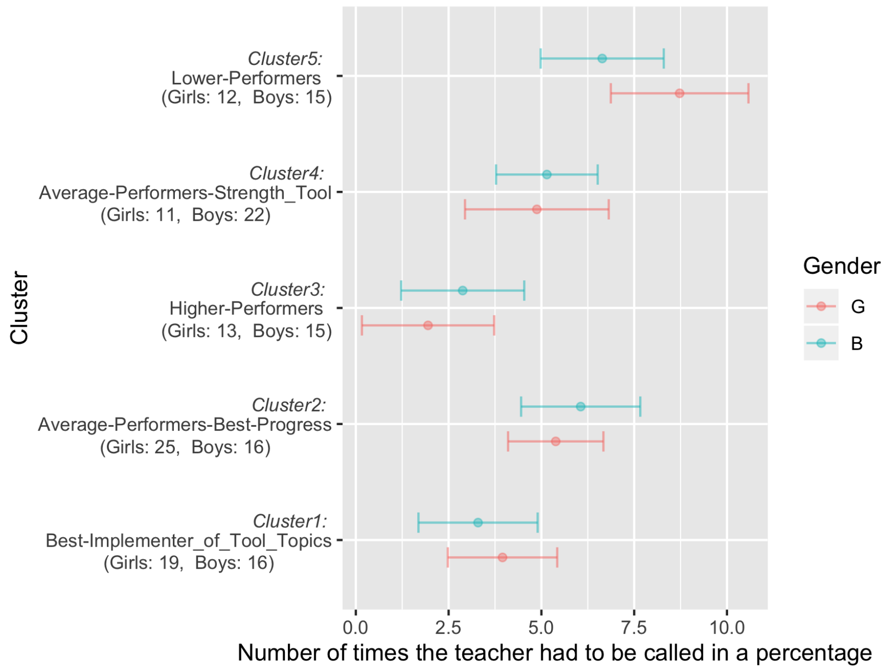


Figure 6.14: **Cluster-based** comparisons regarding the total number of **distinct** learning assignments for which the teacher had to be called by each child, **imposed** by the GEOGEBRATAO tool, in relation to the total number of learning assignments performed by the individual child (expressed as a percentage).
This figure shows 95% CI (confidence interval) bars.
The *percentage* values have been fitted through the Generalized Mixed-Effects Model 6.8.

6.4.4 Conclusion on hypothesis two *(in answer to research question* $RQ3$ *(pg. 6))*

First of all, we establish a summary table (cf. Table 6.45), composed of the Tables 6.34, 6.39, 6.42 and 6.44, that clearly illustrates some common ways of working with our dynamic geometry tool. Therefore, as already mentioned, we ranked the voluntary and forced use of some affordances provided by the GEOGEBRATAO tool on a rating scale from one to $five$, with

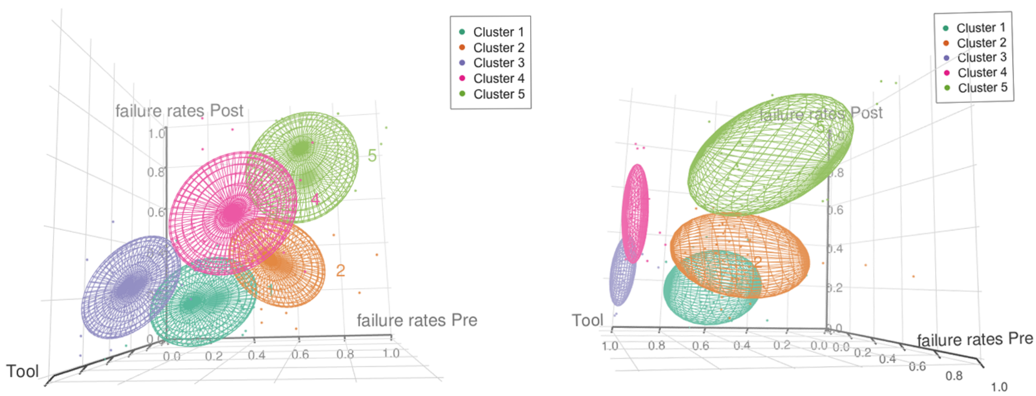
- one representing an affordance that is **used much less**;
- $five$ representing an affordance that is **used much more**

by the cluster in question than by other clusters.

Based on this table, we reflect some cluster characteristics describing its pupils' common ways of working with the GEOGEBRATAO tool and explaining their progress through the sequence of exploratory learning assignments. All the comparisons are made between clusters and include gender information.

Table 6.45: Voluntary and forced use of affordances on a rating scale from 1 to 5, with 1 representing affordances that are used much less by the cluster in question than by the remaining clusters and 5 affordances that are used much more, always considering *Gender* (greater detail in subsubsection 6.4.3.4)

voluntary (*) and forced use of available affordances	*Cluster*1 **Best-Implementer_of_Tool_Topics**		*Cluster*2 **Average-Performers-Best-Progress**		*Cluster*3 **Higher-Performers**		*Cluster*4 **Average-Performers-Strength_Tool**		*Cluster*5 **Lower-Performers**	
(*) different *short* video clips (Table 6.32)	♀: 1	♂: 2	♀: 1	♂: 1	♀: 4	♂: 4	♀: 5	♂: 5	♀: 1	♂: 3
(*) $re-$watchings of *short* video clips (Table 6.33)	♀: 2	♂: 2	♀: 1	♂: 4	♀: 2	♂: 3	♀: 4	♂: 5	♀: 5	♂: 1
different *overlay* video clips (Table 6.37)	♀: 2	♂: 2	♀: 5	♂: 4	♀: 1	♂: 1	♀: 5	♂: 3	♀: 5	♂: 5
$re-$displays of *overlay* video clips (Table 6.38)	♀: 2	♂: 1	♀: 3	♂: 5	♀: 1	♂: 1	♀: 3	♂: 2	♀: 5	♂: 5
different *'repetition loops'* (Table 6.40)	♀: 1	♂: 3	♀: 3	♂: 5	♀: 1	♂: 1	♀: 2	♂: 4	♀: 5	♂: 3
$re-$executings of *'repetition loops'* (Table 6.41)	♀: 2	♂: 1	♀: 5	♂: 5	♀: 1	♂: 1	♀: 1	♂: 5	♀: 3	♂: 5
displays of *Call-Teacher-Messages* (Table 6.43)	♀: 2	♂: 1	♀: 3	♂: 5	♀: 1	♂: 1	♀: 3	♂: 4	♀: 5	♂: 5



(View 1) (View 2)

Figure 6.15: **Reminder** of our $three$-dimensional visualizations of the performance groups nested within the test group; the **restricted** binary post-test data is used

WHAT DOES THE DATA TELL US?

The *Higher-Performers (Cluster3)* **need the least** imposed assistance in some specific error scenarios, i.e. they experienced the **fewest** *overlay* video clips, *'repetition loops'* and *Call-Teacher-Messages*. In contrast, they voluntarily watch, or rather click on nearly the **most** distinct *short* video clips. It seems as though these children take **full advantage** of this optional instant aid and therefore require no supplementary help imposed by the tool. They also $re-$watch an average number of *short* video clips.

WHAT MIGHT THIS MEAN?

It seems as though *Higher-Performers* are able to concentrate and focus on the important information being given by the GEOGEBRATAO tool and consult affordances available on a voluntary basis to fill gaps on their own.

WHAT DOES THE DATA TELL US?

The *Best-Implementer_of_Tool_Topics (Cluster1)* need the help imposed in some specific error scenarios on an **enforced** basis **slightly more frequently** than the pupils in *Cluster3*, but consult distinct *short* video clips **less frequently**. If they used the optional instant aid more often, then they would likely be forced to watch even fewer *overlay* video clips, repeat fewer loops or call their teacher less often.

WHAT MIGHT THIS MEAN?

The *Best-Implementer_of_Tool_Topics* seem to work quite **slowly** while **sustaining attention** on the task demands. In fact, they do **not complete** our sequence of activities, although they use / have to use the available affordances listed in Table 6.45 less than most other clusters. Interestingly, they are the **highest (best) performing** in terms of converting topics they have explored themselves within the GEOGEBRATAO phase into similar paper-and-pencil exercises.

WHAT DOES THE DATA TELL US?

Of those who did **not complete** the sequency of dynamic learning assignments, the *Average-Performers-Strength_Tool (Cluster4)* came **closest** to doing so, despite using / having to use the available affordances (see in Table 6.45) **quite often** compared to the other clusters. In particular, they consulted the **most** *short* video clips voluntarily, more than the children in *Cluster3*. They still have to watch a **fairly high** number of *overlay* video clips, especially the girls. The boys, on the other hand, had to execute nearly the **most** *'repetition loops'*. Moreover, both genders had to call their teacher an average number of times.

WHAT MIGHT THIS MEAN?

It appears that the *Average-Performers-Strength_Tool* move too quickly through the sequence of exploratory learning assignments; perhaps they just click on the video clips without really watching them or do not focus on repeated information when completing a loop. And so, they do **not** acquire **enough knowledge** nor do they learn the different concepts **deeply enough** to perform equally well as the pupils in *Cluster3* and *Cluster1*.

WHAT DOES THE DATA TELL US?

The *Lower-Performers (Cluster5)* **need the most** imposed help in some specific error scenarios. Regarding the different *short* video clips, the girls tend to click on roughly the **same number** of video clips as the girls in *Cluster1*, and the boys consult **somewhat more** video clips than the boys in *Cluster1*. Regarding $re-$watching *short* video clips, a ***'gender gap'*** seems to exist; the girls tend to consult some video clips more often than the remaining test-group children and the boys tend to watch the video clips only once. This may be due to a common observation of *"girls' lack of self-confidence in their own ability in science and mathematics" (OECD, 2015, pg. 63)*.

WHAT MIGHT THIS MEAN?

It seems as if the *Lower-Performer* girls often ***'forget'*** to watch distinct *short* video clips. However, if they consult one, they often tend to $re-$watch it. This is not the case for the boys; they may **not be aware** of the video clips' presence in the tool or utility, or even of their own knowledge gaps. Language does not cause difficulties since only silent video clips are used to present (visual) explanations, and the written text used in the clips is extremely limited.

WHAT DOES THE DATA TELL US?

The *Average-Performers-Best-Progress (Cluster2)* have **similar** ways of working as the children in *Cluster5*. The former children just use / have to use the available affordances listed in Table 6.45 **slightly less / less often** than the latter.

WHAT MIGHT THIS MEAN?

However, the **big difference** is that the *Average-Performers-Best-Progress* made *good / great* **progress** from the pre- to post-test, most likely because they **thoroughly** used the affordances imposed by the GEOGEBRATAO tool. These affordances may help the children to **focus their attention** on a specific task or a relevant concept not yet understood. It seems that these pupils are not able to independently realize when it would be beneficial to use optional instant aid, i.e. to watch *short* video clips voluntarily. They need to **be guided** somewhat by our tool in order to make good progress.

In summary, it seems clear that the *Higher-Performers (Cluster3)* and the *Best-Implementer_of_Tool_Topics (Cluster1)* are able to work independently with the GEOGEBRATAO tool and to learn autonomously.

The *Average-Performers-Strength_Tool (Cluster4)* are able to work independently as well, but their teachers should check for understanding frequently, slow the children down, to prevent students from simply skimming over the concepts. With more guidance, they will likely better focus their attention and deepen and broaden their understanding.

The *Average-Performers-Best-Progress (Cluster2)* and the *Lower-Performers (Cluster5)* probably need the most help from their teacher. In particular, they should regularly be reminded to consult the *short* video clips and thereby be made aware of their importance. In the event of unavoidable intervention, the teachers should follow the recommendations described in subsection 2.2.10 for using our tool in class to develop children's ability to learn autonomously. Perhaps the self-confidence of the pupils in *Cluster2* should be boosted by the teacher, since these pupils are able to make good progress (cf. progress from pre- to post-test). A lack of self-confidence when working on the computer may be responsible for frequently using the imposed supplementary support in the GEOGEBRATAO tool.

It should be noted that all these observations are solely based on numerical data that we try to interpret, i.e. on data collected within each exploratory learning assignment which is a combination of the child inputs, the result(s) and some metrics related to the assessment. Thereby, the human dimension gets somewhat lost. This implies, as already mentioned, that it would be interesting to study children's physical interaction and engagement while using the GEOGEBRATAO tool in a later project. This would enable us to obtain additional precious information relevant to our outstanding issues.

In this chapter, through a creative but time-consuming process, answers to the research questions were best developed; we tried to draw the most accurate and relevant conclusions from the mass of collected data within the sequence of exploratory learning assignments, i.e. through the logs of the TAO platform, and from the pre- and post-test. However, what might actually be the possible barriers that somewhat disturbed during the project implementation?

Chapter 7

Barriers to GEOGEBRATAO implementation

Both theoretical foundations of learning environments and our own experience have taught us to expect a predictable set of barriers when planning implementation of a multimedia project in schools. We have found that our experiences are not unusual and that the following barriers can also be observed in the GEOGEBRATAO project. They fall into $three$ categories:

1. technology barriers;
2. organizational barriers;
3. human-related barriers.

Technology barriers comprise the largest subset and they are the most difficult to eliminate in our research, followed by the organizational barriers. Human-related barriers could, among other things, be addressed by offering teachers more specific **skills training to real classroom situations** and a more supportive school environment. This could greatly improve the chances of a project's success.

Let us list and briefly examine the observed problems by category.

7.1 Technology barriers

As already mentioned, in our study, we had no control over the schools' equipment and ICT infrastructure. Therefore, the schools were equipped quite differently in terms of quality and availability.

All the schools were equipped with sufficient machines to provide children with individual regular access to either computers or laptops or iPads. However, some key differences between these machines led to technology problems:

- operating systems;
- computer performance;
- the installation of computer updates;
- internet connection;
- screen type (touch screen or not);
- screen resolution;
- mouse sensitivity.

First of all, it should be noted that operating systems and their updates significantly impact the operation ability of the devices. Depending on the first $four$ above differences, the learning assignments of the GEOGEBRATAO tool and the video clips were loaded more or less quickly. This might have an impact on the children's motivation and patience. Some impatient children might have tried to perform the activities without the benefit of instant aid. Other children might have watched on a neighbour's or classmate's computer screen while waiting for their next activity to load. Furthermore, these differences might have impacted the 500 available minutes dedicated to the project. These assumptions, though sometimes observed in field notes during participation in GEOGEBRATAO lessons, can be neither verified nor confirmed.

Concerning differences related to the type of computer screen, we noticed that on some touch-screens, the mouseover effect did not work properly: when the child moved the mouse pointer over the French flag icon to switch the language, the GEOGEBRATAO tool did not switch back to the German version afterwards. This is once again information obtained from field notes during participation in GEOGEBRATAO lessons.

Moreover, differences related to screen resolution caused difficulties for some children. In fact, when programming the tool, we allowed the children a certain degree of imprecision in their actions in the GeoGebra frame. These imprecisions correspond more or less to $1,5\,mm$ in paper activities. However, depending on the screen resolution, the GEOGEBRATAO tool was more or less sensitive when the child had to draw a point at a predetermined place in the GeoGebra frame, resulting in some pupils having more room to draw a point 'correctly' than others. This obviously had an impact on the degree of difficulty of some learning assignments.

Finally difficulties arose from some pupil confusion between single clicking versus double clicking on buttons (for moving to the next item) and on points in the GeoGebra area. Double clicking on buttons made pupils reach the maximum number of consecutive errors more quickly. On the other side, in more complex programmed GeoGebra activities, double clicking on points had an impact on the proper functioning of the GEOGEBRATAO tool. It may be that mouse sensitivity was also a factor.

7.2 Organizational barriers

The planning of GEOGEBRATAO lessons was not always easy. Details to be planned included:

- timetabling of the lessons;
- access to computer rooms, laptop carriages or iPads;
- the presence of the teacher (some teachers worked part-time);
- the presence or absence of some children;
- the amount of time available respecting the constraints of the national school curriculum;
- catch-up when a child was absent from school.

In addition, the guidelines of the GEOGEBRATAO project added the constraints of the pre- and post-test dates and of the 500 minute working period, which were identical for **all** the classes in order to create equitable test conditions in real classroom situations. This overall package of constraints was therefore sometimes a rather complex issue.

We should also note that some barriers were related to the Luxembourg school system, which is adapted to the growing societal diversity; some children would normally have attended support classes during the GEOGEBRATAO lessons. Arrangements had to be made with support teachers to ensure that all the children in a class could participate in our project, even the lowest performers.

7.3 Human- and teacher-related barriers

A range of teacher-related barriers were described in subsection 4.1.2 and were also observed over the course of the GEOGEBRATAO project:

- the sustainability of ICT in the teacher's school;
- the teacher's understanding of our research goals;
- the teacher's behavior intentions when using the GEOGEBRATAO tool in their class;
- the teacher's active involvement in our project;
- the teacher's confidence and competence in using the GEOGEBRATAO tool;
- the *"teachers' attitudes and beliefs regarding* ICT *in teaching and learning" (Eickelmann and Vennemann, 2017, pg. 733)*, *"teachers' thinking and self-determination about* ICT *and pedagogy" (Blackberry and Woods, 2014)*;
- the necessary technical support provided to the teacher by the school IT manager.

Most of these barriers are evident from information obtained while participating in GEOGEBRATAO lessons. The majority of teachers were not used to simply supervising their children during computer-based activities or assuming the role of **mediator** between the child and the computer software, i.e. between the child and the 'new' elementary geometric concepts explored by the GEOGEBRATAO tool. They usually limited themselves to using the computer as a complement to their ordinary teaching method or to do internet research with their pupils. Therefore, they seemed to somewhat lack self-confidence when ever children struggled with geometry concepts while working with the GEOGEBRATAO tool. This might have impacted our research goals, possibly through too hasty teacher interventions.

Although we gave two personal user IDs to each teacher, we noticed that only $three$ teachers out of *twelve* tested our tool by themselves before or even during the official GEOGEBRATAO testing phase. This was observed through the logs of the TAO platform. Relatedly, only two teachers out of *twelve* filled in the personal questionnaire about the use, viability and perceived effectiveness of the GEOGEBRATAO tool in their class. Therefore, it was not possible to get an overall view of the teachers' personal views of the project.

On the one hand, we recommend to further develop teachers' Technological Pedagogical Content Knowledge (TPCK). Therefore, researchers and teacher trainers must even better understand how teachers use ICT in classroom practice. Often teachers have *"good basic* ICT *skills, but lack skills to integrate* ICT *into education, due to a lack of technological content knowledge (TCK) and technological pedagogical knowledge (TPK)" (Helppolainen and Aksela, 2015, pg. 783)*. They need ideas that can be directly applied to classroom practice. An analysis shows among others:

> *"how technology incorporation during initial teacher training programmes, in particular GeoGebra, shapes the development of prospective teachers' TPCK" (Dockendorff and Solar, 2018, pg. 17).*

On the other hand, offering teachers more specific skills training for real classroom situations and a more supportive technical environment could enhance their motivation, confidence and competence using a computer software in class as well as their opinions and self-determination about ICT and pedagogy.

7.4 iPad-related barriers

iPad-related barriers obviously belong to technology barriers, but since they only concern two of our test classes ($XC05$, $XC06$), we prefer to devote a separate section to this topic.

The use of iPads was an interesting experience, but connected to another challenge. The children had to zoom in and out on the working screen to be able to move and draw objects (points, lines, segments) in the GeoGebra area. At the same time, they had to be able to locate their position and move within the GEOGEBRATAO tool window to find the assignment's objective, teaching instructions and instant aid when they zoomed out.

These children could not effectively imitate movements in the GeoGebra area while watching a *short* video clip, an advantage provided to the other children in the test classes. Furthermore, the pupils using iPads had to constantly keep in mind that instant aid was available because it did not appear all the time on the iPad screen, another disadvantage compared to the remaining test pupils.

To enhance iPad classes' working conditions, we provided them with touch-screen pencils for more precise use on the screens. This gave the children similar functionality as a mouse would have.

One of the iPad classes had another disadvantage constituting more of an organizational barrier. As one of the largest classes participating in the project, there weren't enough iPads for each child to have individual access. Therefore, the class was regularly split and we participated in most of the GEOGEBRATAO lessons as mediators between the child and the 'new' elementary geometric concepts to be explored by using our tool. This enabled us to make some ethnographic observations concerning the use of our tool in class and the usability of this specific technical equipment.

As we see in Table 6.31, for example:

- The pupils in class $XC05$ are almost equally distributed among the $five$ performance clusters. Only one child belongs to the lower performing children.
- In class $XC06$, two pupils belong to the lower performing children. The remaining children belong either to the children in *Average-Performers-Best-Progress (Cluster*2*)* or to those in *Best-Implementer_of_Tool_Topics (Cluster*1*)*.

So, the use of iPads was not a determining factor on the results of our statistical analysis. With the provided help, i.e. the touch-screen pencils and our assistance in class, the pupils managed to overcome these supplementary barriers.

Of course, our list of barriers is not exhaustive. These are the main barriers we, as well as the teachers and even the pupils, encountered during the project. Some of the barriers were obvious and familiar to us, while others were more unexpected, such as the screen resolution problem.

Finally, based on all the research findings, analysis and our experience, we will skip to the summary conclusion of this PhD thesis.

Chapter 8

Conclusion

In this thesis we presented a simple classroom procedure based on a *booster* computer software, i.e. the GEOGEBRATAO tool, that largely fulfills the role of a *twin* teacher, thereby enhancing student-centered learning and opportunities for getting teachers relieved of some classroom duties. Once some technology challenges are addressed and some pedagogical adaptations of different exploratory learning assignments are achieved, this tool will be *Easy-to-Use*, particularly in blended learning environments. It can be regarded

1. as an alternative to the traditional didactic teaching methods in blended learning environments;
2. as a tool that can be used in combination with traditional didactic teaching methods;
3. as a tool allowing children to continue learning outside the classroom, i.e. as a tool for onsite and remote learning.

On the one hand *(*$RQ1$ *and* $RQ2$ *(pg. 6))*, this procedure's effectiveness has been measured by comparing children who used the GEOGEBRATAO tool to learn 'new' geometric concepts autonomously with those who followed a traditional paper-and-pencil geometry course. On the other hand *(*$RQ3$*)*, according to a thorough analysis of the pupils' ways of working, the challenge of creating a blended learning environment through the use of this student-centered tool has been achieved. We have identified some commonalities and differences in ways of working with the tool, grouped within $five$ clusters based on similar performance on $three$ different instruments (the pre- and post-test, the tool itself) developed specifically to conduct this research.

For the purposes of this study, we mainly analysed the pupils' inputs within each exploratory learning assignment, thereby hiding some interesting and useful information about the *Human Computer Interaction*. For example, an eye tracker could help researchers find out exactly where a child is looking when using the GEOGEBRATAO tool, i.e. where does the child gaze move around on the working screen of the tool; this could help determine which help modus being offered is most effective and whether or not the child is interacting with what we want them to. (McKenzie and Vaughan, n.d.) This is an area ideal for future research.

An important aspect to be considered is the easy usability of our platform. As described above, there are diverse ways to implement the GEOGEBRATAO tool, and therefore it can be adapted to traditional geometry tasks without presenting any additional challenges to the pupils, such as a difficult user interface.

To make our tool even more attractive, we could, from time to time, integrate **a kind of quiz competition** into the sequence of dynamic learning assignments, thereby making full use of the TAO assessment platform. The quiz questions would relate to the last 15 or 20 assignments the child has explored. This would enable the teachers and researchers to get data-driven, holistic insight into each child's progress and adapt the tool to evolve with the children. Quiz-based tools are an innovative teaching approach that improve pupils' engagement and participation in class (Malandrino et al., 2014).

Furthermore, it would be useful to integrate a system into the GEOGEBRATAO tool that classifies the children working with the tool into **our $five$ performance-based clusters**. It seems that some children (Cluster4) tended to move too quickly through the sequence of learning assignments; these children should be slowed down by the tool. Other children's (Cluster2) self-confidence might benefit from being boosted by the tool, e.g. by giving a different kind of feedback sentence or motivational message. It should be noted that confidence in one's abilities generally enhances motivation, love of challenge and persistence in the face of difficulty (Dweck, 2002). This classification system should be executed at the beginning of the sequence of dynamic learning assignments, just after the first few completed activities, and could be re-evaluated later.

Finally, remote learning is a rising phenomenon; the learning environment no longer needs to be exclusively within a classroom, particularly in difficult circumstances. The GEOGEBRATAO tool is an opportunity to ensure that pupils get the explanations they need wherever they are. However, a small **immediate dialogue box** should be implemented into the tool that allows the child to communicate with their teacher when they are in need; the teacher, for their part could see which activity the child is performing when the message is sent. The video clips included in the GEOGEBRATAO tool are beneficial for active and remote learning (Lipomi, 2020).

We suggest **combining onsite (classroom) and remote teaching when using our tool**, i.e. teaching and learning in a kind of blended learning environment, in which the teachers must thoroughly understand the goals, promise and limitations (discipline, careful planning) of our student-centered software; we expect to see continued growth in these areas. Alternatively, a **chat bot system**, referring to a chatting robot (Dahiya, 2017; Winkler and Söllner, 2018), could also be developed and integrated into the tool for immediate help.

A valuable focus of future research might be to investigate how our *booster* computer software acting as a *twin* teacher may broaden the range of teachers who might become interested in using a dynamic geometry student-centered software in blended learning environments.

Future Work: We plan to submit a CORE Project (the central program of The Luxembourg National Research Fund (FNR)) through which the impact of $three$ usability dimensions, i.e. the social (Soc), technological (Tech) and pedagogical (Ped) dimension, on interaction and learning processes with the new technology should be analysed.

A second objective of this CORE Project should target the level of teachers' technology acceptance, by using the Technology Acceptance Model (TAM) to analyse and explain teachers' attitudes and behavior regarding digital learning technologies.

Bibliography

Ahmed M. Abdulla and Mark A. Runco. Who Funds the Future? Federal Funding Support for Creativity and 21st Century Skills. *Business Creativity and the Creative Economy*, 4:1–7, Sept. 2018. ISSN 2374-1414, 2374-1422. doi: 10.18536/bcce.2018.10.8.1.01. ICSC Press, International Center for Studies in Creativity, publisher. Buffalo State, The State University of New York.

Daniel Adler, Duncan Murdoch, and others. rgl: 3D Visualization Using OpenGL, Apr. 2020. URL `https://CRAN.R-project.org/package=rgl`. R package version 0.100.54.

Martha W. Alibali. Does Visual Scaffolding Facilitate Students' Mathematics Learning? Evidence From Early Algebra. Institute of Education Sciences (IES), 2006. URL `https://ies.ed.gov/funding/grantsearch/details.asp?ID=54`. Accessed: 2020-02-06.

Douglas G. Altman and J. Martin Bland. Statistics Notes - Standard deviations and standard errors. *The British Medical Journal (BMJ, `https://thebmj-frontend.bmj.com`)*, 331 (7521):903–904, Oct. 2005. ISSN 0959-8146. doi: 10.1136/bmj.331.7521.903. British Medical Journal Publishing Group, publisher.

Panagiotis Angelopoulos, Alexandros Balatsoukas, and Adina Nistor. The Use of 3D Technologies to Support Computational Thinking in STEM Education. In Michail Kalogiannakis and Stamatios Papadakis, editors, *Handbook of Research on Tools for Teaching Computational Thinking in P-12 Education*, pages 425–459. IGI Global Publisher of Timely Knowledge, Jun. 2020. ISBN 978-1-7998-4576-8, 978-1-7998-4577-5. doi: 10.4018/978-1-7998-4576-8.ch017.

Richard C. Atkinson and Richard M. Shiffrin. Human Memory: A Proposed System and its Control Processes. volume 2 of *The Psychology of Learning and Motivation*, pages 89–195. Academic Press, Elsevier ScienceDirect, 4 edition, 1968. ISBN 978-0-08-086353-5. doi: 10.1016/S0079-7421(08)60422-3. URL `https://www.sciencedirect.com/science/article/pii/S0079742108604223`. ISSN 0079-7421.

R. Harald Baayen, Douglas J. Davidson, and Douglas M. Bates. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4): 390–412, Nov. 2008. ISSN 0749-596X. doi: 10.1016/j.jml.2007.12.005. Elsevier B.V. ScienceDirect, publisher. Special Issue: Emerging Data Analysis.

Kamariah Abu Bakar, Ahmad Fauzi Mohd Ayub, Rohani Ahmad Tarmizi, and Wong Su Luan. Effect of teaching mathematics using GeoGebra on students' with dissimilar spatial visualisation. In *American Institute of Physics (AIP) Conference Proceedings*, volume 1682. AIP Publishing LLC, Oct. 2015. doi: 10.1063/1.4932469.

Arthur Bakker, Jantien Smit, and Rupert Wegerif. Scaffolding and dialogic teaching in mathematics education: introduction and review. *ZDM: International Journal on Mathematics Education*, 47(7):1047–1065, Oct. 2015. ISSN 1863-9704. doi: 10.1007/s11858-015-0738-8. Springer Nature Switzerland AG, open access at Springerlink.com, publisher.

Bandicam Screen Recorder. Bandicam Company, n.d. URL https://www.bandicam.com. Accessed: 2020-09-10.

Dale J. Barr, Roger Levy, Christoph Scheepers, and Harry J. Tily. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68 (3):255–278, Apr. 2013. ISSN 0749-596X. doi: 10.1016/j.jml.2012.11.001. Elsevier B.V. ScienceDirect, publisher.

Mary Lee Barton and Clare Heidema. *Teaching Reading in Mathematics*. Mid-continent Research for Education and Learning (McREL), Eisenhower Regional Consortium for Mathematics and Science, Aurora, CO, 2 edition, 2002. ISBN 978-1-893476-04-2. URL https://www.valees.org/wp-content/uploads/2019/08/math_teaching_reading_barton2002_sample_chapters.pdf. A Supplement to Teaching Reading in the Content Areas: If Not Me, Then Who?, Teacher's Manual. Accessed: 2020-10-10.

Battelle for Kids. Our mission is to realize the power and promise of 21st century learning for every student — in early learning, in school, and beyond school — across the country and around the globe. Battelle for Kids (P21 Partnership for 21st Century Learning A Network of Battelle for Kids), n.d. URL https://www.battelleforkids.org/networks/p21. Columbus Office (Ohio), Tucson Office (Arizona). Accessed: 2020-09-10.

Gary Beauchamp. ICT and assessment. In Sarah Younie, Marilyn Leask, and Kevin Burden, editors, *Teaching and Learning with ICT in the Primary School*, chapter 18, pages 210–224. Routledge, Taylor & Francis Group (London and NY), 2 edition, 2014. ISBN 978-1-138-78315-7, 978-1-138-78314-0.

Brian R. Belland. Scaffolding: Definition, Current Debates, and Future Directions. In J. Michael Spector, M. David Merrill, Jan Elen, and M.J. Bishop, editors, *Handbook of Research on Educational Communications and Technology*, pages 505–518. Springer, New York, NY, 4 edition, Jan. 2014. ISBN 978-1-4614-3184-8, 978-1-4614-3185-5. doi: 10.1007/978-1-4614-3185-5_39.

Brian R. Belland. *Instructional Scaffolding in STEM Education, Strategies and Efficacy Evidence*. SpringerOpen, Springer International Publishing AG Switzerland, Utah State University, Logan, UT, USA, 1 edition, 2017. ISBN 978-3-319-02564-3, 978-3-319-02565-0. doi: 10.1007/978-3-319-02565-0.

Brian R. Belland, Kim ChanMin, and Michael J. Hannafin. A Framework for Designing Scaffolds That Improve Motivation and Cognition. *Educational Psychologist*, 48(4):243–270, Oct. 2013. doi: 10.1080/00461520.2013.838920. Routledge Taylor & Francis Online, publisher.

Marilyn Binkley, Ola Erstad, Joan Herman, Senta Raizen, Martin Ripley, May Miller-Ricci, and Mike Rumble. Defining Twenty-First Century Skills. In Patrick Griffin, Barry McGaw, and Esther Care, editors, *Assessment and Teaching of 21st Century Skills*, pages 17–66. Springer, Dordrecht, Netherlands, 2012. ISBN 978-94-007-2323-8, 978-94-007-2324-5. doi: 10.1007/978-94-007-2324-5_2.

Gina Blackberry and Deb Woods. Teachers and pupils incorporated: developing a co-constructed classroom. In Sarah Younie, Marilyn Leask, and Kevin Burden, editors, *Teaching and Learning with ICT in the Primary School*, chapter 11, pages 130–141. Routledge, Taylor & Francis Group (London and NY), 2 edition, 2014. ISBN 978-1-138-78315-7, 978-1-138-78314-0.

Peter L. Bonate. *Analysis of Pretest-Posttest Designs*. Chapman and Hall / CRC, USA, May 2000. ISBN 978-1-4200-3592-6. CRC Press, editor.

Harry N. Boone and Deborah A. Boone. Analyzing Likert Data. *Journal of Extension (https://tigerprints.clemson.edu/joe/)*, 50(2):1–5, Apr. 2012.

Aibhín Bray and Brendan Tangney. Technology usage in mathematics education research - A systematic review of recent trends. *Computers & Education (https://www.journals.elsevier.com/computers-and-education)*, 114:255–273, Jul. 2017. ISSN 0360-1315. doi: 10.1016/j.compedu.2017.07.004. Sue Bennett, Johan van Braak, Rachelle S. Heller, Miguel Nussbaum and Chin-Chung Tsai, editors. Elsevier Ltd ScienceDirect, publisher.

Guy Brousseau and Michael Otte. The Fragility of Knowledge. In Alan J. Bishop, Stieg Mellin-Olsen, and Van Dormolen Joop, editors, *Mathematical Knowledge: Its Growth Through Teaching*, volume 10 of *Mathematics Education Library*, pages 121–121. Springer, Dordrecht, 1991. ISBN 978-90-481-4095-4, 978-94-017-2195-0. doi: 10.1007/978-94-017-2195-0_2.

William Bryant. Developing a Strategy for Using Technology-Enhanced Items in Large-Scale Standardized Tests. *Practical Assessment, Research, and Evaluation*, 22(1):1–10, Feb. 2017. ISSN 1531-7714. doi: 10.7275/70yb-dj34. URL https://scholarworks.umass.edu/pare/vol22/iss1/1. ScholarWorks@UMass Amherst, publisher.

Lingguo Bu and Markus Hohenwarter. Modeling for Dynamic Mathematics: Toward Technology-Integrated Aesthetic Experiences in School Mathematics. In Xun Ge, Dirk Ifenthaler, and J. Michael Spector, editors, *Emerging Technologies for STEAM Education*, pages 355–379. Springer, Cham, Sept. 2015. ISBN 978-3-319-02572-8, 978-3-319-02573-5. doi: 10.1007/978-3-319-02573-5_19. Part of the Educational Communications and Technology: Issues and Innovations book series (ECTII).

Zhihui Cai, Xitao Fan, and Jianxia Du. Gender and attitudes toward technology use: A meta-analysis. *Computers & Education*, 105:1–13, Feb. 2017. ISSN 0360-1315. doi: 10.1016/j.compedu.2016.11.003. Elsevier Ltd ScienceDirect, publisher.

Martha Cantrell. Rethinking Proof with The Geometer's Sketchpad. *The Mathematics Teacher (MT), official journal of the National Council of Teachers of Mathematics*, 93(3):254–254, Mar. 2000. URL https://www.jstor.org/stable/27971359. National Council of Teachers of Mathematics, publisher. Accessed: 2021-01-30.

Rogelio E. Cardona-Rivera and R. Michael Young. A Cognitivist Theory of Affordances for Games. In *Digital Games Research Association (DiGRA) Conference*, North Carolina State University, NC, USA, 27695, 2013. URL http://www.rogel.io/content/2-publications/cardona-rivera2013cognitivist.pdf. Proceedings of DiGRA 2013: DeFragging Game Studies.

David Carless. Innovation in Language Teaching and Learning. *The Encyclopedia of Applied Linguistics*, Nov. 2012. doi: 10.1002/9781405198431.wbeal0540. Carol A. Chapelle, editor. Blackwell Publishing Ltd.

Susana Carreira, Keith Jones, Nélia Amado, Hélia Jacinto, and Sandra Nobre. Youngsters Solving Mathematical Problems with Technology: Summary and Implications. In *Youngsters Solving Mathematical Problems with Technology*, pages 209–239. Springer, Cham, Feb. 2016. ISBN 978-3-319-24908-7, 978-3-319-24910-0. doi: 10.1007/978-3-319-24910-0_8. Part of the Mathematics Education in the Digital Era book series (MEDE, volume 5).

Stephanie Carretero, Riina Vuorikari, and Yves Punie. DigComp 2.1: The Digital Competence Framework for Citizens, with eight proficiency levels and examples of use. May 2017. ISSN 1831-9424, 1018-5593. doi: 10.2760/38842,10.2760/836968. URL https://publications.jrc.ec.europa.eu/repository/bitstream/JRC106281/web-digcomp2.1pdf_(online).pdf. ISBN 978-92-79-68006-9, 978-92-79-68005-2. Publications Office of the European Union. Joint Research Centre, institution. Accessed: 2021-01-26.

Victoria Cartwright and Michael Hammond. 'Fitting it in': A study exploring ICT use in a UK primary school. *Australasian Journal of Educational Technology (AJET)*, 23(3):390–407, Sept. 2007. doi: 10.14742/ajet.1259.

European Centre for the Development of Vocational Training CEDEFOP. Luxembourg: Simply digital, a comprehensive digital education strategy for all. CEDEFOP, 2020. URL https://www.cedefop.europa.eu/it/news-and-press/news/luxembourg-simply-digital-comprehensive-digital-education-strategy-all. Accessed: 2020-09-10.

Malika Charrad, Nadia Ghazzali, Véronique Boiteau, and Azam Niknafs. NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *Journal of Statistical Software*, 61(6):1–36, Oct. 2014. doi: 10.18637/jss.v061.i06. URL http://www.jstatsoft.org/. https://cran.r-project.org/package=NbClust. R package version >= 3.1.0. Accessed: 2021-01-25.

Michelene T.H. Chi. Knowledge Structures and Memory Development. In Robert S. Siegler, editor, *Children's Thinking: What Develops?*, pages 73–96. Lawrence Erlbaum Associates (LEA), Hillsdale, New Jersey, 1978.

Inés Márquez Chisholm. Computer Use in a Multicultural Classroom. *Journal of Research (on Technology) on Computing in Education (JRTE)*, 28(2):162–174, 1995. doi: 10.1080/08886504.1995.10782158. International Society for Technology in Education (ISTE), editor. Routledge, Taylor & Francis Online, publisher.

Kyeong-Sik Choi. Motivating students in learning mathematics with GeoGebra. *Annals. Computer Science Series*, 8(2):65–76, 2010. URL https://users.qsm.ac.il/Assistant/geogebra/articles/geogebra1.pdf. Accessed: 2021-01-29.

Felicia W. Chu, Kristy VanMarle, and David C. Geary. Quantitative deficits of preschool children at risk for mathematical learning disability. *Frontiers in Psychology*, 4:195, May 2013. ISSN 1664-1078. doi: 10.3389/fpsyg.2013.00195. URL https://www.frontiersin.org/article/10.3389/fpsyg.2013.00195. Korbinian Moeller, editor.

Christos Chytas, Alexandros Tsilingiris, and Ira Diethelm. Exploring Computational Thinking Skills in 3D Printing: A Data Analysis of an Online Makerspace. In *2019 IEEE Global Engineering Education Conference (EDUCON)*, pages 1173–1179, Dubai, United Arab Emirates, Apr. 2019. IEEE, the world's largest technical professional organization dedicated to advancing technology for the benefit of humanity. ISBN 978-1-5386-9507-4, 978-1-5386-9506-7. doi: 10.1109/EDUCON.2019.8725202. ISSN 2165-9559, 2165-9567.

Michael Clark. Mixed Models with R: Getting started with random effects. Consulting for Statistics, Computing and Analytics Research (CSCAR), Advanced Research Computing (ARC), 2019. URL https://m-clark.github.io/mixed-models-with-R/. University of Michigan, USA. Accessed: 2020-03-14.

Norman Cliff. An Improved Internal Consistency Reliability Estimate. *Journal of Educational (and Behavioral) Statistics (JEBS)*, 9(2):151–161, Jun. 1984. doi: 10.2307/1164718. American Educational Research Association (AERA), publisher.

Benõ Csapó, John Ainley, Randy E Bennett, Thibaud Latour, and Nancy Law. Technological Issues for Computer-Based Assessment. In Patrick Griffin, Barry McGaw, and Esther Care, editors, *Assessment and Teaching of 21st Century Skills*, pages 143–230. Springer, Dordrecht, 2012. ISBN 978-94-007-2323-8, 978-94-007-2324-5. doi: 10.1007/978-94-007-2324-5_4.

Benő Csapó, Gyöngyvér Molnár, and Krisztina Farkasné Tóth. Comparing Paper-and-Pencil and Online Assessment of Reasoning Skills: A Pilot Study for Introducing Electronic Testing in Large-scale Assessment in Hungary. In Friedrich Scheuermann and Julius Björnsson, editors, *The Transition to Computer-Based Assessment: New Approaches to Skills Assessment and Implications for Large-scale Testing*, pages 120–125. Luxembourg: Office for Official Publications of the European Communities, Institute of Education, University of Szeged, Szeged, Hungary, 2009. ISBN 978-92-79-11110-5. doi: 10.2788/60083. ISSN 1018-5593.

Geoff Cumming, Fiona Fidler, and David L. Vaux. Error bars in experimental biology. *The Journal of Cell Biology (JCB)*, 177(1):7–11, Apr. 2007. ISSN 0021-9525. doi: 10.1083/jcb.200611141. URL https://rupress.org/jcb/article-pdf/177/1/7/1329781/jcb_200611141.pdf. The Rockefeller University Press, publisher.

Menal Dahiya. A Tool of Conversation: Chatbot. *International Journal of Computer Sciences and Engineering (IJCSE)*, 5(5):158–161, May 2017. ISSN 2347-2693. URL https://www.ijcseonline.org/pdf_paper_view.php?paper_id=1283&27-IJCSE-02149.pdf.

Michael D. de Villiers. *Rethinking Proof with The Geometer's Sketchpad*. Key Curriculum Press, Emeryville, CA, 1 edition, Aug. 2003. ISBN 978-1559536462. Number of pages: 214.

Rishi Desai and Shiv Gaglani. How to Study in Medical School. Osmosis.org, MSD Manual, Professional Version, The trusted provider of medical information since 1899, Jan. 2018. URL https://www.msdmanuals.com/professional/special-subjects/how-to-study-in-medical-school/how-to-study-in-medical-school. Accessed: 2020-03-30.

Digital Luxembourg innovative initiatives. digital4education. Digital Luxembourg, n.d. URL https://digital-luxembourg.public.lu/initiatives/digital4education. Accessed: 2020-09-10.

Monika Dockendorff and Horacio Solar. ICT integration in mathematics initial teacher training and its impact on visualization: the case of GeoGebra. *International Journal of Mathematical Education in Science and Technology (iJMEST, https://www.tandfonline.com/loi/tmes20)*, 49(1):66–84, 2018. ISSN 0020-739X, 1464-5211. doi: 10.1080/0020739X.2017.1341060. Taylor & Francis Group, publisher.

Carol S. Dweck. Messages that Motivate: How Praise Molds Students' Beliefs, Motivation, and Performance (in Surprising Ways). In Joshua Aronson, editor, *Improving Academic Achievement: Impact of Psychological Factors on Education*, Educational Psychology, chapter 3, pages 37–39. Academic Press, San Diego, California, 2002. ISBN 978-0-12-064455-1. doi: 10.1016/B978-012064455-1/50006-3. ISSN 1871-6148.

Charles Dziuban, Charles R. Graham, Patsy D. Moskal, Anders Norberg, and Nicole Sicilia. Blended learning: the new normal and emerging technologies. *International Journal of Educational Technology in Higher Education*, 15:1–16, Feb. 2018. ISSN 2365-9440. doi: 10.1186/s41239-017-0087-5. Springer Nature Switzerland AG, publisher. Article number: 3.

Steve Easterbrook. From Computational Thinking to Systems Thinking: A conceptual toolkit for sustainability computing. In *Proceedings of the 2014 conference ICT for Sustainability (ICT4S-14, 2nd International Conference on ICT for Sustainability)*, Advances in Computer Science Research, pages 235–244. Atlantis Press, Aug. 2014. ISBN 978-94-62520-22-6. doi: 10.2991/ict4s-14.2014.28. ISSN 2352-538X.

Robert L. Ebel. The Relation of Item Discrimination to Test Reliability. *Journal of Educational Measurement*, 4(3):125–128, 1967. URL https://www.jstor.org/stable/1434085. National Council on Measurement in Education, JSTOR, publisher. Accessed: 2019-06-30.

Leovy Echeverría, Ruth Cobos, Mario Morales, Fernando Moreno, and Victor Negrete. Promoting Computational Thinking Skills in Primary School Students to Improve Learning of Geometry. In *2019 IEEE Global Engineering Education Conference (EDUCON)*, pages 424–429, Dubai, United Arab Emirates, May 2019. IEEE, the world's largest technical professional organization dedicated to advancing technology for the benefit of humanity. ISBN 978-1-5386-9507-4, 978-1-5386-9506-7. doi: 10.1109/EDUCON.2019.8725088. ISSN 2165-9559, 2165-9567.

Carrefour pédagogique pour l'enseignement et l'apprentissage à l'ère numérique eduSphere. Digital (4) Education. WordPress, 2015. URL https://info.edusphere.lu/digital-letzebuerg/. Accessed: 2020-09-10.

Birgit Eickelmann and Mario Vennemann. Teachers' attitudes and beliefs regarding ICT in teaching and learning in European countries. *European Educational Research Journal (EERJ: Research for the benefit of education and society)*, 16(6):733–761, Nov. 2017. ISSN EISSN-1474-9041. doi: 10.1177/1474904117725899. SAGE Publications (http://sagepub.com) SAGE UK: London, England.

Claus Th. Ekstrom. *The R Primer*. CRC Press, imprint of the Taylor and Francis Group, University of Copenhagen, Denmark, 2 edition, Mar. 2017. ISBN 978-1-138-63197-7, 978-1-4987-7255-6. a Chapman and Hall Book / CRC The R Series. Number of pages: 426.

Susan E. Embretson and Scott L. Hershberger. *The New Rules of Measurement: What Every Psychologist and Educator Should Know*, pages 39–40. Psychology Press, 1999. ISBN 978-1-135-70660-9.

Lasse Eronen and Eija Kärnä. Students Acquiring Expertise through Student-Centered Learning in Mathematics Lessons. *Scandinavian Journal of Educational Research (https://www.tandfonline.com/toc/csje20/62/5?nav=tocList)*, 62(5):682–700, Oct. 2018. ISSN 0031-3831, 1470-1170. doi: 10.1080/00313831.2017.1306797. Routledge, Taylor & Francis Group, publisher.

Cho Eunseong and Kim Seonghoon. Cronbach's Coefficient Alpha: Well Known but Poorly Understood. *SAGE Journals: Organizational Research Methods*, 18(2):207–230, Apr. 2015. doi: 10.1177/1094428114555994. SAGE Publications SAGE CA: Los Angeles, CA, publisher. 2015 Best Paper Winner, special collection(s).

European Commission. Digital Education Action Plan (2018 - 2020). European Union, 2018. URL https://ec.europa.eu/education/education-in-the-eu/digital-education-action-plan_en. Education and Training, institution. Accessed: 2020-09-10.

Rossana Falcade. *Théorie des situations, médiation sémiotique et discussions collectives dans des séquences d'enseignement qui utilisent Cabri-géomètre et qui visent à l'apprentissage des notions de fonction et graphe de fonction [math]*. PhD thesis, Thèse en co-tutelle pour l'obtention du Diplôme de Docteur de l'Université Joseph-Fourier, Grenoble I, 2006. language: French.

Antoine Fischbach and Uwe Hentschel. Pisa-Studie: Trotz aller Kritik der beste Vergleich. science.lu, Wëssenschaft fir jiddereen, Fonds National de la Recherche Luxembourg, Dec. 2019. URL https://www.science.lu/de/bildung/pisa-studie-trotz-aller-kritik-der-beste-vergleich. Accessed: 2020-01-02.

Bethany Fishbein, Michael O. Martin, Ina V.S. Mullis, and Pierre Foy. The TIMSS 2019 Item Equivalence Study: examining mode effects for computer-based assessment and implications for measuring trends. *Large-scale Assessments in Education (an IEA-ETS Research Institute Journal)*, 6:1–23, Oct. 2018. ISSN 2196-0739. doi: 10.1186/s40536-018-0064-z. Article number: 11. SpringerOpen, publisher.

Five in a Row - My FIAR. Did you ever wonder why Five in a Row is based on reading the same book five days in a row? https://fiveinarow.com/2019/10/28/why-read-the-same-book-5-days-in-a-row/, n.d. 312 SW Greenwich Dr. Suite 220, Lee's Summit, MO 64082, (816) 866-8500. Accessed: 2020-10-01.

John Fox and Sanford Weisberg. *An R Companion to Applied Regression*. SAGE Publications, Thousand Oaks, CA, 3 edition, Sept. 2018. ISBN 978-1-5443-3647-3. URL https://socialsciences.mcmaster.ca/jfox/Books/Companion/. Number of pages: 608.

Norm Friesen. Report: Defining Blended Learning. Technical report, Aug. 2012. URL https://www.normfriesen.info/papers/Defining_Blended_Learning_NF.pdf. Number of pages: 10. Accessed: 2020-08-31.

Joseph M. Furner. Tackling Math Anxiety through Photography while using GeoGebra. *Transformations*, 5(1):59–75, 2019. ISSN 2474-610X. URL https://nsuworks.nova.edu/transformations/vol5/iss1/5. Article 5. Masthead Logo.

R. Michael Furr and Verne R. Bacharach. *Psychometrics: An Introduction*. SAGE Publications, Inc., Wake Forest University, Appalachian State University, 2 edition, Feb. 2013. ISBN 978-1-4522-5680-1. Number of pages: 442.

Eveline Gebhardt, Sue Thomson, John Ainley, and Kylie Hillman. What Have We Learned About Gender Differences in ICT. In IEA Research for Education (A Series of In-depth Analyses Based on Data of the International Association for the Evaluation of Educational Achievement (IEA)), editor, *Gender Differences in Computer and Information Literacy (An In-depth Analysis of Data from ICILS)*, volume 8, pages 69–73. Springer-Verlag, Cham, Sept. 2019. ISBN 978-3-030-26202-0. doi: 10.1007/978-3-030-26203-7_6.

Kurt F. Geisinger. 21st Century Skills: What Are They and How Do We Assess Them? *Applied Measurement in Education (21st Century Skill Assessment)*, 29(4):245–249, Jul. 2016. doi: 10.1080/08957347.2016.1209207. Routledge, Taylor & Francis Online, publisher.

Katerina V. Glezou. Robotics as a Powerful Vehicle Toward Learning and Computational Thinking in Secondary Education of 21st Century. In Stamatios Papadakis and Michail Kalogiannakis, editors, *Handbook of Research on Using Educational Robotics to Facilitate Student Learning*, pages 1–40. IGI Global Publisher of TIMELY KNOWLEDGE, Hershey, PA 17033, USA, Dec. 2020. ISBN 978-1-79986-717-3, 978-1-79986-719-7. doi: 10.4018/978-1-7998-6717-3.ch001. ISSN 2326-8905, 2326-8913. Advances in Educational Technologies and Instructional Design (AETID), collection.

Jeffrey A. Gliner, George A. Morgan, and Robert J. Harmon. Pretest-Posttest Comparison Group Designs: Analysis and Interpretation. *Journal of the American Academy of Child and Adolescent Psychiatry*, 42(4):500–503, Apr. 2003. doi: 10.1097/01.CHI.0000046809.95464.BE. Elsevier, publisher.

Chelsea Goforth. Using and Interpreting Cronbach's Alpha. University of Virginia Library, https://data.library.virginia.edu/using-and-interpreting-cronbachs-alpha/, Nov. 2015. Accessed: 2019-10-26.

Bart H.H. Golsteyn and Trudie Schils. Gender Gaps in Primary School Achievement: A Decomposition into Endowments and Returns to IQ and Non-cognitive Factors. *Economics of Education Review (https://www.journals.elsevier.com/economics-of-education-review)*, 41:176–187, Aug. 2014. ISSN 0272-7757. doi: 10.1016/j.econedurev.2014.04.001. Elsevier B.V. ScienceDirect, publisher. Working Paper https://www.econstor.eu/bitstream/10419/98968/1/dp8201.pdf. Accessed: 2021-06-06.

Katherine R. Gordon. How Mixed-Effects Modeling Can Advance Our Understanding of Learning and Memory and Improve Clinical and Educational Practice. *Journal of Speech, Language, and Hearing Research (JSLHR)*, 62(3):507–524, Mar. 2019. doi: 10.1044/2018_JSLHR-L-ASTM-18-0240. American Speech-Language-Hearing Association (ASHA), publisher.

Frederick J. Gravetter and Lori-Ann B. Forzano. *Research Methods for the Behavioral Sciences*. Wadsworth, Cengage Learning, USA, 5 edition, 2016. ISBN 978-1-305-10413-6. Number of pages: 656.

Meghan Cox Gurdon. *The Enchanted Hour: The Miraculous Power of Reading Aloud in the Age of Distraction*. Episodes, 130. Piatkus, London, Jan. 2019. ISBN 978-0-349-42294-7. Hachette UK, editor.

Dörte Haftendorn. Dynamische Mathematik - Bewegung beflügelt Verstehen. In Günter Graumann, editor, *Beiträge zum Mathematikunterricht 2005: Vorträge auf der 39. Tagung für Didaktik der Mathematik vom 28.2. bis 4.3.2005 in Bielefeld*, pages 235–238. Verlag Franzbecker, Hildesheim, Germany, 2005. ISBN 978-3-88120-404-0.

Julia Handl and Joshua Knowles. Multiobjective clustering around medoids. In *2005 IEEE (Institute of Electrical and Electronics Engineers) Congress on Evolutionary Computation*, volume 1, pages 632–639, Edinburgh, Scotland, UK, Oct. 2005. IEEE Xplore. ISBN 0-7803-9363-5. doi: 10.1109/CEC.2005.1554742. ISSN 1089-778X.

Joanne Hardman. Towards a pedagogical model of teaching with ICTs for mathematics attainment in primary school: A review of studies 2008-2018. *Heliyon (http://www.heliyon.com/)*, 5(5), May 2019. ISSN 2405-8440. doi: 10.1016/j.heliyon.2019.e01726. Elsevier B.V. ScienceDirect, publisher.

John Hattie and Helen Timperley. The Power of Feedback. *Review of Educational Research*, 77(1):81–112, Mar. 2007. doi: 10.3102/003465430298487. SAGE Publications SAGE CA: Thousand Oaks, CA. University of Auckland.

Jiří Haviger and Iva Vojkůvková. The van Hiele geometry thinking levels: gender and school type differences. In Zafer Bekirogullari and Melis Yagmur Minas, editors, *Procedia - Social and Behavioral Sciences*, volume 112, pages 977–981. Elsevier Ltd. ScienceDirect, Feb. 2014. doi: 10.1016/j.sbspro.2014.01.1257. ISSN 1877-0428. International Conference on Education & Educational Psychology 2013 (ICEEPSY 2013).

Stephen Hegedus and Luis Moreno-Armella. Information and Communication Technology (ICT) Affordances in Mathematics Education. In Stephen Lerman, editor, *Encyclopedia of Mathematics Education*, pages 380–384. Springer International Publishing, Cham, Switzerland, Feb. 2020. ISBN 978-3-030-15788-3, 978-3-030-15789-0. doi: 10.1007/978-3-030-15789-0_78.

Satu Helppolainen and Maija Aksela. Science teachers' ICT use from a viewpoint of Technological Pedagogical Content Knowledge (TPCK). *LUMAT: International Journal on Math, Science and Technology Education*, 3(6):783–799, Oct. 2015. ISSN 2323-7112. doi: 10.31129/lumat.v3i6.998. URL https://journals.helsinki.fi/lumat/article/view/998/991. University of Helsinki, Finland and LUMA Centre Finland, publishers. Accessed: 2021-05-28.

Christian Hennig. Clustering strategy and method selection. arXiv:1503.02059, arXivLabs: experimental projects with community collaborators, Cornell University, Mar. 2015. URL https://arxiv.org/pdf/1503.02059.pdf. in the Handbook of Cluster Analysis. Department of Statistical Science, University College London. Accessed: 2021-01-20.

Marian Henry. Learning in the Digital Age: Developing critical, creative and collaborative skills. In Sarah Younie, Marilyn Leask, and Kevin Burden, editors, *Teaching and Learning with ICT in the Primary School*, chapter 1, pages 1–12. Routledge, Taylor & Francis Group (London and NY), 2 edition, 2014. ISBN 978-1-138-78315-7, 978-1-138-78314-0.

Khe Foon Hew and Thomas Brush. Integrating technology into K-12 teaching and learning: current knowledge gaps and recommendations for future research. *Educational Technology Research and Development*, 55(3):223–252, Jun. 2007. ISSN 1556-6501. doi: 10.1007/

s11423-006-9022-5. Tristan E. Johnson and Lin Lin, editors. Springer Nature Switzerland AG, publisher.

Kathleen Hogan and Michael Pressley. *Scaffolding Student Learning: Instructional Approaches & Issues. Advances in Teaching and Learning Series.* Brookline Books, original from the University of Michigan, Ann Arbor, MI 48109 USA, 1997. ISBN 978-1-57129-036-6. Steve Graham, Karen R. Harris and Michael Pressley, editors. Number of pages: 198.

Judith Hohenwarter, Selay Arkün Kocadere, and Markus Hohenwarter. Maths teachers' adventure of ICT integration: From an open online course towards an online teacher community. In *CERME 10 - Tenth Congress of the European Society for Research in Mathematics Education*, pages 2390–2397, Dublin, Ireland, Feb. 2017. URL https://hal.archives-ouvertes.fr/hal-01942136/file/TWG15_08.pdf.

Markus Hohenwarter, Judith Preiner, and Taeil Yi. Incorporating GeoGebra into Teaching Mathematics at the College Level. In *Proceedings of the International Conference for Technology in Collegiate Mathematics*. ICTCM, Boston, MA, Jan. 2007.

Pi-Sui Hsu. Examining Current Beliefs, Practices and Barriers About Technology Integration: A Case Study. *TechTrends, Linking Research and Practice to Improve Learning*, 60(1):30–40, Jan. 2016. ISSN 1559-7075. doi: 10.1007/s11528-015-0014-3. Association for Educational Communications & Technology, editor. Springer Nature Switzerland AG, publisher.

Kim Huett. Designing K-12 Student-centered Blended Learning Environments. *Journal of Online Learning Research*, 4(3):295–326, Dec. 2018. ISSN 2374-1473. URL https://www.learntechlib.org/primary/p/181357/. The Learning & Technology Library (LearnTechLib), Association for the Advancement of Computing in Education (AACE), publisher. Waynesville, NC USA. Accessed: 2021-02-03.

Ren-Hung Hwang, Hsin-Tung Lin, Jerry Chih-Yuan Sun, and Jang-Jiin Wu. Improving Learning Achievement in Science Education for Elementary School Students via Blended Learning. *International Journal of Online Pedagogy and Course Design (IJOPCD, https://www.igi-global.com/journal/international-journal-online-pedagogy-course/1183)*, 9(2):44–62, Apr.-Jun. 2019. ISSN 2155-6873, 2155-6881. doi: 10.4018/IJOPCD.2019040104. Chia-Wen Tsai and Pei-Di Shen, editors. IGI Global Publisher of TIMELY KNOWLEDGE, publisher.

IEA. 21st Century Skills Mapping: Mapping 21st Century Skills in a Changing World. IEA, International Association for the Evaluation of Educational Achievement, Researching education, improving learning, 2021. Amsterdam (NL), Hamburg (DE). Accessed: 2021-02-11.

IEA Study TIMSS. Trends in International Mathematics and Science Study. IEA, Researching education, improving learning: International Association for the Evaluation of Educational Achievement, n.d. URL https://www.iea.nl/studies/iea/timss. Amsterdam (NL), Hamburg (DE). Accessed: 2020-08-25.

I. Isrokatun, D. Anggita, B.S. Purwono, C. Sunaengsih, and A.A. Syahid. Scaffolding in conceptual science. *Journal of Physics: Conference Series, 4th Annual Applied Science and Engineering Conference*, 1402(4), Dec. 2019. doi: 10.1088/1742-6596/1402/4/044079. IOP (science) Publishing Ltd.

Hélia Jacinto and Susana Carreira. Mathematical Problem Solving with Technology: the Techno-Mathematical Fluency of a Student-with-GeoGebra. *International Journal of Science and Mathematics Education*, 15(6):1115–1136, Aug. 2017. ISSN 1573-1774. doi: 10.1007/s10763-016-9728-8. Springer Nature Switzerland AG, publisher.

Sue Johnston-Wilder and John Mason. *Fundamental Constructs in Mathematics Education*. Researching Mathematics Learning S. RoutledgeFalmer, Taylor & Francis Group, London and NY, Jan. 2004. ISBN 978-1-134-33890-0. Psychology Press.

Ian T. Jolliffe and Jorge Cadima. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065), Apr. 2016. ISSN 1364-503X. doi: 10.1098/rsta.2015.0202. URL https://royalsocietypublishing.org/doi/pdf/10.1098/rsta.2015.0202. The Royal Society Publishing.

Mitchell Kapor. A Software Design Manifesto. *Dr. Dobb's Journal*, 16(1), Nov. 1990. doi: 10.1145/229868.230026. Association for Computing Machinery, ACM Digital Library. in booktitle: Bringing Design to Software (April 1996, pg. 1–6).

Victor Kaptelinin. *The Encyclopedia of Human-Computer Interaction*, chapter 44. Affordances. The Interaction Design Foundation, Aarhus, Denmark, 2 edition, 2014. URL https://www.interaction-design.org/encyclopedia/affordances_and_design.html. Mads Soegaard and Rikke F. Dam, editors. Accessed: 2020-03-26.

Rasha Kashef and Mohamed S. Kamel. Efficient Bisecting k-Medoids and Its Application in Gene Expression Analysis. In A. Campilho and M. Kamel, editors, *International Conference Image Analysis and Recognition (ICIAR 2008)*, pages 423–424. Springer-Verlag Berlin Heidelberg, 2008. ISBN 978-3-540-69811-1, 978-3-540-69812-8. doi: 10.1007/978-3-540-69812-8_42. Lecture Notes in Computer Science (LNCS), vol 5112.

Alboukadel Kassambara. *Practical Guide To Cluster Analysis in R: Unsupervised Machine Learning, Multivariate Analysis*, volume 1. STHDA (http://www.sthda.com), Aug. 2017. ISBN 978-1-5424-6270-9. Number of pages: 187.

Alboukadel Kassambara and Fabian Mundt. *factoextra: Extract and Visualize the Results of Multivariate Data Analyses*, Apr. 2020. URL https://CRAN.R-project.org/package=factoextra. R package version 1.0.7. Accessed: 2020-09-10.

Leonard Kaufman and Peter J. Rousseeuw. *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley Series in Probability and Statistics. Wiley-Interscience Online Library, a John Wiley & Sons, Inc., publication, Hoboken, New Jersey, USA, Mar. 1990. ISBN 978-0471878766. doi: 10.1002/9780470316801.

Elizabeth R. Kazakoff, Paul Macaruso, and Pam Hook. Efficacy of a blended learning approach to elementary school reading instruction for students who are English Learners. *Educational Technology Research and Development*, 66(2):429–449, Apr. 2018. ISSN 1556-6501. doi: 10.1007/s11423-017-9565-7. Association for Educational Communications & Technology (AECT), editor. Springer Nature Switzerland AG, publisher.

Lisa Kervin and Jessica Mantei. Incorporating technology within classroom literacy experiences. *Journal of Literacy and Technology*, 11(3):77–100, Jul. 2010. ISSN 1535-0975. URL https://ro.uow.edu.au/edupapers/313. University of Wollongong Australia, Research Online (RO), Part of the Education Digital Commons. Accessed: 2021-05-31.

Muhammad Khalil, Umair Khalil, and Zahoor ul Haq. Geogebra as a Scaffolding Tool for Exploring Analytic Geometry Structure and Developing Mathematical Thinking of Diverse Achievers. *International Electronic Journal of Mathematics Education (IEJME, http://www.iejme.com/)*, 14(2):427–434, May 2019. ISSN 1306-3030. doi: 10.29333/iejme/5746. Modestum, UK.

Mugenyi Justice Kintu, Chang Zhu, and Edmond Kagambe. Blended learning effectiveness: the relationship between student characteristics, design features and outcomes. *International Journal of Educational Technology in Higher Education*, 14:1–20, Feb. 2017. ISSN 2365-9440. doi: 10.1186/s41239-017-0043-4. Article number: 7. Springer Nature, BioMed Central Ltd, publisher.

Sotaro Kita and Asli Özyürek. How Does Spoken Language Shape Iconic Gestures? In Susan D. Duncan, Justine Cassell, and Elena T. Levy, editors, *Gesture and the Dynamic Dimension of Language: Essays in Honor of David McNeill*, volume 1, pages 67–75. John Benjamins Publishing Company, Amsterdam and Philadelphia, PA, Jun. 2007. ISBN 978-90-272-2841-3. doi: 10.1075/gs.1.07kit. ISSN 1874-6829. Gesture studies.

Charles Kivunja. Exploring the Pedagogical Meaning and Implications of the 4Cs "Super Skills" for the 21st Century through Bruner's 5E Lenses of Knowledge Construction to Improve Pedagogies of the New Learning Paradigm. *Creative Education (CE, https://www.scirp.org/journal/ce/)*, 6(2):224–239, Feb. 2015. ISSN 2151-4755, 2151-4771. doi: 10.4236/ce.2015.62021. Scientific Research Publishing Inc (SciRes).

Paul Kline. *The Handbook of Psychological Testing*. Routledge, imprint of the Taylor and Francis Group, GB and NY, 2 edition, 2000. ISBN 978-0-415-21158-1. Number of pages: 744.

Siu-Cheung Kong, Ming Ming Chiu, and Ming Lai. A study of primary school students' interest, collaboration attitude, and programming empowerment in computational thinking education. *Computers & Education*, 127:178–189, Dec. 2018. ISSN 0360-1315. doi: 10.1016/j.compedu.2018.08.026. Elsevier Ltd ScienceDirect, publisher.

Yves Kreis, Carole Dording, and Ulrich Keller. GeoGebraPrim - GeoGebra in der Grundschule. In Anke Lindmeier and Stefan Ufer, editors, *Beiträge zum Mathematikunterricht 2010, 44. Jahrestagung der Gesellschaft für Didaktik der Mathematik vom 08. bis 12. März 2010 in München*, pages 513–516, Munich, Germany, 2010a. Gesellschaft für Didaktik der Mathematik (GDM). ISBN 978-3-942197-03-8. doi: 10.17877/DE290R-13545. URL https://orbilu.uni.lu/bitstream/10993/46729/1/BzMU10_KREIS_Yves_GeoGebraPrim.pdf. Accessed: 2021-05-31.

Yves Kreis, Carole Dording, Ulrich Keller, Vincent Porro, and Raynald Jadoul. Dynamic mathematics and computer-assisted testing: Geogebra inside TAO. In Ulrich Kortenkamp, Pavel Pech, Christian Dohrmann, and Roman Hašek, editors, *Proceedings of the Interoperable Interactive Geometry (I2GEO) Conference 2010*, University of South Bohemia, Hluboká nad Vltavou, Czech Republic, 2010b. URL https://cermat.org/i2geo2010/downloads/files/I2GEO2010-Kreis.pdf. Accessed: 2020-10-01.

Yves Kreis, Carole Dording, Ulrich Keller, Vincent Porro, and Raynald Jadoul. Dynamische Mathematik und computergestützte Tests: GeoGebra in TAO. In Silke Ladel, Julia Knopf, and Armin Weinberger, editors, *Digitalisierung und Bildung*, pages 59–80. Springer VS, Wiesbaden, Germany, 2018. ISBN 978-3-658-18332-5, 978-3-658-18333-2. doi: 10.1007/978-3-658-18333-2_4.

Bjarnheiður (Bea) Kristinsdóttir, Freyja Hreinsdóttir, and Zsolt Lavicza. Realizing students' ability to use technology with silent video tasks. In Hans-Georg Weigand, Alison Clark-Wilson, Ana Donevska-Todorova, Eleonora Faggiano, Niels Grønbæk, and Jana Trgalova, editors, *Proceedings of the 5th ERME Topic Conference Mathematics Education in the Digital Age (MEDA)*, pages 163–170, Copenhagen, Denmark, Sept. 2018. ERME, European Society for Research in Mathematics Education. ISBN 978-87-7078-798-7.

Bjarnheiður (Bea) Kristinsdóttir, Freyja Hreinsdóttir, and Zsolt Lavicza. Silent video tasks: Towards a definition. In Uffe Th. Jankvist, Marja Van den Heuvel-Panhuizen, and Michiel Veldhuis, editors, *CERME 11 - Eleventh Congress of the European Society for Research in Mathematics Education*, volume TWG15 of *Proceedings of the 11th Congress of the European Society for Research in Mathematics Education (CERME11)*, Utrecht, Netherlands, Feb. 2019. Utrecht University, Freudenthal Group. URL https://hal.archives-ouvertes.fr/hal-02417067/file/TWG15_11.pdf. Accessed: 2021-05-31.

Ranjit Kumar. *Research Methodology: A Step-by-Step Guide for Beginners*. SAGE Publications Limited, 3 edition, Nov. 2010. ISBN 978-0-85702-345-2.

Ranjit Kumar. *Research Methodology: A Step-by-Step Guide for Beginners*. SAGE Publications Limited, 5 edition, Dec. 2018. ISBN 978-1-5264-5708-0, 978-1-5264-4989-4. Aly Owen, editor.

Emily R. Lai and Michaela Viering. Assessing 21st Century Skills: Integrating Research Findings. *Pearson, World's learning company (http://www.pearsoned.com/)*, pages 1–67, Apr. 2012. Institute of Education Sciences, ERIC. Paper presented at the Annual Meeting of The National Council on Measurement in Education. Vancouver, British Columbia (B.C.), Canada.

Stephen Lamb, Quentin Maire, and Esther Doecke. Key Skills for the 21st Century: an evidence-based review. Technical report, Education: Future Frontiers, NSW Department of Education, State of New South Wales, Australia, 2017. URL http://vuir.vu.edu.au/35865/1/Key-Skills-for-the-21st-Century-Analytical-Report.pdf. Analytical Report. Number of pages: 70. Accessed: 2021-02-18.

Susan M. Land, Michael J. Hannafin, and Kevin M. Oliver. Student-Centered Learning Environments: Foundations, Assumptions and Design. In David H. Jonassen and Susan M. Land, editors, *Theoretical Foundations of Learning Environments*, chapter 1, pages 3–25. Routledge, Taylor & Francis Group (NY and London), 2 edition, Mar. 2012. ISBN 978-0-415-89422-7, 978-0-415-89421-0.

Mike Lansdown. *In the Zone: Helping Children Rise to the Challenge of Learning*. Critical Publishing Ltd, University of Buckingham, Apr. 2020. ISBN 978-1-913063-85-6, 978-1-913063-88-7. URL https://books.google.lu/books?id=ltnaDwAAQBAJ. Number of pages: 160. Accessed: 2021-05-31.

Shahram Latifi. *16th International Conference on Information Technology-New Generations (ITNG 2019), Advances in Intelligent Systems and Computing*, volume 800. Springer Nature Switzerland AG, 1 edition, May 2019. ISBN 978-3-030-14069-4. doi: 10.1007/978-3-030-14070-0. Number of pages: 611. ISSN 2194-5357.

Thibaud Latour and Romain Martin. TAO: An Open and Versatile Computer-Based Assessment Platform Based on Semantic Web Technology. *ERCIM News, Special theme: Technology-Enhanced Learning*, 71:32–33, Oct. 2007. URL http://ercim-news.ercim.eu/tao-an-open-and-versatile-assessment-platform-based-on-semantic-web-technology. Accessed: 2020-10-01.

Zsolt Lavicza, Ben Haas, and Yves Kreis. Discovering Everyday Mathematical Situations Outside the Classroom with MathCityMap and GeoGebra 3D. In Matthias Ludwig, Simone Jablonski, Amélia Caldeira, and Ana Moura, editors, *Research on Outdoor STEM Education in the digiTal Age: Proceedings of the ROSETA Online Conference in June 2020*, volume 6, pages 23–30, Münster, Germany, Jun. 2020. WTM Publishing and Communications. ISBN 978-3-95987-144-0. doi: 10.37626/GA9783959871440.0.03. URL https://orbilu.uni.lu/bitstream/10993/43471/1/Lavicza%20et%20al.%20-%202020%20-%20Discovering%20Everyday%20Mathematical%20Situations%20Outsi.pdf. Conference Proceedings in Mathematics Education. Accessed: 2021-01-29.

George Leckie. Cross-classified multilevel models. arXiv:1907.02569, arXivLabs: experimental projects with community collaborators, Cornell University, 2019a. URL https://arxiv.org/pdf/1907.02569.pdf. Centre for Multilevel Modelling and School of Education, University of Bristol. Accessed: 2021-01-20.

George Leckie. Multiple membership multilevel models. arXiv:1907.04148, arXivLabs: experimental projects with community collaborators, Cornell University, 2019b. URL https://arxiv.org/pdf/1907.04148.pdf. Centre for Multilevel Modelling and School of Education, University of Bristol. Accessed: 2021-01-20.

Günther Lehnert and Hubert D. Zimmer. Auditory and visual spatial working memory. *Memory and Cognition*, 34(5):1080–1090, Jul. 2006. doi: 10.3758/BF03193254. Psychonomic Society, Inc., Springer Link, publisher.

Dajiang Lei, Qingsheng Zhu, Jun Chen, Hai Lin, and Peng Yang. Automatic PAM Clustering Algorithm for Outlier Detection. *Journal of Software (JSW)*, 7(5):1045–1046, May 2012. ISSN 1796-217X. doi: 10.4304/jsw.7.5.1045-1051. URL http://www.jsoftware.us/vol7/jsw0705-15.pdf. Academy Publisher. Accessed: 2021-05-31.

Thomas Lenz, Isabell E. Baumann, and et al. Rapport national sur l'éducation au Luxembourg 2018. We-Print, Luxembourg, Esch-sur-Alzette, L, Dec. 2018. ISBN 978-99959-1-140-9. URL https://script.lu/sites/default/files/publications/2019-12/Bildungsbericht%202018.pdf. SCRIPT, Université du Luxembourg (LUCET), editor. Pages: 30-33 & 158-167. Accessed: 2021-05-31.

Samantha Leonard. Infographic: Building Engaging Online Assessments with Technology Enhanced Items (TEIs). TAO, Open Assessment Technologies, Apr. 2020a. URL https://www.taotesting.com/blog/building-engaging-online-assessments-with-technology-enhanced-items/. Accessed: 2020-09-01.

Samantha Leonard. Empowering Online Education with Digital Testing Technology. TAO, Open Assessment Technologies, Mar. 2020b. URL https://www.taotesting.com/blog/empowering-online-education-with-digital-testing-technology/. Accessed: 2020-09-01.

Lexia CORE5. Lexia CORE5. Family Literacy Centers, 2017. URL https://www.lexiaforhome.com/lexia-core5. Reading Director Deanna Weierholt, contact. Accessed: 2020-08-31.

Tao Li. A Unified View on Clustering Binary Data. *Machine Learning*, 62:199–215, Jan. 2006. doi: 10.1007/s10994-005-5316-9. Jennifer Dy, editor. Springer Science + Business Media, Inc. Manufactured in The Netherlands, publisher.

Yanjun Li, Meng Zhang, Yinghe Chen, Zhijun Deng, Xiaoshuang Zhu, and Shijia Yan. Children's Non-symbolic and Symbolic Numerical Representations and Their Associations With Mathematical Ability. *Frontiers in Psychology*, 9:1035, Jun. 2018. ISSN 1664-1078. doi: 10.3389/fpsyg.2018.01035. URL https://www.frontiersin.org/article/10.3389/fpsyg.2018.01035. Ann Dowker, editor. Accessed: 2021-05-31.

Rensis Likert. A Technique for the Measurement of Attitudes. *Archives of Psychology New York*, 22(140):5–55, Jun. 1932. Robert S. Woodworth, editor.

Serge Linckels and Christoph Meinel. *E-Librarian Service: User-Friendly Semantic Search in Digital Libraries*. X.media.publishing. Springer-Verlag Berlin Heidelberg, 1 edition, 2011. ISBN 978-3-642-17742-2, 978-3-642-17743-9. doi: 10.1007/978-3-642-17743-9. ISSN 1612-1449.

Serge Linckels, Carole Dording, and Christoph Meinel. Better results in mathematics lessons with a virtual personal teacher. In *Proceedings of the 34th annual ACM SIGUCCS Fall Conference: Expanding the Boundaries*, pages 201–209, Edmonton, Alberta, Canada, Nov. 2006. Association for Computing Machinery, ACM Digital Library. ISBN 1-59593-438-3. doi: 10.1145/1181216.1181261.

Serge Linckels, Carole Dording, and Christoph Meinel. Bessere Schulnoten mit MatES, dem e-Bibliothekardienst für den Mathematikunterricht. In Christian Eibl, Johannes Magenheim, Sigrid Schubert, and Martin Wessner, editors, *DeLFI 2007: 5. e-Learning Fachtagung Informatik der Gesellschaft für Informatik e.V. (GI)*, pages 91–102, Bonn, Germany, 2007. Gesellschaft für Informatik e.V.

Serge Linckels et al. Digital (4) Education - A National Digital Strategy in the Context of the Transformation of a Country into a Digital Nation. In *European Distance and E-Learning Network (EDEN) Annual Conference Proceedings*, number 1, pages 335–342. European Distance and E-Learning Network (EDEN), Jun. 2019. ISBN 978-615-5511-27-1. doi: 10.38069/edenconf-2019-ac-0037. ISSN 2707-2819. Connecting through Educational Technology.

Darren J. Lipomi. Video for Active and Remote Learning. *Trends in Chemistry*, 2(6):483–485, Jun. 2020. ISSN 2589-7209. doi: 10.1016/j.trechm.2020.03.003. Cell Press, Elsevier, publisher. Scientific Life.

Chris Little. Interactive Geometry in the Classroom: Old Barriers and New Opportunities. *Mathematics in School*, 38(2):9–11, Mar. 2009. URL https://www.jstor.org/stable/20696871. The Mathematical Association, JSTOR, publisher. Accessed: 2021-01-27.

Todd D. Little. *The Oxford Handbook of Quantitative Methods*, volume 2 (Statistical Analysis). Oxford (Library of Psychology) University Press, 2013. ISBN 978-0-19-993487-4, 978-0-19-993489-8. pages: 718-728. Peter E. Nathan, editor.

Martin Maechler, Peter J. Rousseeuw, Anja Struyf, Mia Hubert, and Kurt Hornik. cluster: Cluster Analysis Basics and Extensions, Jun. 2019. URL https://CRAN.R-project.org/package=cluster. R package version 2.1.0 — For new features, see the 'Changelog' file (in the package source). Accessed: 2021-05-31.

Oded Maimon and Lior Rokach. *Data Mining and Knowledge Discovery Handbook*. Series in Solid-State Sciences. Springer Science & Business Media, 2 edition, Oct. 2010. ISBN 978-0-387-09822-7. doi: 10.1007/978-0-387-09823-4. Number of pages: 1285.

Delfina Malandrino, Ilaria Manno, Giuseppina Palmieri, Vittorio Scarano, and Giovanni Filatrella. How quiz-based tools can improve students' engagement and participation in the classroom. In *2014 International Conference on Collaboration Technologies and Systems (CTS)*, pages 379–386, Minneapolis, MN, USA, Jul. 2014. IEEE Xplore, Institute of Electrical and Electronics Engineers (IEEE). ISBN 978-1-4799-5157-4, 978-1-4799-5158-1. doi: 10.1109/CTS.2014.6867592.

Ronald C. Martella, J. Ron Nelson, Robert L. Morgan, and Nancy E. Marchand-Martella. *Understanding and Interpreting Educational Research*, pages 135–175. The Guilford Press, New York and London, Apr. 2013. ISBN 978-1-4625-0962-1, 978-1-4625-0974-4.

Elizabeth Masterman and Yvonne Rogers. A framework for designing interactive multimedia to scaffold young children's understanding of historical chronology. *Instructional Science*, 30(3):221–241, May 2002. doi: 10.1023/A:1015133106888. Kluwer Academic Publishers (Netherlands), Springer-Verlag.

Charles Max. The development of initial teacher education focusing on multilingualisms: the innovative approach of Luxembourg. In Friederike Klippel and Gudrun Ziegler, editors, *ForumSprache*, volume 5, pages 59–78. Hueber Verlag, Ismaning, Germany, Mar. 2011. ISBN 978-3-19-476100-1. URL https://www.hueber.de/media/36/ForumSprache_05_2011_978-3-19-436100-3.pdf. ISSN 1868-0852. Accessed: 2021-02-01.

Natalie McCosker and Carmel M. Diezmann. Scaffolding Students' Thinking in Mathematical Investigations. *Australian Primary Mathematics Classroom*, 14(3):27–32, Dec. 2009. ISSN 1326-0286. Australian Association of Mathematics Teachers Inc., publisher.

Patrick McKenzie and Kerry Vaughan. Digital Library Usability Laboratory. Charles Sturt University, School of Information Studies, n.d. URL https://arts-ed.csu.edu.au/schools/sis/research-in-the-school-of-information-studies/lab. Accessed: 2020-09-06.

Sabine Meinck and Falk Brese. Trends in gender gaps: using 20 years of evidence from TIMSS. *Large-scale Assessments in Education, IEA-ETS Research Institute Journal*, 7(1):Article number: 8, Jul. 2019. ISSN 2196-0739. doi: 10.1186/s40536-019-0076-3. SpringerOpen, publisher.

Debra K. Meyer and Julianne C. Turner. Scaffolding Emotions in Classrooms. In Paul A. Schutz and Reinhard Pekrun, editors, *Emotion in Education*, Educational Psychology, chapter 14, pages 235–250. Academic Press, Elsevier B.V. ScienceDirect, Burlington, Vermont, Sept. 2007. ISBN 978-0-12-372545-5. doi: 10.1016/B978-012372545-5/50015-0. ISSN 1871-6148.

Michigan Virtual University. Teacher Boost, Helping teachers replace traditional top-down, teacher-led instruction with more student-centered learning. https://michiganvirtual.org/professionals/blended-learning/teacher-boost/, n.d. Michigan Virtual. 920 Municipal Way, Lansing, MI 48917-6872, USA. Accessed: 2020-01-02.

Paul M. Morsink, Michelle Sch. Hagerman, Anne Heintz, D. Matthew Boyer, Robin Harris, Kristen Kereluik, Douglas K. Hartman, Anne Wood, Amber White, Carmen Woodruff, et al. Professional Development to Support TPACK Technology Integration: The Initial Learning Trajectories of Thirteen Fifth- and Sixth-Grade Educators. *Journal of Education*, 191(2):3–16, Apr. 2011. ISSN 0022-0574, 2515-5741. doi: 10.1177/002205741119100203. SAGE Publications.

Eddie M. Mulenga and José M. Marbán-Prieto. Teachers' ICT Skills, Beliefs and Attitudes towards ICT Integration in the Teaching and Learning of Mathematics. A local study in Kabwe district in Zambia. *Journal of Global Research in Education and Social Science (JOGRESS)*, 11(4):176–189, Aug. 2018. ISSN 2454-1834. Ignatius Isaac Dambudzo, editor.

Nathan Naparstek. *Learning Solutions: What To Do If Your Child Has Trouble With Schoolwork*. Information Age Publishing (IAP), Inc., Mar. 2010. ISBN 978-1-60752-321-5, 978-1-60752-322-2. Number of pages: 155.

Diem M. Nguyen. Geogebra in the Classroom. In G. Donald Allen and Amanda Ross, editors, *Pedagogy and Content in Middle and High School Mathematics*, pages 47–48. Springer-Verlag, SensePublishers, Rotterdam, Jan. 2017. ISBN 978-94-6351-137-7. doi: 10.1007/978-94-6351-137-7_13.

Open Assessment Technologies OAT. Using Adaptive Testing in Digital Assessment to Support Learning. TAO, Open Assessment Technologies, Apr. 2020a. URL https://www.taotesting.com/blog/using-adaptive-testing-in-digital-assessment-to-support-learning/. Accessed: 2020-09-01.

Open Assessment Technologies OAT. Breaking Down Barriers: Designing Digital Assessment Systems for Accessibility from the Ground Up. Eventbrite, Sept. 2020b. URL https://www.eventbrite.com/e/designing-digital-assessment-systems-for-accessibility-from-the-ground-up-tickets-120293590367?utm. Accessed: 2020-09-17.

Open Assessment Technologies OAT. How to Boost Your Digital Assessment Platform with Integrations. TAO, Open Assessment Technologies, Jan. 2020c. URL https://www.taotesting.com/blog/how-to-boost-your-digital-assessment-platform-with-integrations/. Accessed: 2021-03-17.

Open Assessment Technologies OAT. How Digital Assessment Tools Make it Easier to Prove Learning Outcomes. TAO, Open Assessment Technologies, Apr. 2020d. URL https://www.taotesting.com/blog/how-digital-assessment-tools-make-it-easier-to-prove-learning-outcomes/. Accessed: 2020-09-01.

OAT, Open Assessment Technologies. Why Online Testing Tools are the Future of Assessments. TAO, Open Assessment Technologies, Dec. 2019. URL https://www.taotesting.com/blog/why-online-testing-tools-are-the-future-of-assessments/. Accessed: 2020-09-01.

Open Assessment Technologies OAT Practical Guide. The Practical Guide to Digital Assessment. TAO (https://www.taotesting.com), Open Assessment Technologies, n.d. URL https://www.accelerate.taotesting.com/hubfs/ebook-practical-guide-to-digital-assessment.pdf. Number of pages: 14. Accessed: 2021-03-17.

OECD. Girls' lack of self-confidence. In *The ABC of Gender Equality in Education: Aptitude, Behaviour, Confidence*, pages 63–95. OECD publishing Paris, 3 2015. doi: 10.1787/9789264229945-6-en.

OECD. *The ABC of Gender Equality in Education: Aptitude, Behaviour, Confidence*. OECD (Better Policies for Better Lives), OECD publishing, Paris Cedex 16, France, Mar. 2015. ISBN 978-92-64-23002-6, 978-92-64-22994-5. doi: 10.1787/9789264229945-en. URL https://www.oecd-ilibrary.org/content/publication/9789264229945-en. Number of pages: 180. PISA, editor. Accessed: 2021-02-03.

OECD. *PISA 2015 Results (Volume I)*, volume 1, chapter Excellence and Equity in Education. OECD (Better Policies for Better Lives) Publishing Paris, Dec. 2016. ISBN 978-92-64-26732-9, 978-92-64-26649-0. doi: 10.1787/9789264266490-en. URL https://www.oecd-ilibrary.org/content/publication/9789264266490-en. ISSN 1990-8539, 1996-3777. Accessed: 2021-05-31.

Office of Educational Assessment. Understanding Item Analyses. http://www.washington.edu/assessment/scanning-scoring/scoring/reports/item-analysis/, n.d. University of Washington, Seattle, WA. Accessed: 2019-06-30.

Jeroen Ooms. The jsonlite Package: A Practical and Consistent Mapping Between JSON Data and R Objects. *Cornell University, arXiv preprint arXiv:1403.2805v1*, Mar. 2014. URL https://arxiv.org/pdf/1403.2805.pdf. Number of pages: 29. Accessed: 2020-05-03.

Carlos Ordonez. Clustering Binary Data Streams with K-means. In *Proceedings of the 8th ACM Special Interest Group on Management of Data (SIGMOD) Workshop on Research Issues in Data Mining and Knowledge Discovery - DMKD*, pages 12–19. Association for Computing Machinery (ACM), Jun. 2003. doi: 10.1145/882082.882087.

Roy D. Pea. The Social and Technological Dimensions of Scaffolding and Related Theoretical Concepts for Learning, Education, and Human Activity. *The Journal of the Learning Sciences*, 13(3):423–451, Jul. 2004. ISSN 1050-8406, 1532-7809. doi: 10.1207/s15327809jls1303_6. Routledge Taylor & Francis Group, publisher.

Patrick Plichart, Gilbert Busana, Romain Martin, and Thibaud Latour. Computer Based School System Monitoring with Feedback to Teachers. In J. Luca and E. Weippl, editors, *Proceedings of ED-MEDIA 2008 - World Conference on Educational Multimedia, Hypermedia and Telecommunications (EdMedia + Innovate Learning)*, number 1, pages 5065–5070, Chesapeake, USA, Jun. 2008. Vienna, Austria: American Association for the Advancement of Computing in Education (AACE). ISBN 978-1-880094-65-5.

Sadhana Puntambekar and Roland Hubscher. Tools for Scaffolding Students in a Complex Learning Environment: What Have We Gained and What Have We Missed? *Educational Psychologist*, 40(1):1–12, 2005. ISSN 0046-1520. doi: 10.1207/s15326985ep4001_1.

Luis Radford. Why Do Gestures Matter? Sensuous Cognition and the Palpability of Mathematical Meanings. *Educational Studies in Mathematics: Gestures and Multimodality in the Construction of Mathematical Meaning*, 70(2):111–126, Mar. 2009. doi: 10.1007/s10649-008-9127-3. URL http://www.jstor.org/stable/40284564. Springer, publisher. Accessed: 2021-05-31.

Slaviša Radović, Marija Radojičić, Kristina Veljković, and Miroslav Marić. Examining the effects of Geogebra applets on mathematics learning using interactive mathematics textbook. *Interactive Learning Environments (https://www.tandfonline.com/loi/nile20)*, 28(1):32–49, 2020. ISSN 1049-4820, 1744-5191. doi: 10.1080/10494820.2018.1512001. Routledge, Taylor & Francis Group, publisher.

Noor Hanim Rahmat. Innovation in Education: Barriers and Facilitating Factors. *European Journal of Education Studies (https://www.oapub.org/edu/)*, 6(10):55–66, 2020. ISSN 2501-1111. doi: 10.5281/zenodo.3596994. URL https://www.oapub.org/edu/index.php/ejes/article/view/2784. Open Access Publishing Group (OAPub), European Open Access Research Association (EOARA). Accessed: 2021-05-31.

Eric Ras, Judith Swietlik, Patrick Plichart, and Thibaud Latour. TAO - A Versatile and Open Platform for Technology-Based Assessment. In Martin Wolpers, Paul A Kirschner, Maren Scheffel, Stefanie Lindstaedt, and Vania Dimitrova, editors, *European Conference on Technology Enhanced Learning, EC-TEL 2010: Sustaining TEL: From Innovation to Learning and Practice*, volume 6383, pages 644–649. Springer-Verlag, Berlin, Heidelberg, Germany, 2010. ISBN 978-3-642-16019-6, 978-3-642-16020-2. doi: 10.1007/978-3-642-16020-2_68. Lecture Notes in Computer Science.

Eric Ras, Valérie Maquil, Muriel Foulonneau, and Thibaud Latour. Using tangible user interfaces for technology-based assessment - Advantages and challenges. *International Computer Assisted Assessment Conference (CAA 2012) in Southampton*, Jul. 2012. URL https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.475.3393&rep=rep1&type=pdf. Accessed: 2021-05-31.

Brian J. Reiser. Scaffolding Complex Learning: The Mechanisms of Structuring and Problematizing Student Work. *The Journal of the Learning Sciences*, 13(3):273–304, 2004. doi: 10.1207/s15327809jls1303_2. Taylor & Francis Online, publisher.

William Revelle. *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University, Evanston, Illinois, USA, Dec. 2019. URL https://CRAN.R-project.org/package=psych. https://personality-project.org/r/psych-manual.pdf. R package version 1.9.12. Accessed: 2021-05-31.

Kirt Saville. Strategies for Using Repetition as a Powerful Teaching Tool. *Music Educators Journal*, 98(1):69–75, Sept. 2011. ISSN 0027-4321. doi: 10.1177/0027432111414432. SAGE Publications: Los Angeles, CA.

John W. Saye and Thomas Brush. Scaffolding critical reasoning about history and social issues in multimedia-supported learning environments. *Educational Technology Research (ETR) and Development (D)*, 50(3):77–96, Sept. 2002. ISSN 1042-1629. doi: 10.1007/BF02505026. Springer, publisher.

Scaffolding. Scaffolding. The Glossary of Education Reform for Journalists, Parents, and Community Members. Great Schools Partnership, Jun. 2015. URL `https://www.edglossary.org/scaffolding/`. Portland, ME, USA. Accessed: 2020-08-29.

Janessa Schilmoeller, Lori Griswold, and Neal Strudler. Multicultural Technology Education: The Need to Teach Digital Technologies to All Students. In Christine Clark, Amanda VandeHei, Kenneth J. Fasching-Varner, and Zaid M. Haddad, editors, *Multicultural Curriculum Transformation in Science, Technology, Engineering, and Mathematics*, volume 1, pages 205–220. Lexington Books, The Rowman & Littlefield Publishing Group, 2018. ISBN 978-1-4985-8051-9, 978-1-4985-8052-6.

Peter Scrimshaw. Enabling teachers to make successful use of ICT. (1):1–45, Jun. 2004. URL `https://dera.ioe.ac.uk/1604/1/becta_2004_enablingsuccessfuluse_litrev.pdf`. Available at: `http://www.becta.org.uk`, British Educational Communications and Technology Agency (Becta ICT Research). Accessed: 2021-01-26.

Steven J. Seage and Mehmet Türegün. The Effects of Blended Learning on STEM Achievement of Elementary School Students. *International Journal of Research in Education and Science (IJRES, `https://www.ijres.net/`)*, 6(1):133–140, 2020. ISSN 2148-9955. doi: 10.46328/ijres.v6i1.728. URL `https://files.eric.ed.gov/fulltext/EJ1231349.pdf`. Wilfried Admiraal, editor. IJRES, publisher. Accessed: 2021-05-31.

Fazilat Siddiq and Ronny Scherer. Is there a gender gap? A meta-analysis of the gender differences in students' ICT literacy. *Educational Research Review*, 27:205–217, Jun. 2019. ISSN 1747-938X. doi: 10.1016/j.edurev.2019.03.007. Elsevier Ltd ScienceDirect, publisher.

Marjorie Siegel and Raffaella Borasi. Toward a New Integration of Reading in Mathematics Instruction. *Focus on Learning Problems in Mathematics*, 14(2):18–36, Spring 1992. ISSN 0272-8893. Institute of Education Sciences, ERIC, publisher.

Nathalie Sinclair and Victoria Guyevskey. Geometric Coding Environment: Integration of Computational Thinking and Mathematics in Primary School. `http://www.sfu.ca/education.html`, 2018. URL `http://ctmath.ca/wp-content/uploads/2018/02/Guyevskey-CT-Symposium-Poster.pdf`. Simon Fraser University (SFU). Accessed: 2021-05-31.

Nathalie Sinclair, Maria G. Bartolini Bussi, Michael de Villiers, Keith Jones, Ulrich Kortenkamp, Allen Leung, and Kay Owens. Geometry Education, Including the Use of New Technologies: A Survey of Recent Research. In Gabriele Kaiser, editor, *Proceedings of the 13th International Congress on Mathematical Education (ICME-13 Monographs)*, pages 277–287, ICME-13 Hamburg 2016, 2017. SpringerOpen, Cham. ISBN 978-3-319-62596-6, 978-3-319-62597-3. doi: 10.1007/978-3-319-62597-3_18. ISSN 2520-8322, 2520-8330.

Henrik Singmann and David Kellen. An Introduction to Mixed Models for Experimental Psychology. In Daniel Spieler and Eric Schumacher, editors, *New Methods in Cognitive Psychology*, chapter 1, pages 4–31. Routledge, Taylor & Francis Group (NY and London), Oct. 2019. ISBN 978-1-84872-630-7, 978-1-84872-631-4. doi: 10.4324/9780429318405. Frontiers of Cognitive Psychology. preprint: http://singmann.org/download/publications/singmann_kellen-introduction-mixed-models.pdf. Accessed: 2021-05-31.

Anders Skrondal and Sophia Rabe-Hesketh. Prediction in multilevel generalized linear models. *Journal of the Royal Statistical Society. Series A: Statistics in Society*, 172(3):659–687, Jun. 2009. ISSN 0964-1998. doi: 10.1111/j.1467-985X.2009.00587.x. URL http://www.gllamm.org/JRSSApredict_09.pdf. Wiley Online Library, publisher. Accessed: 2020-04-19.

Rachael Smyth and Andrew Johnson. Item Analysis. Western University, London, Ontario, CA, https://www.uwo.ca/fhs/tc/labs/09.ItemAnalysis.pdf, Oct. 2015. Number of pages: 9. Accessed: 2020-05-14.

Bridget Somekh. Designing software to maximize learning. *Research in Learning Technology Journal (RLT-J, https://journal.alt.ac.uk/index.php/rlt/about)*, 4(3):4–16, 1996. ISSN 0968-7769, 1741-1629. doi: 10.3402/rlt.v4i3.9974. URL https://journal.alt.ac.uk/index.php/rlt/article/view/827/1078. Association for Learning Technology (ALT), publisher. Accessed: 2021-05-31.

Erica D. Spangenberg and Gabrielle De Freitas. Mathematics teachers' levels of technological pedagogical content knowledge and information and communication technology integration barriers. *Pythagoras - Journal of the Association for Mathematics Education of South Africa*, 40(1), Dec. 2019. ISSN 1012-2346, 2223-7895. doi: 10.4102/pythagoras.v40i1.431. URL https://files.eric.ed.gov/fulltext/EJ1241178.pdf. AOSIS, 15 Oxford Street, Durbanville, Cape Town, 7550 South Africa. Accessed: 2021-05-31.

Natasha M. Speer and Joseph F. Wagner. Knowledge Needed by a Teacher to Provide Analytic Scaffolding during Undergraduate Mathematics Classroom Discussions. *Journal for Research in Mathematics Education*, 40(5):530–530, Nov. 2009. ISSN 00218251, 19452306. doi: 10.5951/jresematheduc.40.5.0530. URL http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.611.691&rep=rep1&type=pdf. The National Council of Teachers of Mathematics, publisher. Accessed: 2021-05-31.

Marthese Spiteri and Shu-Nu Chang Rundgren. Literature Review on the Factors Affecting Primary Teachers' Use of Digital Technology. *Technology, Knowledge and Learning*, 25(1): 115–128, Jul. 2018. ISSN 2211-1670. doi: 10.1007/s10758-018-9376-x. URL https://rdcu.be/cf1si. Springer Nature Switzerland AG, publisher. Department of Education, Stockholm University, Sweden. Accessed: 2021-05-31.

Elizabeth Stacey and Philippa Gerbic. Teaching for blended learning – Research perspectives from on-campus and distance students. *Education and Information Technologies*, 12(3):165–174, Jun. 2007. ISSN 1573-7608. doi: 10.1007/s10639-007-9037-5. Springer Nature Switzerland AG, publisher.

Gerry Stahl and The VMT Project Team. Dynamic-Geometry Activities with GeoGebra for Virtual Math Teams (VMT), Jan. 2013. URL http://GerryStahl.net/vmt/activities.pdf. The Math Forum at Drexel University. Accessed: 2020-08-26.

Heather Staker and Michael B. Horn. Classifying K-12 Blended Learning. Technical report, The Clayton Christensen Institute, formerly Innosight Institute, Inc (https://www.christenseninstitute.org), San Mateo, California, May 2012. URL https://files.eric.ed.gov/fulltext/ED535180.pdf. Number of pages: 22. Accessed: 2021-02-02.

Institut national de la statistique et des études économiques STATEC. Indice Socio-Économique par Commune. 64(2), Jul. 2017. ISSN 0076-1583. URL https://statistiques.public.lu/catalogue-publications/bulletin-Statec/2017/PDF-Bulletin2-2017.pdf. Dr Allegrezza, Serge. Centre Administratif Pierre Werner, L-1468 LUXEMBOURG. Accessed: 2020-10-20.

Institut national de la statistique et des études économiques STATEC. Atlas démographique du Luxembourg. 2(1), Feb. 2019. ISSN 2658-9648. URL https://integratioun.lu/wp-content/uploads/2019/03/AtlasDemographiqueLuxembourg.pdf. Dr Allegrezza, Serge. Centre Administratif Pierre Werner, L-1468 LUXEMBOURG. Accessed: 2020-10-20.

Rudolf Straesser. Cabri-géomètre: Does Dynamic Geometry Software (DGS) Change Geometry and its Teaching and Learning? *International Journal of Computers for Mathematical Learning*, 6(3):319–333, Jan. 2002. ISSN 1573-1766. doi: 10.1023/A:1013361712895. Springer Nature Switzerland AG, publisher.

Walter W. Stroup. *Generalized Linear Mixed Models: Modern Concepts, Methods and Applications*. CRC Press, Taylor & Francis Group, 1 edition, Oct. 2012. ISBN 978-1-4398-1512-0. doi: 10.1201/b13151. Chapman & Hall/CRC Texts in Statistical Science. Number of pages: 555.

TAOtesting. The Importance of IMS Certification and QTI Compliance. Open Assessment Technologies S.A. (OAT), https://www.taotesting.com/blog/importance-ims-certification-qti-compliance/, Aug. 2018. Capellen, Luxembourg. Accessed: 2021-05-31.

TAOtesting. Redefining Digital Assessment - Open Source. Open Standards. Open Possibilities., TAO – the Leading Assessment Platform for Education, Career, and Industry Advancement. Open Assessment Technologies S.A. (OAT), https://www.taotesting.com/wp-content/uploads/2019/05/TAO_Data_Sheet_2017.pdf, 2019. Capellen, Luxembourg. Accessed: 2020-01-10.

Mohsen Tavakol and Reg Dennick. Making sense of Cronbach's alpha. *International Journal of Medical Education (IJME)*, 2:53–55, Jun. 2011. ISSN 2042-6372. doi: 10.5116/ijme.4dfb.8dfd. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4205511/pdf/ijme-2-53.pdf. Accessed: 2021-05-31.

Melissa Taylor. *Book Love: Help Your Child Grow from Reluctant to Enthusiastic Reader*. Imagination Soup, Oct. 2012. ISBN 978-0-9884124-1-5. language: en-US.

Elli Theobald. Students Are Rarely Independent: When, Why, and How to Use Random Effects in Discipline-Based Education Research. *LSE: CBE—Life Sciences Education (Research Methods)*, 17(3):1–12, Sept. 2018. doi: 10.1187/cbe.17-12-0280. Erin L. Dolan, monitoring editor. The American Society for Cell Biology (ascb), and CBE-Life Sciences Education (LSE). Bethesda, MD 20852, USA.

Sue Thomson. Achievement at school and socioeconomic background — an educational perspective. *Nature Partner Journals (npj) Science of Learning*, 3:5, Mar. 2018. ISSN 2056-7936. doi: 10.1038/s41539-018-0022-0. Nature Publishing Group.

Catlin R. Tucker. *Blended Learning in Grades 4–12: Leveraging the Power of Technology to Create Student-Centered Classrooms*. Corwin Press, A SAGE Publications Company, Thousand Oaks, CA, 1 edition, Jun. 2012. ISBN 978-1-4522-4086-2. Corwin Teaching Essentials, collection. Number of pages: 272.

U.S. General Services Administration. Visual Design Basics. usability.gov, Improving the User Experience, n.d. URL https://www.usability.gov/what-and-why/visual-design.html. Accessed: 2020-03-26.

User Guide TAOtesting. User Guide 3.0 - Creating Test-takers. Open Assessment Technologies S.A. (OAT), https://www.taotesting.com/user-guide/users/test-takers/creating-test-taker/, 2015. Capellen, Luxembourg. Accessed: 2021-05-31.

User Guide TAOtesting. User Guide 3.1. Open Assessment Technologies S.A. (OAT), 2016. Capellen, Luxembourg.

User Guide TAOtesting. User Guide 3.2 - Portable Custom (PC) Interactions. Open Assessment Technologies S.A. (OAT), https://www.taotesting.com/user-guide/users/interactions/portable-custom-interactions/, 2017. Capellen, Luxembourg. Accessed: 2021-05-31.

User Guide TAOtesting. User Guide 3.3 - Creating a New Delivery. Open Assessment Technologies S.A. (OAT), https://www.taotesting.com/user-guide/users/deliveries/create-a-new-delivery/, 2019. Capellen, Luxembourg. Accessed: 2021-05-31.

Rachel R. Van Der Stuyf. Scaffolding as a Teaching Strategy. *Adolescent learning and development*, 52(3):5–18, Nov. 2002.

Daisy E. Vasquez. *Enhancing Student Achievement Using GeoGebra in a Technology Rich Environment*. PhD thesis, California State Polytechnic University, Pomona, 2015. URL https://core.ac.uk/download/pdf/48498993.pdf. Accessed: 2021-01-29.

Rolland Viau. *La motivation en contexte scolaire*. De Boeck, 5 edition, May 2009. ISBN 978-2-8041-1148-9.

Olga Viberg, Åke Grönlund, and Annika Andersson. Integrating digital technology in mathematics education: a Swedish case study. *Interactive Learning Environments (https://www.tandfonline.com/loi/nile20)*, 28:1–12, Jun. 2020. doi: 10.1080/10494820.2020.1770801. Informa UK Limited, Routledge, Taylor & Francis Group, publisher.

Iva Vojkůvková. The van Hiele Model of Geometric Thinking. *'Week of Doctoral Students 2012' (WDS'12) Proceedings of Contributed Papers*, 1:72–75, 2012. ISBN 978-80-7378-224-5. MATFYZPRESS, publisher. Accessed: 2021-01-29.

Fabio Votta. Multilevel Modelling mit R. RPubs by RStudio: R-Group Stuttgart, Jan. 2017. URL https://rpubs.com/favstats/multilevel. Accessed: 2020-02-26.

Rose K. Vukovic, Nonie K. Lesaux, and Linda S. Siegel. The mathematics skills of children with reading difficulties. *Elsevier, Learning and Individual Differences*, 20(6):639–643, Dec. 2010. ISSN 1041-6080. doi: 10.1016/j.lindif.2010.08.004. URL http://www.sciencedirect.com/science/article/pii/S1041608010000865. ScienceDirect, publisher. Accessed: 2020-10-10.

Markus Welz. Spot on the Mouse, n.d. URL http://www.markuswelz.de/software/. Accessed: 2020-09-10.

Wayne A. Wickelgren. *Models of Human Memory*, chapter 3. Multitrace Strength Theory, pages 65–102. Academic Press New York and London, 1970. ISBN 978-1-4832-5820-1.

Glenn Williams. R for Psych, Chapter 9: Mixed Effects Models. bookdown.org, Aug. 2018. URL https://glennwilliams.me/r4psych/mixed-effects-models.html. Accessed: 2020-02-17.

Helen J. Williams. *The relevance of role play to the learning of mathematics in the primary classroom*. PhD thesis, School of Education, University of Roehampton, Jun. 2014. URL https://pure.roehampton.ac.uk/portal/en/studentTheses/the-relevance-of-role-play-to-the-learning-of-mathematics-in-the-. language: en-UK. Accessed: 2021-02-28.

Jeannette M. Wing. Computational Thinking. *Communications of the ACM (the Association for Computing Machinery)*, 49(3):33–35, Mar. 2006. doi: 10.1145/1118178.1118215. URL https://www.cs.cmu.edu/~15110-s13/Wing06-ct.pdf. ACM (DL) Digital Library, publisher. Accessed: 2021-05-31.

Jeannette M. Wing. Computational thinking and thinking about computing. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 366 (1881):3717–3725, Jul. 2008. ISSN 1364-503X, 1471-2962. doi: 10.1098/rsta.2008.0118. The Royal Society London, publisher.

Rainer Winkler and Matthias Söllner. Unleashing the Potential of Chatbots in Education: A State-Of-The-Art Analysis. In *Academy of Management Annual Meeting (AOM)*, Chicago, USA, Aug. 2018. URL https://www.alexandria.unisg.ch/254848/1/JML_699.pdf. Accessed: 2021-05-31.

Bodo Winter. Linear models and linear mixed effects models in R with linguistic applications. *Cornell University, arXiv preprint arXiv:1308.5499*, Aug. 2013. URL https://arxiv.org/ftp/arxiv/papers/1308/1308.5499.pdf. Number of pages: 42. Accessed: 2020-04-21.

David Wood, Jerome S. Bruner, and Gail Ross. The Role of Tutoring in Problem Solving. *Journal of Child Psychology and Psychiatry*, 17(2):89–100, Apr. 1976. doi: 10.1111/j.1469-7610.1976.tb00381.x. Pergamon Press, Wiley Online Library. Nottingham, Oxford and Harvard, GB and MA.

Mohd Azli Yeop, Kung-Teck Wong, and Pauline Swee Choo Goh. Blended learning: pedagogy, learning styles, and assessment activities in the classroom. *International Journal of Advanced and Applied Sciences (http://www.science-gate.com/IJAAS.html)*, 3(11):36–39, Nov. 2016. ISSN 2313-626X. doi: 10.21833/ijaas.2016.11.007. Institute of Advanced Studies in Education (IASE), publisher.

Sarah Younie, Marilyn Leask, and Kevin Burden. *Teaching and Learning with ICT in the Primary School*. Routledge, Taylor & Francis Group (London and NY), 2 edition, Sept. 2014. ISBN 978-1-138-78315-7, 978-1-138-78314-0. Number of pages: 268.

Allyson Zalewski and Miranda Cairns. Aligning With the School Day With Y4Y Click and Go. You for Youth (Y4Y), Illinois Quality Afterschool Program: 2017 Illinois 21st CCLC Spring Conference, 2011 & 2015. URL https://iqa.airprojects.org/events/archive/2017-spring-conference/y4y_click_and_go.pdf. U.S. Department of Education, Office of Elementary and Secondary Education. Accessed: 2020-09-10.

Giovanna Zanolla et al. The gender gap in math. Evidences of a study in the primary school in the Swiss canton of Ticino. *International Journal of Teaching and Education*, 6(1):103–125, Apr. 2018. doi: 10.20472/TE.2018.6.1.007. URL https://ideas.repec.org/a/sek/jijote/v6y2018i1p103-125.html. IDEAS, bibliographic database dedicated to Economics, International Institute of Social and Economic Sciences (http://www.iises.net/), publisher. Accessed: 2021-02-03.

Appendix A

Socioeconomic index by Commune in 2017

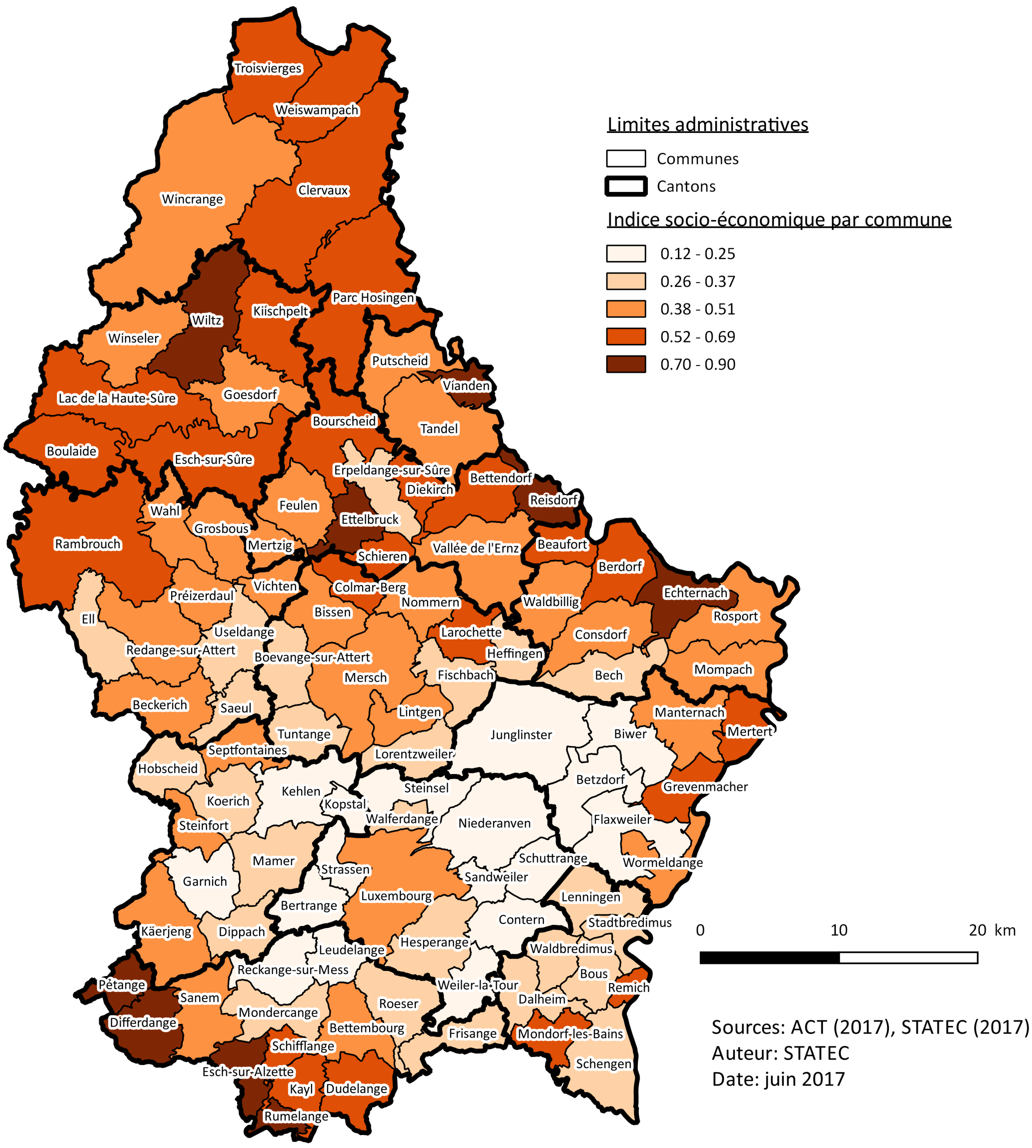


Figure A.1: Socioeconomic index by Commune in 2017 (STATEC, 2017, pg. 26)

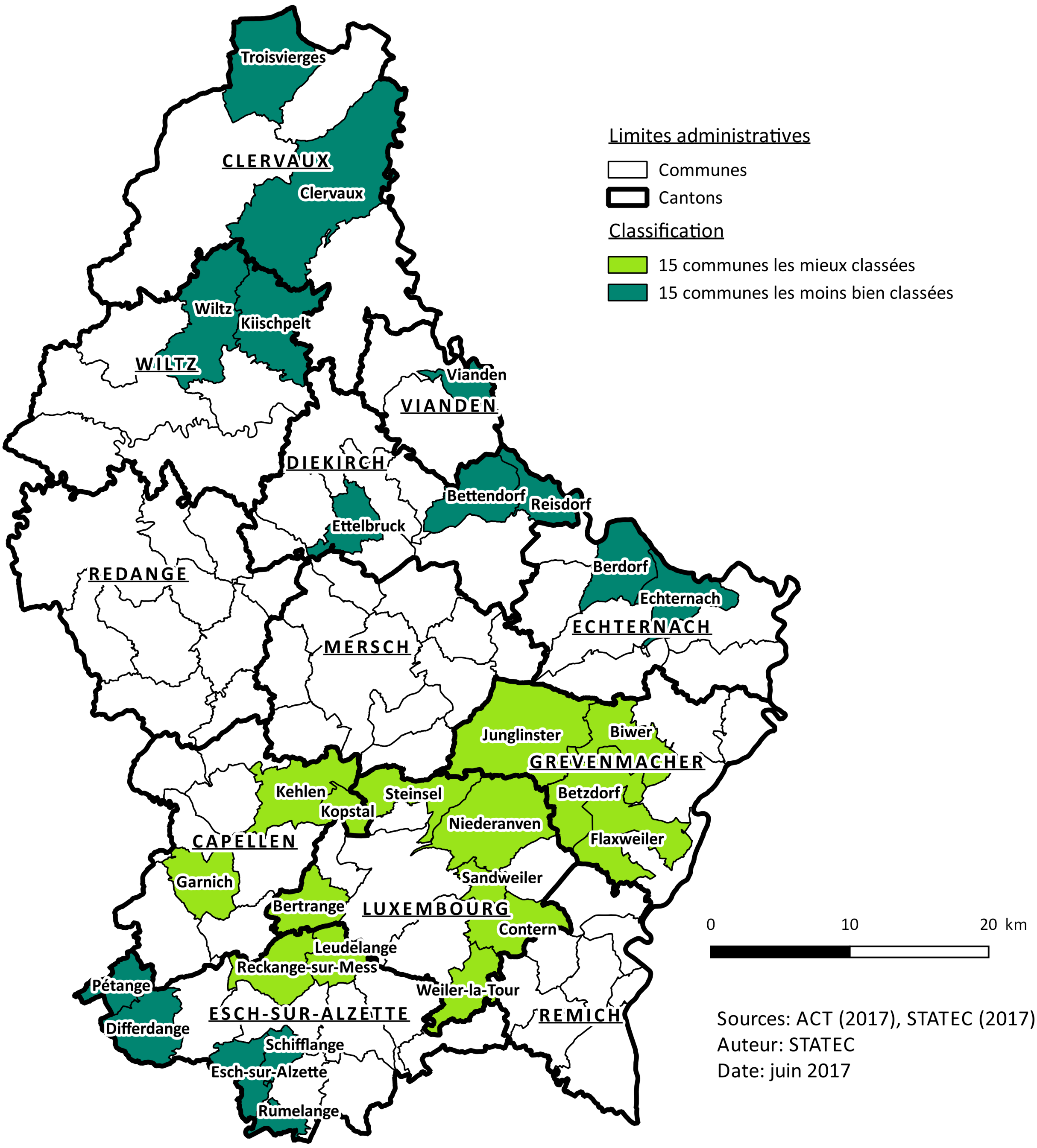


Figure A.2: Socioeconomic index by Commune in 2017 (synthesis) (STATEC, 2017, pg. 27)

Appendix B

Proportion of foreigners per Commune 2018

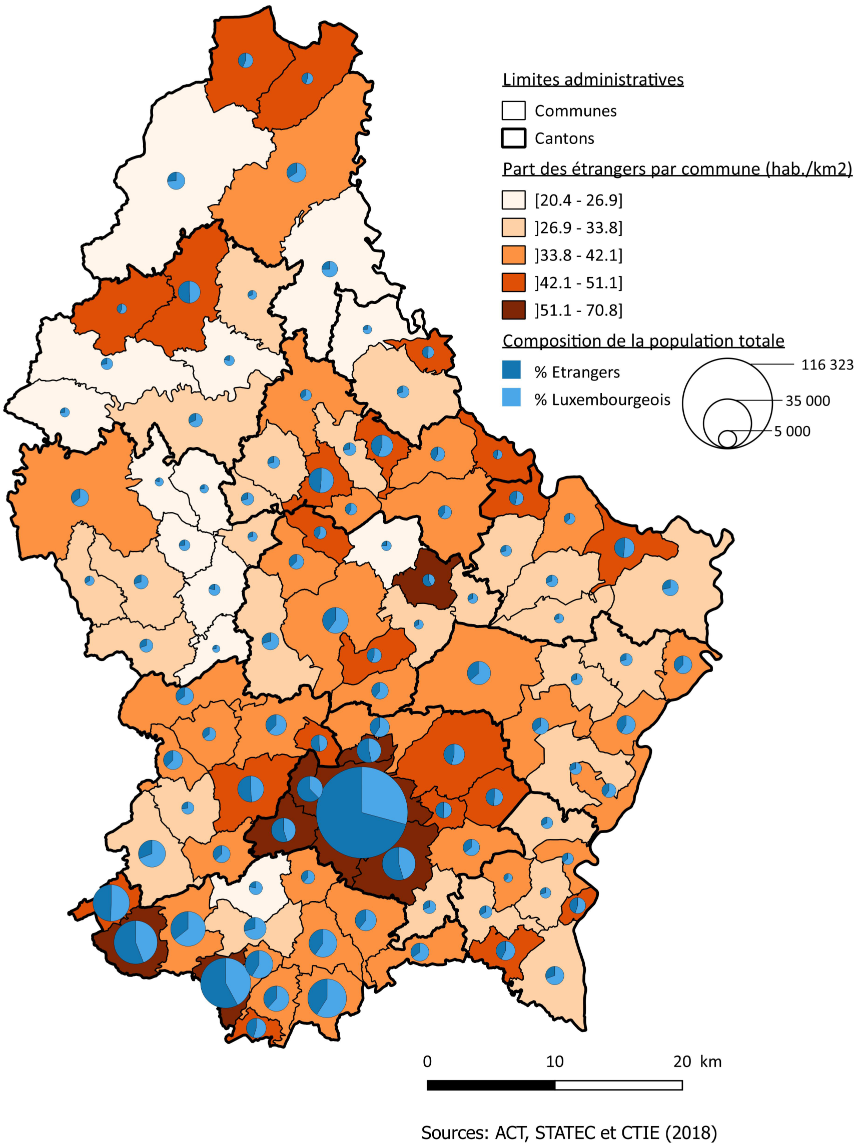


Figure B.1: Proportion of foreigners per Commune 2018 (STATEC, 2019, pg. 19)

Appendix C

Children's orientation after fundamental school

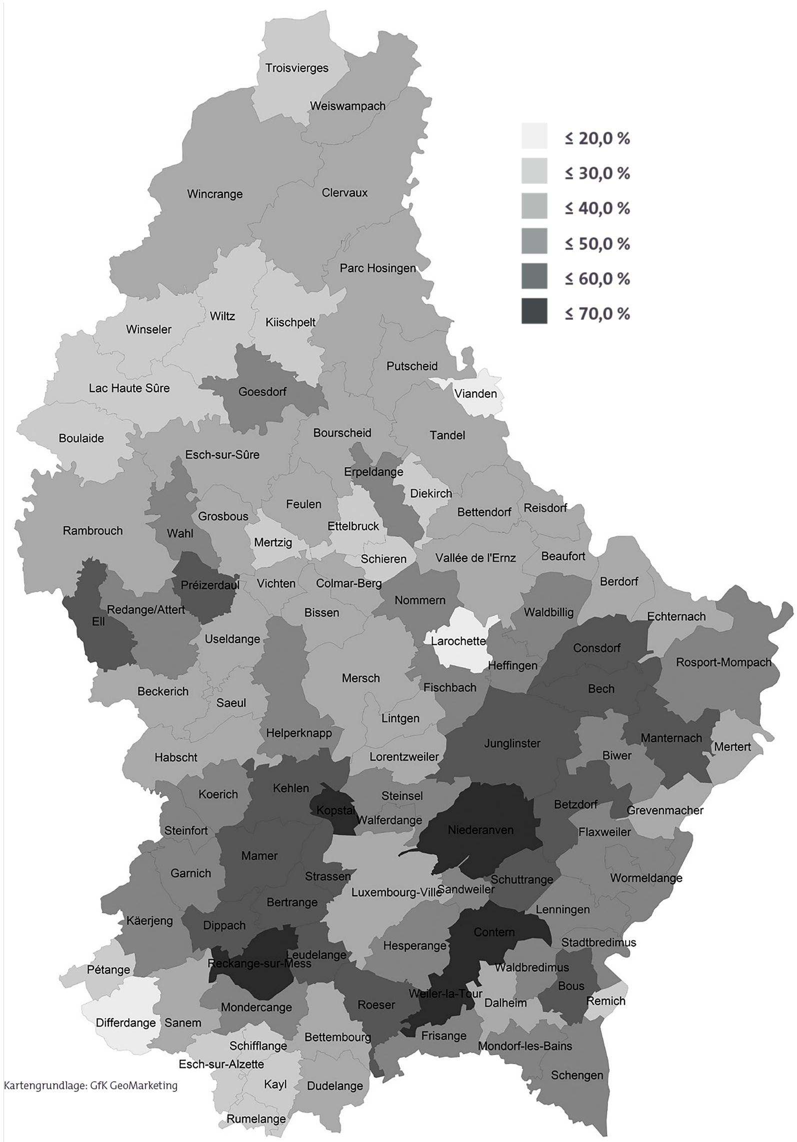


Figure C.1: Proportions of children's orientation to **classic** secondary education during the school years 2009/2010 to 2016/2017 (Lenz et al., 2018, pg. 31)

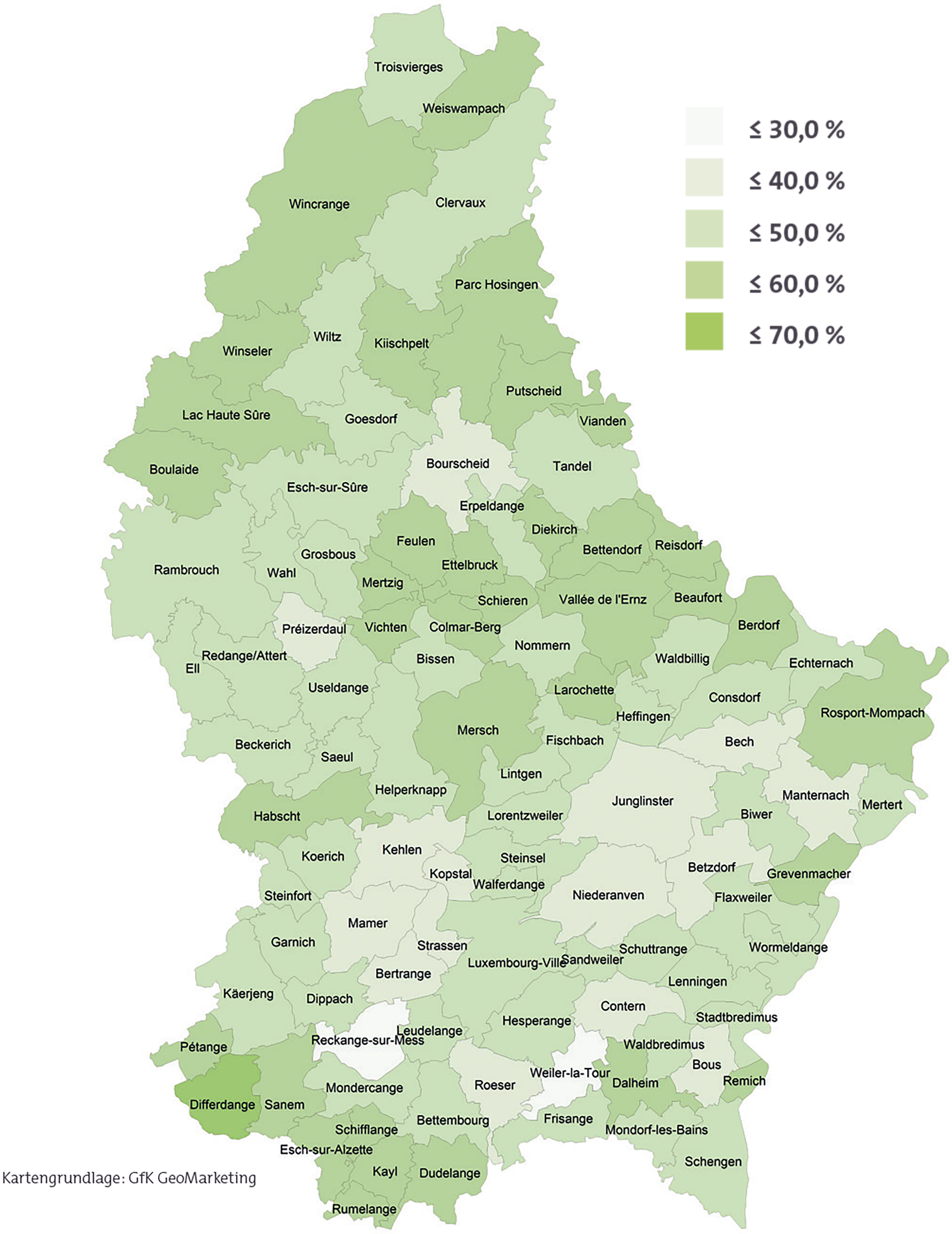
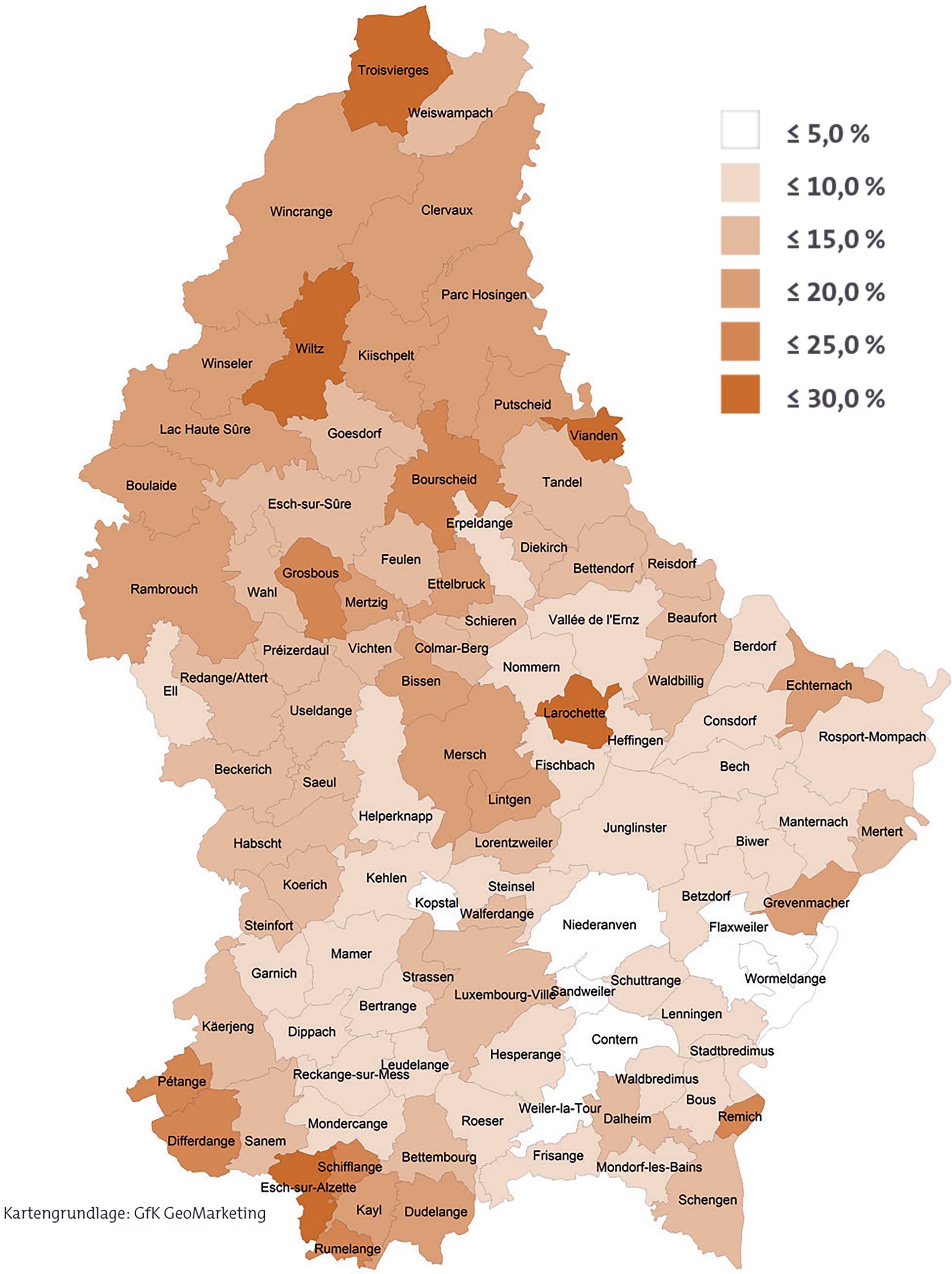


Figure C.2: Proportions of children's orientation to **general** secondary education **guidance route** during the school years 2009/2010 to 2016/2017 (Lenz et al., 2018, pg. 32)



Kartengrundlage: GfK GeoMarketing

Figure C.3: Proportions of children's orientation to **general** secondary education **preparatory route** during the school years 2009/2010 to 2016/2017 (Lenz et al., 2018, pg. 32)

Appendix D

Workflows defined by the project GEOGEBRATAO

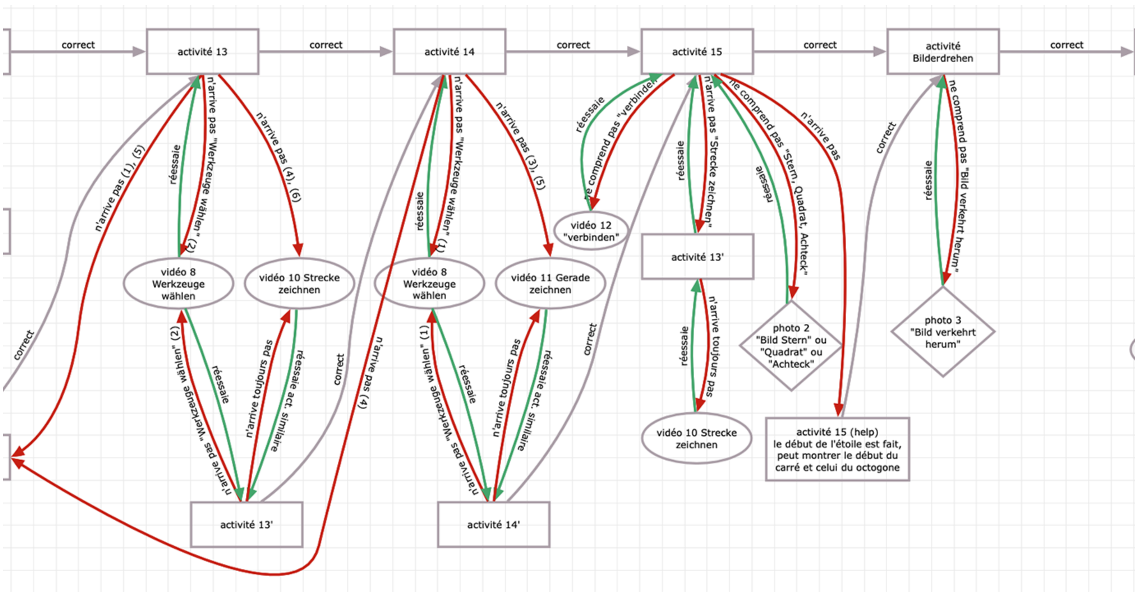


Figure D.1: Extract of the initial version of the GEOGEBRATAO tool's workflow;
well-developed version, for internal use only, for programming the individual exploratory learning assignments

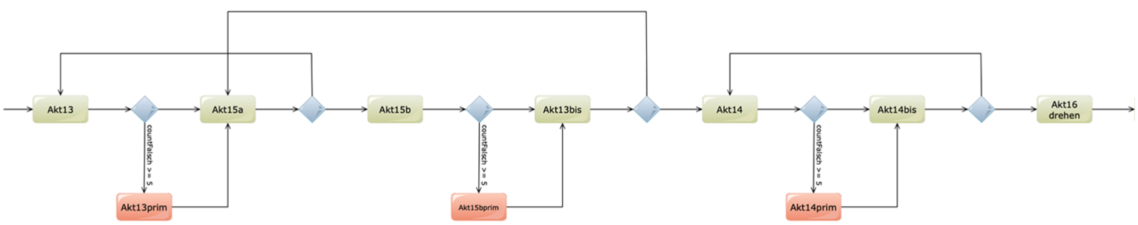


Figure D.2: The same extract as in Figure D.1, but workflow less developed, clearer for the integration of the learning assignments into the computer-assisted testing framework TAO. Some activities were split to facilitate their integration and the order of some activities inside the sequence was changed for logical structure.

Appendix E

Reflected post-test data and fitting its distribution

We use a Generalized Mixed-Effects Model to compare the complete binary post-test data set to the restricted one, because Generalized Mixed-Effects Models allow response variables from different distributions. Our response variable, the post-test data (success scores), is data distorted from normality and its distribution is mostly left-skewed, which means that many large data values are concentrated together at the top of the distribution and fewer small values spread out to the left.

To specify an appropriate distributional *family* for our planned model, we reflect the success scores (*Pct*) and generate a new response variable $(100,01-Pct)$, called *Pct_refl*. *Pct_refl* represents the failure rates in the post-test in percent to within $0,01\%$ upward. Its distribution is moderately right-skewed (skewness equal to $0,8655$). This allows us to specify $family = Gamma$ for the Generalized Mixed-Effects Model we implement and to apply a *log* link function (therefore $(Pct\_refl = 100,01-Pct)$ and not $(100-Pct)$).

In Figure E.1, we chose *Gamma* to fit the distribution of *Pct_refl*.

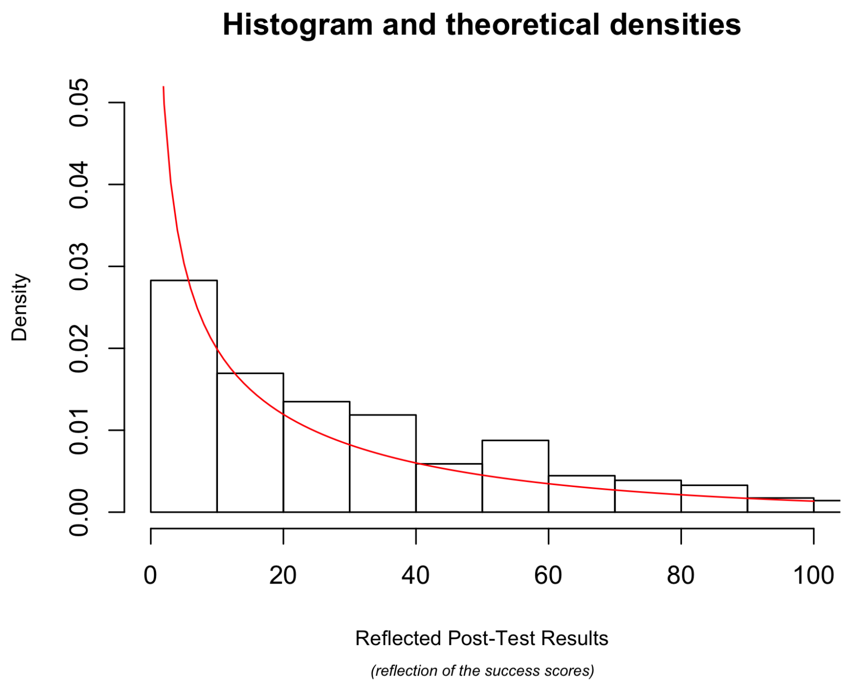


Figure E.1: *Gamma* fitting the distribution of the failure rates in the post-test in percent (*Pct_refl*)

Appendix F

Outputs of model $glmer\_Comp\_Post$

F.1 Analysis of Deviance Table

```
ANOVA_glmer_Comp_Post

## Analysis of Deviance Table (Type II Wald chisquare tests)
##
## Response: Pct_refl
##                          Chisq Df Pr(>Chisq)
## Restrict                4.5445  1    0.03302 *
## Group                   0.7237  1    0.39493
## Domain                447.2210  5  < 2.2e-16 ***
## Restrict:Group          1.9990  1    0.15740
## Restrict:Domain         2.5765  5    0.76494
## Group:Domain           40.7682  5  1.045e-07 ***
## Restrict:Group:Domain   1.1384  5    0.95065
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

F.2 Summary of fitted model

```
Summary_glmer_Comp_Post

## Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
##  Family: Gamma  ( log )
## Formula: Pct_refl ~ Restrict * Group * Domain + (1 | ClassNbr)
##    Data: .
## Control: glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 1e+05))
##
##      AIC      BIC   logLik deviance df.resid
##  23374.9  23529.5 -11661.4  23322.9     2806
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.1374 -0.6879 -0.1670  0.3765  6.4261
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  ClassNbr (Intercept) 0.0536   0.2315
##  Residual             0.7728   0.8791
## Number of obs: 2832, groups:  ClassNbr, 17
##
```

```
## Fixed effects:
##                                        Estimate Std. Error t value Pr(>|z|)
## (Intercept)                           3.325e+00  1.907e-01  17.435  < 2e-16 ***
## RestrictTRUE                          3.823e-06  2.129e-01   0.000  0.99999
## Grouptreatment                        6.279e-02  2.281e-01   0.275  0.78308
## DomainCo                             -1.449e+00  2.161e-01  -6.703 2.04e-11 ***
## DomainDS                              9.799e-01  2.135e-01   4.589 4.46e-06 ***
## DomainFi                             -1.484e-01  2.146e-01  -0.691  0.48939
## DomainLS                             -3.963e-01  2.145e-01  -1.848  0.06464 .
## DomainRS                             -1.316e-01  2.134e-01  -0.617  0.53748
## RestrictTRUE:Grouptreatment          -2.917e-01  2.554e-01  -1.142  0.25340
## RestrictTRUE:DomainCo                 6.707e-06  3.014e-01   0.000  0.99998
## RestrictTRUE:DomainDS                -4.030e-05  3.015e-01   0.000  0.99989
## RestrictTRUE:DomainFi                -1.230e-06  3.015e-01   0.000  1.00000
## RestrictTRUE:DomainLS                -2.318e-06  3.013e-01   0.000  0.99999
## RestrictTRUE:DomainRS                 7.881e-06  3.015e-01   0.000  0.99998
## Grouptreatment:DomainCo               7.562e-01  2.584e-01   2.926  0.00343 **
## Grouptreatment:DomainDS              -1.943e-01  2.561e-01  -0.759  0.44804
## Grouptreatment:DomainFi               3.016e-01  2.571e-01   1.173  0.24072
## Grouptreatment:DomainLS              -7.122e-02  2.570e-01  -0.277  0.78164
## Grouptreatment:DomainRS               4.678e-02  2.561e-01   0.183  0.85502
## RestrictTRUE:Grouptreatment:DomainCo  2.917e-01  3.615e-01   0.807  0.41973
## RestrictTRUE:Grouptreatment:DomainDS  1.743e-01  3.616e-01   0.482  0.62982
## RestrictTRUE:Grouptreatment:DomainFi  1.669e-01  3.617e-01   0.462  0.64442
## RestrictTRUE:Grouptreatment:DomainLS  2.347e-01  3.615e-01   0.649  0.51622
## RestrictTRUE:Grouptreatment:DomainRS -5.312e-03  3.617e-01  -0.015  0.98828
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Appendix G

Outputs of model $glmer\_Comp\_PrePost$

G.1 Analysis of Deviance Table

```
ANOVA_glmer_Comp_PrePost

## Analysis of Deviance Table (Type II Wald chisquare tests)
##
## Response: Pct_refl
##                         Chisq Df Pr(>Chisq)
## Test_3               455.8082  2  < 2.2e-16 ***
## Group                  0.2889  1    0.59096
## Domain              1637.2614  5  < 2.2e-16 ***
## Test_3:Group           6.1185  2    0.04692 *
## Test_3:Domain         62.1288 10  1.430e-09 ***
## Group:Domain          62.4378  5  3.806e-12 ***
## Test_3:Group:Domain   21.7041 10    0.01668 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

G.2 Summary of fitted model

```
Summary_glmer_Comp_PrePost

## Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
##  Family: Gamma  ( log )
## Formula: Pct_refl ~ Test_3 * Group * Domain + (1 | ClassNbr/ChildInClass)
##    Data: .
## Control: glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 5e+05))
##
##      AIC      BIC   logLik deviance df.resid
##  73090.4  73365.3 -36506.2  73012.4     8457
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.3566 -0.5953 -0.0960  0.4062  6.4496
##
## Random effects:
##  Groups                Name        Variance Std.Dev.
##  ChildInClass:ClassNbr (Intercept) 0.10791  0.3285
##  ClassNbr              (Intercept) 0.02636  0.1624
##  Residual                          0.54328  0.7371
## Number of obs: 8496, groups:  ChildInClass:ClassNbr, 236; ClassNbr, 17
```

```
##
## Fixed effects:
##                                    Estimate Std. Error t value Pr(>|z|)
## (Intercept)                         3.64969    0.14762  24.723  < 2e-16 ***
## Test_3Post                         -0.38475    0.13584  -2.832 0.004619 **
## Test_3Rest                         -0.38472    0.13576  -2.834 0.004599 **
## Grouptreatment                      0.05559    0.17643   0.315 0.752680
## DomainCo                           -0.73171    0.13709  -5.337 9.43e-08 ***
## DomainDS                            0.75326    0.13561   5.554 2.79e-08 ***
## DomainFi                            0.24963    0.13605   1.835 0.066518 .
## DomainLS                           -0.22209    0.13597  -1.633 0.102395
## DomainRS                           -0.18048    0.13564  -1.331 0.183319
## Test_3Post:Grouptreatment          -0.02421    0.16286  -0.149 0.881806
## Test_3Rest:Grouptreatment          -0.33475    0.16295  -2.054 0.039948 *
## Test_3Post:DomainCo                -0.80926    0.19513  -4.147 3.36e-05 ***
## Test_3Rest:DomainCo                -0.80930    0.19508  -4.149 3.35e-05 ***
## Test_3Post:DomainDS                 0.26755    0.19158   1.397 0.162548
## Test_3Rest:DomainDS                 0.26751    0.19160   1.396 0.162657
## Test_3Post:DomainFi                -0.44925    0.19323  -2.325 0.020073 *
## Test_3Rest:DomainFi                -0.44926    0.19316  -2.326 0.020025 *
## Test_3Post:DomainLS                -0.24081    0.19227  -1.252 0.210414
## Test_3Rest:DomainLS                -0.24085    0.19223  -1.253 0.210229
## Test_3Post:DomainRS                 0.04197    0.19172   0.219 0.826726
## Test_3Rest:DomainRS                 0.04192    0.19169   0.219 0.826879
## Grouptreatment:DomainCo             0.17653    0.16443   1.074 0.283020
## Grouptreatment:DomainDS            -0.01408    0.16275  -0.087 0.931051
## Grouptreatment:DomainFi            -0.01589    0.16306  -0.097 0.922394
## Grouptreatment:DomainLS            -0.03413    0.16324  -0.209 0.834390
## Grouptreatment:DomainRS            -0.02228    0.16280  -0.137 0.891152
## Test_3Post:Grouptreatment:DomainCo  0.47529    0.23355   2.035 0.041846 *
## Test_3Rest:Grouptreatment:DomainCo  0.78584    0.23356   3.365 0.000766 ***
## Test_3Post:Grouptreatment:DomainDS -0.14266    0.22980  -0.621 0.534732
## Test_3Rest:Grouptreatment:DomainDS  0.01388    0.22999   0.060 0.951893
## Test_3Post:Grouptreatment:DomainFi  0.33892    0.23121   1.466 0.142697
## Test_3Rest:Grouptreatment:DomainFi  0.51709    0.23128   2.236 0.025365 *
## Test_3Post:Grouptreatment:DomainLS -0.09472    0.23064  -0.411 0.681304
## Test_3Rest:Grouptreatment:DomainLS  0.14691    0.23092   0.636 0.524653
## Test_3Post:Grouptreatment:DomainRS  0.09453    0.23004   0.411 0.681131
## Test_3Rest:Grouptreatment:DomainRS  0.08857    0.23015   0.385 0.700346
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Appendix H

Outputs of model $glmer\_Comp\_PrePost\_Gen$

H.1 Analysis of Deviance Table

```
ANOVA_glmer_Comp_PrePost_Gen

## Analysis of Deviance Table (Type II Wald chisquare tests)
##
## Response: Pct_refl
##                                Chisq Df Pr(>Chisq)
## Test_3                      462.6349  2  < 2.2e-16 ***
## Domain                     1650.7032  5  < 2.2e-16 ***
## Test_3:Domain                64.7863 10  4.453e-10 ***
## Group:Gender                  0.9717  3     0.8081
## Test_3:Group:Gender           6.7073  6     0.3488
## Domain:Group:Gender          87.5064 15  2.888e-12 ***
## Test_3:Domain:Group:Gender   35.8754 30     0.2122
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

H.2 Summary of fitted model

```
Summary_glmer_Comp_PrePost_Gen

## Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
##  Family: Gamma  ( log )
## Formula: Pct_refl ~ Test_3 * Domain * Group:Gender + (1 | ClassNbr/ChildInClass)
##    Data: .
## Control: glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 5e+05))
##
##      AIC      BIC   logLik deviance df.resid
##  73127.5  73656.0 -36488.7  72977.5     8421
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.3553 -0.5886 -0.0964  0.3947  6.5980
##
## Random effects:
##  Groups                 Name        Variance Std.Dev.
##  ChildInClass:ClassNbr (Intercept) 0.10959  0.3310
##  ClassNbr              (Intercept) 0.02701  0.1643
##  Residual                          0.54432  0.7378
## Number of obs: 8496, groups:  ChildInClass:ClassNbr, 236; ClassNbr, 17
```

```
##
## Fixed effects:
##                                             Estimate Std. Error t value Pr(>|z|)
## (Intercept)                                3.640e+00  1.710e-01  21.287  < 2e-16 ***
## Test_3Post                                -3.766e-01  1.716e-01  -2.195 0.028162 *
## Test_3Rest                                -3.766e-01  1.710e-01  -2.202 0.027663 *
## DomainCo                                  -8.121e-01  1.742e-01  -4.662 3.14e-06 ***
## DomainDS                                   7.854e-01  1.720e-01   4.567 4.96e-06 ***
## DomainFi                                   2.483e-01  1.733e-01   1.433 0.151867
## DomainLS                                  -2.052e-01  1.732e-01  -1.185 0.236061
## DomainRS                                  -2.042e-01  1.719e-01  -1.188 0.234918
## Test_3Post:DomainCo                       -1.035e+00  2.496e-01  -4.147 3.37e-05 ***
## Test_3Rest:DomainCo                       -1.035e+00  2.492e-01  -4.153 3.27e-05 ***
## Test_3Post:DomainDS                        2.136e-01  2.437e-01   0.876 0.380782
## Test_3Rest:DomainDS                        2.136e-01  2.432e-01   0.878 0.379787
## Test_3Post:DomainFi                       -3.696e-01  2.466e-01  -1.499 0.133848
## Test_3Rest:DomainFi                       -3.696e-01  2.459e-01  -1.503 0.132733
## Test_3Post:DomainLS                       -2.112e-01  2.456e-01  -0.860 0.389959
## Test_3Rest:DomainLS                       -2.112e-01  2.455e-01  -0.860 0.389789
## Test_3Post:DomainRS                        9.827e-02  2.432e-01   0.404 0.686155
## Test_3Rest:DomainRS                        9.826e-02  2.432e-01   0.404 0.686214
## Grouptreatment:GenderG                     7.594e-02  2.089e-01   0.364 0.716209
## Groupcontrol:GenderG                       2.960e-02  2.174e-01   0.136 0.891666
## Grouptreatment:GenderB                     5.797e-02  2.077e-01   0.279 0.780131
## Test_3Post:Grouptreatment:GenderG         -1.176e-02  2.124e-01  -0.055 0.955850
## Test_3Rest:Grouptreatment:GenderG         -3.715e-01  2.117e-01  -1.755 0.079227 .
## Test_3Post:Groupcontrol:GenderG           -2.104e-02  2.664e-01  -0.079 0.937045
## Test_3Rest:Groupcontrol:GenderG           -2.107e-02  2.668e-01  -0.079 0.937057
## Test_3Post:Grouptreatment:GenderB         -5.216e-02  2.107e-01  -0.248 0.804453
## Test_3Rest:Grouptreatment:GenderB         -3.170e-01  2.102e-01  -1.508 0.131514
## DomainCo:Grouptreatment:GenderG            3.025e-01  2.153e-01   1.405 0.160055
## DomainDS:Grouptreatment:GenderG           -4.478e-02  2.129e-01  -0.210 0.833405
## DomainFi:Grouptreatment:GenderG            2.984e-03  2.141e-01   0.014 0.988881
## DomainLS:Grouptreatment:GenderG           -9.311e-02  2.143e-01  -0.434 0.663957
## DomainRS:Grouptreatment:GenderG            2.916e-05  2.125e-01   0.000 0.999891
## DomainCo:Groupcontrol:GenderG              1.793e-01  2.709e-01   0.662 0.508042
## DomainDS:Groupcontrol:GenderG             -7.678e-02  2.675e-01  -0.287 0.774100
## DomainFi:Groupcontrol:GenderG              9.369e-03  2.694e-01   0.035 0.972254
## DomainLS:Groupcontrol:GenderG             -4.100e-02  2.697e-01  -0.152 0.879178
## DomainRS:Groupcontrol:GenderG              5.269e-02  2.679e-01   0.197 0.844073
## DomainCo:Grouptreatment:GenderB            2.129e-01  2.140e-01   0.995 0.319950
## DomainDS:Grouptreatment:GenderB           -4.567e-02  2.114e-01  -0.216 0.828976
## DomainFi:Grouptreatment:GenderB           -3.103e-02  2.128e-01  -0.146 0.884060
## DomainLS:Grouptreatment:GenderB           -1.583e-02  2.128e-01  -0.074 0.940706
## DomainRS:Grouptreatment:GenderB            1.846e-03  2.110e-01   0.009 0.993020
## Test_3Post:DomainCo:Grouptreatment:GenderG  6.386e-01  3.074e-01   2.078 0.037727 *
## Test_3Rest:DomainCo:Grouptreatment:GenderG  9.984e-01  3.067e-01   3.255 0.001134 **
## Test_3Post:DomainDS:Grouptreatment:GenderG -1.027e-01  3.017e-01  -0.340 0.733498
## Test_3Rest:DomainDS:Grouptreatment:GenderG  9.320e-02  3.011e-01   0.310 0.756885
## Test_3Post:DomainFi:Grouptreatment:GenderG  2.077e-01  3.043e-01   0.683 0.494857
## Test_3Rest:DomainFi:Grouptreatment:GenderG  4.279e-01  3.035e-01   1.410 0.158581
## Test_3Post:DomainLS:Grouptreatment:GenderG  4.880e-02  3.038e-01   0.161 0.872389
## Test_3Rest:DomainLS:Grouptreatment:GenderG  3.561e-01  3.037e-01   1.173 0.240980
## Test_3Post:DomainRS:Grouptreatment:GenderG  1.556e-02  3.008e-01   0.052 0.958760
## Test_3Rest:DomainRS:Grouptreatment:GenderG  2.886e-02  3.009e-01   0.096 0.923591
## Test_3Post:DomainCo:Groupcontrol:GenderG    4.380e-01  3.854e-01   1.136 0.255819
## Test_3Rest:DomainCo:Groupcontrol:GenderG    4.380e-01  3.857e-01   1.135 0.256215
## Test_3Post:DomainDS:Groupcontrol:GenderG    1.319e-01  3.782e-01   0.349 0.727228
## Test_3Rest:DomainDS:Groupcontrol:GenderG    1.320e-01  3.784e-01   0.349 0.727253
## Test_3Post:DomainFi:Groupcontrol:GenderG   -2.106e-01  3.819e-01  -0.551 0.581302
## Test_3Rest:DomainFi:Groupcontrol:GenderG   -2.106e-01  3.826e-01  -0.550 0.582018
## Test_3Post:DomainLS:Groupcontrol:GenderG   -8.563e-02  3.804e-01  -0.225 0.821909
## Test_3Rest:DomainLS:Groupcontrol:GenderG   -8.562e-02  3.818e-01  -0.224 0.822589
## Test_3Post:DomainRS:Groupcontrol:GenderG   -1.331e-01  3.785e-01  -0.352 0.725135
## Test_3Rest:DomainRS:Groupcontrol:GenderG   -1.331e-01  3.794e-01  -0.351 0.725855
## Test_3Post:DomainCo:Grouptreatment:GenderB  7.599e-01  3.056e-01   2.487 0.012884 *
## Test_3Rest:DomainCo:Grouptreatment:GenderB  1.025e+00  3.051e-01   3.359 0.000782 ***
```

```
## Test_3Post:DomainDS:Grouptreatment:GenderB -7.630e-02  2.994e-01  -0.255 0.798821
## Test_3Rest:DomainDS:Grouptreatment:GenderB  4.185e-02  2.988e-01   0.140 0.888614
## Test_3Post:DomainFi:Grouptreatment:GenderB  3.048e-01  3.021e-01   1.009 0.313008
## Test_3Rest:DomainFi:Grouptreatment:GenderB  4.433e-01  3.015e-01   1.471 0.141418
## Test_3Post:DomainLS:Grouptreatment:GenderB -3.069e-01  3.014e-01  -1.018 0.308494
## Test_3Rest:DomainLS:Grouptreatment:GenderB -1.291e-01  3.014e-01  -0.428 0.668501
## Test_3Post:DomainRS:Grouptreatment:GenderB  6.086e-02  2.985e-01   0.204 0.838466
## Test_3Rest:DomainRS:Grouptreatment:GenderB  3.601e-02  2.984e-01   0.121 0.903961
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## fit warnings:
## fixed-effect model matrix is rank deficient so dropping 18 columns / coefficients
```

Appendix I

Outputs of model $glmer\_Comp\_PrePost\_LikeM$

I.1 Analysis of Deviance Table

```
ANOVA_glmer_Comp_PrePost_LikeM

## Analysis of Deviance Table (Type II Wald chisquare tests)
##
## Response: Pct_refl
##                              Chisq Df Pr(>Chisq)
## Test_3                    203.6728  2  < 2.2e-16 ***
## Domain                    759.1329  5  < 2.2e-16 ***
## Test_3:Domain              27.7822 10  0.0019560 **
## Group:Like                 14.8051  3  0.0019910 **
## Test_3:Group:Like           2.9723  6  0.8123186
## Domain:Group:Like          44.1797 15  0.0001031 ***
## Test_3:Domain:Group:Like  12.6497 30  0.9977005
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

I.2 Summary of fitted model

```
Summary_glmer_Comp_PrePost_LikeM

## Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
##  Family: Gamma  ( log )
## Formula: Pct_refl ~ Test_3 * Domain * Group:Like + (1 | ClassNbr/ChildInClass)
##    Data: .
## Control: glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 5e+05))
##
##      AIC      BIC   logLik deviance df.resid
##  36833.9  37310.5 -18342.0  36683.9     4173
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.3366 -0.5853 -0.0739  0.3854  6.6482
##
## Random effects:
##  Groups                Name        Variance Std.Dev.
##  ChildInClass:ClassNbr (Intercept) 0.06997  0.2645
##  ClassNbr              (Intercept) 0.02636  0.1624
##  Residual                          0.55966  0.7481
## Number of obs: 4248, groups:  ChildInClass:ClassNbr, 236; ClassNbr, 17
```

```
##
## Fixed effects:
##                                             Estimate Std. Error t value Pr(>|z|)
## (Intercept)                                 3.607807   0.185069  19.494  < 2e-16 ***
## Test_3Post                                 -0.376944   0.212772  -1.772 0.076464 .
## Test_3Rest                                 -0.376959   0.212697  -1.772 0.076348 .
## DomainCo                                   -0.763665   0.214643  -3.558 0.000374 ***
## DomainDS                                    0.760948   0.212828   3.575 0.000350 ***
## DomainFi                                    0.226647   0.213414   1.062 0.288234
## DomainLS                                   -0.221522   0.213497  -1.038 0.299464
## DomainRS                                   -0.173684   0.213091  -0.815 0.415033
## Test_3Post:DomainCo                        -0.766820   0.305813  -2.507 0.012160 *
## Test_3Rest:DomainCo                        -0.766785   0.305874  -2.507 0.012181 *
## Test_3Post:DomainDS                         0.263484   0.300852   0.876 0.381142
## Test_3Rest:DomainDS                         0.263514   0.300913   0.876 0.381185
## Test_3Post:DomainFi                        -0.424734   0.302552  -1.404 0.160367
## Test_3Rest:DomainFi                        -0.424725   0.302525  -1.404 0.160338
## Test_3Post:DomainLS                        -0.214137   0.302069  -0.709 0.478387
## Test_3Rest:DomainLS                        -0.214122   0.302037  -0.709 0.478369
## Test_3Post:DomainRS                         0.015693   0.301341   0.052 0.958468
## Test_3Rest:DomainRS                         0.015689   0.301314   0.052 0.958475
## Grouptreatment:LikeDislikeM                 0.248206   0.260137   0.954 0.340015
## Groupcontrol:LikeDislikeM                   0.151313   0.358795   0.422 0.673226
## Grouptreatment:LikeLikeM                    0.012911   0.225196   0.057 0.954281
## Test_3Post:Grouptreatment:LikeDislikeM     -0.009330   0.314176  -0.030 0.976310
## Test_3Rest:Grouptreatment:LikeDislikeM     -0.355370   0.314381  -1.130 0.258315
## Test_3Post:Groupcontrol:LikeDislikeM        0.035559   0.483385   0.074 0.941359
## Test_3Rest:Groupcontrol:LikeDislikeM        0.035601   0.482255   0.074 0.941153
## Test_3Post:Grouptreatment:LikeLikeM        -0.012238   0.260993  -0.047 0.962601
## Test_3Rest:Grouptreatment:LikeLikeM        -0.298295   0.261122  -1.142 0.253305
## DomainCo:Grouptreatment:LikeDislikeM        0.514169   0.315965   1.627 0.103674
## DomainDS:Grouptreatment:LikeDislikeM       -0.158672   0.314617  -0.504 0.614028
## DomainFi:Grouptreatment:LikeDislikeM       -0.062791   0.315034  -0.199 0.842016
## DomainLS:Grouptreatment:LikeDislikeM        0.053099   0.315500   0.168 0.866346
## DomainRS:Grouptreatment:LikeDislikeM       -0.049933   0.314960  -0.159 0.874033
## DomainCo:Groupcontrol:LikeDislikeM          0.296808   0.485213   0.612 0.540732
## DomainDS:Groupcontrol:LikeDislikeM         -0.099321   0.483497  -0.205 0.837243
## DomainFi:Groupcontrol:LikeDislikeM          0.070764   0.483624   0.146 0.883668
## DomainLS:Groupcontrol:LikeDislikeM          0.050844   0.484038   0.105 0.916343
## DomainRS:Groupcontrol:LikeDislikeM         -0.047303   0.483367  -0.098 0.922042
## DomainCo:Grouptreatment:LikeLikeM           0.075025   0.263592   0.285 0.775932
## DomainDS:Grouptreatment:LikeLikeM          -0.007544   0.261172  -0.029 0.976956
## DomainFi:Grouptreatment:LikeLikeM           0.043785   0.261641   0.167 0.867096
## DomainLS:Grouptreatment:LikeLikeM          -0.035509   0.262119  -0.135 0.892242
## DomainRS:Grouptreatment:LikeLikeM          -0.007603   0.261381  -0.029 0.976795
## Test_3Post:DomainCo:Grouptreatment:LikeDislikeM  0.373145   0.448697   0.832 0.405624
## Test_3Rest:DomainCo:Grouptreatment:LikeDislikeM  0.719161   0.448939   1.602 0.109175
## Test_3Post:DomainDS:Grouptreatment:LikeDislikeM -0.076887   0.444530  -0.173 0.862680
## Test_3Rest:DomainDS:Grouptreatment:LikeDislikeM  0.117301   0.444917   0.264 0.792053
## Test_3Post:DomainFi:Grouptreatment:LikeDislikeM  0.328830   0.445726   0.738 0.460673
## Test_3Rest:DomainFi:Grouptreatment:LikeDislikeM  0.496309   0.445995   1.113 0.265789
## Test_3Post:DomainLS:Grouptreatment:LikeDislikeM -0.072467   0.446259  -0.162 0.871001
## Test_3Rest:DomainLS:Grouptreatment:LikeDislikeM  0.240495   0.446502   0.539 0.590148
## Test_3Post:DomainRS:Grouptreatment:LikeDislikeM  0.118135   0.445144   0.265 0.790711
## Test_3Rest:DomainRS:Grouptreatment:LikeDislikeM  0.243697   0.446091   0.546 0.584863
## Test_3Post:DomainCo:Groupcontrol:LikeDislikeM   -0.120778   0.689303  -0.175 0.860909
## Test_3Rest:DomainCo:Groupcontrol:LikeDislikeM   -0.120852   0.688299  -0.176 0.860624
## Test_3Post:DomainDS:Groupcontrol:LikeDislikeM   -0.009192   0.683115  -0.013 0.989264
## Test_3Rest:DomainDS:Groupcontrol:LikeDislikeM   -0.009267   0.682627  -0.014 0.989169
## Test_3Post:DomainFi:Groupcontrol:LikeDislikeM    0.035788   0.687387   0.052 0.958478
## Test_3Rest:DomainFi:Groupcontrol:LikeDislikeM    0.035766   0.685685   0.052 0.958400
## Test_3Post:DomainLS:Groupcontrol:LikeDislikeM   -0.004910   0.684788  -0.007 0.994279
## Test_3Rest:DomainLS:Groupcontrol:LikeDislikeM   -0.004967   0.683706  -0.007 0.994203
## Test_3Post:DomainRS:Groupcontrol:LikeDislikeM    0.131465   0.683578   0.192 0.847492
## Test_3Rest:DomainRS:Groupcontrol:LikeDislikeM    0.131440   0.683247   0.192 0.847448
## Test_3Post:DomainCo:Grouptreatment:LikeLikeM     0.534214   0.374765   1.425 0.154023
## Test_3Rest:DomainCo:Grouptreatment:LikeLikeM     0.820247   0.374816   2.188 0.028641 *
```

```
## Test_3Post:DomainDS:Grouptreatment:LikeLikeM    -0.156356   0.369161  -0.424 0.671898
## Test_3Rest:DomainDS:Grouptreatment:LikeLikeM    -0.004410   0.369443  -0.012 0.990475
## Test_3Post:DomainFi:Grouptreatment:LikeLikeM     0.321614   0.370627   0.868 0.385529
## Test_3Rest:DomainFi:Grouptreatment:LikeLikeM     0.495102   0.370696   1.336 0.181679
## Test_3Post:DomainLS:Grouptreatment:LikeLikeM    -0.132677   0.370606  -0.358 0.720342
## Test_3Rest:DomainLS:Grouptreatment:LikeLikeM     0.072194   0.370917   0.195 0.845678
## Test_3Post:DomainRS:Grouptreatment:LikeLikeM     0.105418   0.369629   0.285 0.775492
## Test_3Rest:DomainRS:Grouptreatment:LikeLikeM     0.034986   0.369727   0.095 0.924610
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## fit warnings:
## fixed-effect model matrix is rank deficient so dropping 18 columns / coefficients
```

Appendix J

Outputs of model $glmmPQL\_shVid$

J.1 Analysis of Deviance Table

```
ANOVA_glmer_Comp_ShortVid

## Analysis of Deviance Table (Type II tests)
##
## Response: zz
##                          Chisq Df Pr(>Chisq)
## WhichPct                0.0025  1   0.960098
## Cluster                40.8310  4  2.913e-08 ***
## Gender                  0.4235  1   0.515202
## WhichPct:Cluster       14.9189  4   0.004872 **
## WhichPct:Gender         0.1926  1   0.660733
## Cluster:Gender          2.0844  4   0.720235
## WhichPct:Cluster:Gender 25.9591  4  3.225e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

J.2 Summary of fitted model

```
Summary_glmer_Comp_ShortVid


##                                                          Value Std.Error  DF      t-value      p-value
## (Intercept)                                        13.42377496  2.227516 154  6.026341510 1.189639e-08
## WhichPctMultiple_Clicks_Considered                  3.81528410  1.834714 154  2.079498181 3.922906e-02
## Cluster2                                           -0.22829610  2.868556 154 -0.079585722 9.366701e-01
## Cluster3                                            5.82168448  4.105086 154  1.418163722 1.581634e-01
## Cluster4                                            7.53916907  4.545522 154  1.658592630 9.923329e-02
## Cluster5                                           -0.57533636  3.578083 154 -0.160794582 8.724659e-01
## GenderB                                            -0.98538226  3.180611 154 -0.309809097 7.571250e-01
## WhichPctMultiple_Clicks_Considered:Cluster2        -3.81521413  1.834729 154 -2.079443011 3.923420e-02
## WhichPctMultiple_Clicks_Considered:Cluster3         0.02749757  4.636041 154  0.005931262 9.952752e-01
## WhichPctMultiple_Clicks_Considered:Cluster4         6.86057443  6.163951 154  1.113015774 2.674360e-01
## WhichPctMultiple_Clicks_Considered:Cluster5        13.77695883  4.853193 154  2.838741455 5.140490e-03
## WhichPctMultiple_Clicks_Considered:GenderB         -1.97020165  2.096259 154 -0.939865345 3.487585e-01
## Cluster2:GenderB                                   -3.84658925  4.241573 154 -0.906878055 3.658881e-01
## Cluster3:GenderB                                    2.08823205  5.754094 154  0.362912402 7.171679e-01
## Cluster4:GenderB                                    3.36171995  5.912053 154  0.568621430 5.704413e-01
## Cluster5:GenderB                                    4.67229294  4.919889 154  0.949674501 3.437656e-01
## WhichPctMultiple_Clicks_Considered:Cluster2:GenderB   6.52083687  2.640938 154  2.469136354 1.463627e-02
## WhichPctMultiple_Clicks_Considered:Cluster3:GenderB   1.72523613  6.305007 154  0.273629538 7.847360e-01
## WhichPctMultiple_Clicks_Considered:Cluster4:GenderB  -2.07931773  7.504716 154 -0.277068159 7.820993e-01
## WhichPctMultiple_Clicks_Considered:Cluster5:GenderB -15.62199084  4.957992 154 -3.150870777 1.956327e-03
```

Appendix K

Outputs of model $glmmPQL\_shVid$ at *Class* level

K.1 Analysis of Deviance Table

```
ANOVA_glmer_Comp_ShortVidClass

## Analysis of Deviance Table (Type II tests)
##
## Response: zz
##                          Chisq Df Pr(>Chisq)
## WhichPct                0.0014  1   0.969869
## Class                  21.3631 11   0.029799 *
## Gender                  0.1501  1   0.698413
## WhichPct:Class         44.0873 11  7.016e-06 ***
## WhichPct:Gender         0.3568  1   0.550270
## Class:Gender            3.9253 11   0.972052
## WhichPct:Class:Gender 29.8462 11   0.001675 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

K.2 Summary of fitted model

```
Summary_glmer_Comp_ShortVidClass

##                                                         Value Std.Error  DF      t-value      p-value
## (Intercept)                                       14.67454070  3.980344 140  3.686752038 0.0003240039
## WhichPctMultiple_Clicks_Considered                 5.02221360  3.256793 140  1.542073105 0.1253131606
## ClassXA02                                         -1.67832538  5.070151 140 -0.331020818 0.7411236913
## ClassXB03                                          6.16261522  6.647877 140  0.927004958 0.3555193010
## ClassXB04                                          3.20298730  6.656317 140  0.481195098 0.6311290693
## ClassXC05                                         -6.28964748  6.787114 140 -0.926704316 0.3556748666
## ClassXC06                                         -3.70191726  5.163945 140 -0.716877730 0.4746438136
## ClassXD07                                          6.24814578  6.957685 140  0.898020824 0.3707162459
## ClassXD08                                          1.07520150  5.575802 140  0.192833508 0.8473688171
## ClassXE09                                         -3.32624025  5.357818 140 -0.620819949 0.5357274941
## ClassXE10                                          1.90984696  5.506911 140  0.346809152 0.7292554402
## ClassXF11                                          1.54982562  6.581520 140  0.235481409 0.8141792175
## ClassXF12                                          2.48652194  5.610903 140  0.443158961 0.6583347960
## GenderB                                            0.28903202  5.378215 140  0.053741254 0.9572179255
## WhichPctMultiple_Clicks_Considered:ClassXA02      -4.51400032  3.535074 140 -1.276918250 0.2037449340
## WhichPctMultiple_Clicks_Considered:ClassXB03       5.64320692  7.964770 140  0.708521019 0.4797998342
## WhichPctMultiple_Clicks_Considered:ClassXB04      -0.01033269  6.784662 140 -0.001522948 0.9987870314
## WhichPctMultiple_Clicks_Considered:ClassXC05      14.46818995  8.173124 140  1.770215349 0.0788679019
## WhichPctMultiple_Clicks_Considered:ClassXC06       8.28809424  4.837218 140  1.713400989 0.0888522340
## WhichPctMultiple_Clicks_Considered:ClassXD07       0.78854565  7.503218 140  0.105094334 0.9164513766
## WhichPctMultiple_Clicks_Considered:ClassXD08      -3.78276378  3.635477 140 -1.040513780 0.2998948627
## WhichPctMultiple_Clicks_Considered:ClassXE09      -5.02221305  3.256800 140 -1.542069678 0.1253139931
## WhichPctMultiple_Clicks_Considered:ClassXE10       1.45801555  5.325845 140  0.273762320 0.7846707638
## WhichPctMultiple_Clicks_Considered:ClassXF11      -4.24567721  4.104189 140 -1.034474029 0.3026982205
## WhichPctMultiple_Clicks_Considered:ClassXF12       3.09115742  5.138059 140  0.601619663 0.5484003211
```

```
## WhichPctMultiple_Clicks_Considered:GenderB             2.70317686  4.963902 140  0.544566977 0.5869177244
## ClassXA02:GenderB                                     -0.26396191  7.151699 140 -0.036908979 0.9706101566
## ClassXB03:GenderB                                      2.11248071  9.318281 140  0.226702824 0.8209855115
## ClassXB04:GenderB                                     -0.05795950  8.743225 140 -0.006629075 0.9947202386
## ClassXC05:GenderB                                      6.40448017  8.496966 140  0.753737323 0.4522727538
## ClassXC06:GenderB                                     -1.58217280  7.213916 140 -0.219322309 0.8267184240
## ClassXD07:GenderB                                     -2.30899758  8.991165 140 -0.256807394 0.7977044824
## ClassXD08:GenderB                                     -1.56039457  7.721559 140 -0.202082846 0.8401453131
## ClassXE09:GenderB                                      3.83315619  7.583838 140  0.505437484 0.6140464110
## ClassXE10:GenderB                                     -0.51166084  8.035698 140 -0.063673476 0.9493210478
## ClassXF11:GenderB                                      0.95074659  9.094822 140  0.104537135 0.9168927156
## ClassXF12:GenderB                                     -1.92516545  7.690009 140 -0.250346306 0.8026865560
## WhichPctMultiple_Clicks_Considered:ClassXA02:GenderB  -1.67531595  5.359242 140 -0.312603172 0.7550473206
## WhichPctMultiple_Clicks_Considered:ClassXB03:GenderB  -3.02757011 11.325189 140 -0.267330644 0.7896080066
## WhichPctMultiple_Clicks_Considered:ClassXB04:GenderB   1.02946409  9.234351 140  0.111482015 0.9113937985
## WhichPctMultiple_Clicks_Considered:ClassXC05:GenderB -12.43389248  9.742601 140 -1.276239461 0.2039840205
## WhichPctMultiple_Clicks_Considered:ClassXC06:GenderB -16.01344951  6.118197 140 -2.617347672 0.0098356134
## WhichPctMultiple_Clicks_Considered:ClassXD07:GenderB  -8.16355242  8.487719 140 -0.961807562 0.3378045437
## WhichPctMultiple_Clicks_Considered:ClassXD08:GenderB  -0.85201032  5.866422 140 -0.145235078 0.8847341583
## WhichPctMultiple_Clicks_Considered:ClassXE09:GenderB  -0.82852814  5.479817 140 -0.151196312 0.8800385039
## WhichPctMultiple_Clicks_Considered:ClassXE10:GenderB  13.16648605  9.943527 140  1.324126350 0.1876188535
## WhichPctMultiple_Clicks_Considered:ClassXF11:GenderB   1.73692857  7.556212 140  0.229867644 0.8185301337
## WhichPctMultiple_Clicks_Considered:ClassXF12:GenderB -10.81654425  6.358719 140 -1.701057032 0.0911523191
```

Appendix L

Outputs of model $glmmPQL\_ovlayVid$

L.1 Analysis of Deviance Table

```
ANOVA_glmer_Comp_OvlayVid

## Analysis of Deviance Table (Type II tests)
##
## Response: zz
##                            Chisq Df Pr(>Chisq)
## WhichPct                 333.2510  1  < 2.2e-16 ***
## Cluster                   31.9286  4  1.978e-06 ***
## Gender                     0.3389  1     0.5605
## WhichPct:Cluster          45.1876  4  3.634e-09 ***
## WhichPct:Gender            0.0013  1     0.9707
## Cluster:Gender             1.5143  4     0.8241
## WhichPct:Cluster:Gender    4.6140  4     0.3292
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

L.2 Summary of fitted model

```
Summary_glmer_Comp_OvlayVid

##                                                        Value Std.Error  DF    t-value      p-value
## (Intercept)                                       21.4877433  2.419391 154  8.8814694 1.611108e-15
## WhichPctMultiple_Views_Considered                 18.0114710  2.864987 154  6.2867552 3.182744e-09
## Cluster2                                           4.1713082  3.291863 154  1.2671571 2.070120e-01
## Cluster3                                          -1.8117025  3.718917 154 -0.4871586 6.268386e-01
## Cluster4                                           4.4982429  4.177937 154  1.0766661 2.833132e-01
## Cluster5                                           5.1505160  4.069428 154  1.2656609 2.075460e-01
## GenderB                                            0.8198198  3.588359 154  0.2284665 8.195867e-01
## WhichPctMultiple_Views_Considered:Cluster2         4.0237122  4.067794 154  0.9891632 3.241349e-01
## WhichPctMultiple_Views_Considered:Cluster3        -7.7123517  3.848177 154 -2.0041572 4.680644e-02
## WhichPctMultiple_Views_Considered:Cluster4         5.3030213  5.465065 154  0.9703492 3.333940e-01
## WhichPctMultiple_Views_Considered:Cluster5        21.3747759  6.405313 154  3.3370385 1.061572e-03
## WhichPctMultiple_Views_Considered:GenderB         -2.7083305  4.049411 154 -0.6688209 5.046111e-01
## Cluster2:GenderB                                   1.3288042  5.145927 154  0.2582244 7.965785e-01
## Cluster3:GenderB                                  -0.9903804  5.276962 154 -0.1876800 8.513745e-01
## Cluster4:GenderB                                  -2.1991497  5.480340 154 -0.4012798 6.887704e-01
## Cluster5:GenderB                                   3.0058835  5.797168 154  0.5185089 6.048475e-01
## WhichPctMultiple_Views_Considered:Cluster2:GenderB 10.9577449  6.793267 154  1.6130302 1.087853e-01
## WhichPctMultiple_Views_Considered:Cluster3:GenderB  4.3106619  5.480112 154  0.7866011 4.327242e-01
## WhichPctMultiple_Views_Considered:Cluster4:GenderB -0.8531141  6.853394 154 -0.1244805 9.010972e-01
## WhichPctMultiple_Views_Considered:Cluster5:GenderB -4.9675424  8.692271 154 -0.5714896 5.685010e-01
```

Appendix M

Outputs of model $glmmPQL\_repeat$

M.1 Analysis of Deviance Table

```
ANOVA_glmer_Comp_Repeat

## Analysis of Deviance Table (Type II tests)
##
## Response: zz
##                            Chisq Df Pr(>Chisq)
## WhichPct                  0.0006  1   0.979928
## Cluster                  24.0430  4  7.831e-05 ***
## Gender                    0.9332  1   0.334022
## WhichPct:Cluster         42.0873  4  1.600e-08 ***
## WhichPct:Gender           0.0000  1   0.999418
## Cluster:Gender            5.4442  4   0.244675
## WhichPct:Cluster:Gender 16.3293  4   0.002608 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

M.2 Summary of fitted model

```
Summary_glmer_Comp_Repeat

##                                                              Value Std.Error  DF     t-value      p-value
## (Intercept)                                             15.3813874  2.695466 154  5.70639243 5.751341e-08
## WhichPctRepeat_Same_Loops_Considered                     3.3002273  1.697527 154  1.94413806 5.370143e-02
## Cluster2                                                 5.4358549  3.816721 154  1.42422117 1.564056e-01
## Cluster3                                                -2.5467695  4.073919 154 -0.62513992 5.328045e-01
## Cluster4                                                 1.3943064  4.395301 154  0.31722664 7.515013e-01
## Cluster5                                                13.3271339  5.010196 154  2.66000225 8.641111e-03
## GenderB                                                  3.1816950  3.991341 154  0.79714932 4.265916e-01
## WhichPctRepeat_Same_Loops_Considered:Cluster2           11.7920334  3.918025 154  3.00968805 3.056496e-03
## WhichPctRepeat_Same_Loops_Considered:Cluster3           -3.3002135  1.697540 154 -1.94411581 5.370414e-02
## WhichPctRepeat_Same_Loops_Considered:Cluster4           -3.3001971  1.697540 154 -1.94410617 5.370531e-02
## WhichPctRepeat_Same_Loops_Considered:Cluster5            4.6036304  3.986691 154  1.15474969 2.499816e-01
## WhichPctRepeat_Same_Loops_Considered:GenderB            -3.3002085  1.697540 154 -1.94411289 5.370449e-02
## Cluster2:GenderB                                        -0.5272898  6.018210 154 -0.08761572 9.302959e-01
## Cluster3:GenderB                                        -0.9013511  5.804654 154 -0.15528076 8.768033e-01
## Cluster4:GenderB                                         1.2920863  6.032491 154  0.21418785 8.306839e-01
## Cluster5:GenderB                                       -11.8583743  6.705879 154 -1.76835503 7.898152e-02
## WhichPctRepeat_Same_Loops_Considered:Cluster2:GenderB   1.2988607  6.153648 154  0.21107167 8.331104e-01
## WhichPctRepeat_Same_Loops_Considered:Cluster3:GenderB   3.3002249  1.697564 154  1.94409428 5.370676e-02
## WhichPctRepeat_Same_Loops_Considered:Cluster4:GenderB  14.8244455  3.871945 154  3.82868137 1.869973e-04
## WhichPctRepeat_Same_Loops_Considered:Cluster5:GenderB   6.8723505  5.464406 154  1.25765734 2.104195e-01
```

Appendix N

Outputs of model $glmmPQL\_mes$

N.1 Analysis of Deviance Table

```
ANOVA_glmer_Comp_Message

## Analysis of Deviance Table (Type II tests)
##
## Response: zz
##                  Chisq Df Pr(>Chisq)
## Cluster        43.5340  4  8.017e-09 ***
## Gender          0.0532  1     0.8176
## Cluster:Gender  4.2913  4     0.3680
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

N.2 Summary of fitted model

```
Summary_glmer_Comp_Message

##                      Value Std.Error  DF    t-value      p-value
## (Intercept)      3.9526316 0.7518342 154  5.2573182 4.807291e-07
## Cluster2         1.4337684 0.9974208 154  1.4374760 1.526116e-01
## Cluster3        -2.0072470 1.1795745 154 -1.7016703 9.083517e-02
## Cluster4         0.9255502 1.2416133 154  0.7454416 4.571410e-01
## Cluster5         4.7723684 1.2084044 154  3.9493140 1.189627e-04
## GenderB         -0.6595066 1.1119778 154 -0.5930933 5.539887e-01
## Cluster2:GenderB  1.3318566 1.5288322 154  0.8711594 3.850231e-01
## Cluster3:GenderB  1.5934553 1.6669209 154  0.9559274 3.406071e-01
## Cluster4:GenderB  0.9290520 1.6434774 154  0.5652965 5.726947e-01
## Cluster5:GenderB -1.4268268 1.6874449 154 -0.8455546 3.991131e-01
```

Appendix O

Entire pre- and post-test

Geometrie

Zyklus 4

Schüler-Code: ☐ ☐ ☐ ☐

Carole DORDING
Universität Luxemburg

LUXEMBOURG INSTITUTE OF SCIENCE AND TECHNOLOGY | LIST

uni.lu UNIVERSITÉ DU LUXEMBOURG

LUCET | LUXEMBOURG CENTRE FOR EDUCATIONAL TESTING

Liebe Schülerin,
lieber Schüler,

hier bekommst du ein Heft mit einer ganzen Reihe von Mathe-Aufgaben. Es ist kein Problem, wenn du nicht alle Aufgaben lösen kannst; einige sind momentan noch zu schwierig für dich. Versuche die Aufgaben so gut und so genau wie möglich zu machen. Die Geraden und Strecken mit **Lineal** zeichnen! Viel Spaß dabei!

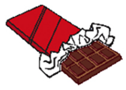
Chère étudiante,
cher étudiant,

voici un cahier contenant une série d'exercices mathématiques. Ce ne sera pas grave si tu n'arrives pas à résoudre tous ces exercices; certains sont encore difficiles pour toi en ce moment. Essaie de faire ces exercices aussi bien et d'une manière aussi précise que possible. Les droites et les segments doivent être tracés en utilisant **une règle** *(une latte)*. Amuse-toi bien !

Aufgabe 1

In welchem Kästchen befinden sich folgende Esswaren im Gitternetz?
Ergänze die nebenstehenden Sätze.

Dans quelle case les aliments suivants se trouvent-ils dans la grille ?
Complète les phrases figurant à la page ci-contre.

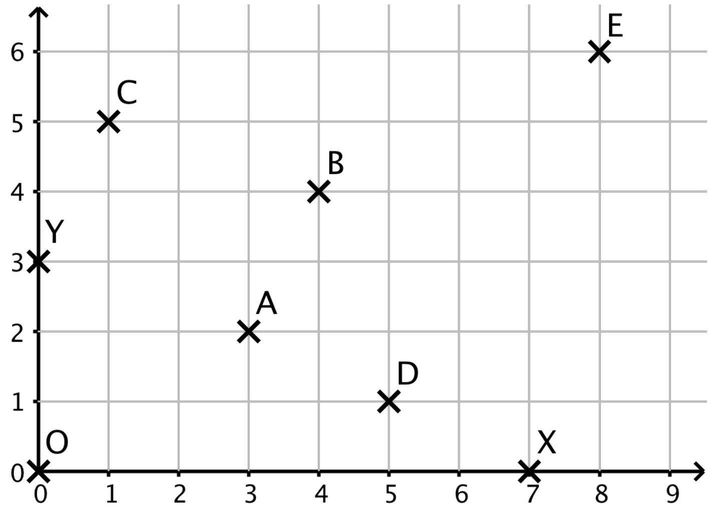
5					
4					
3					
2					
1					
	A	*B*	*C*	*D*	*E*

		Buchstabe / Lettre	Zahl / Nombre
		↓	↓
	Das Bonbon befindet sich im Kästchen: Le bonbon se trouve dans la case :	____	____
	Der Geburtstagskuchen befindet sich im Kästchen: Le gâteau d'anniversaire se trouve dans la case :	____	____
	Der Hamburger befindet sich im Kästchen: Le hamburger se trouve dans la case :	____	____
	Das Popcorn befindet sich im Kästchen: Le pop-corn se trouve dans la case :	____	____
	Die Schokolade befindet sich im Kästchen: Le chocolat se trouve dans la case :	____	____
	Die Pommes frites befinden sich im Kästchen: Les pommes frites se trouvent dans la case :	____	____
	Das Eis befindet sich im Kästchen: La glace se trouve dans la case :	____	____
	Der Lebkuchenmann befindet sich im Kästchen: Le bonhomme de pain d'épices se trouve dans la case:	____	____

Aufgabe 2

Wo befinden sich folgende Punkte auf dem Gitternetz?
Ergänze die nebenstehenden Sätze.

Où les points suivants se trouvent-ils sur la grille ?
Complète les phrases figurant à la page ci-contre.



Zahl	Zahl
Nombre	Nombre
↓	↓

a)	Der Punkt A befindet sich auf dem Gitternetzpunkt: Le point A se trouve sur le point de la grille :	(____ , ____)
b)	Der Punkt B befindet sich auf dem Gitternetzpunkt: Le point B se trouve sur le point de la grille :	(____ , ____)
c)	Der Punkt C befindet sich auf dem Gitternetzpunkt: Le point C se trouve sur le point de la grille :	(____ , ____)
d)	Der Punkt D befindet sich auf dem Gitternetzpunkt: Le point D se trouve sur le point de la grille :	(____ , ____)
e)	Der Punkt E befindet sich auf dem Gitternetzpunkt: Le point E se trouve sur le point de la grille :	(____ , ____)
f)	Der Punkt O befindet sich auf dem Gitternetzpunkt: Le point O se trouve sur le point de la grille :	(____ , ____)
g)	Der Punkt X befindet sich auf dem Gitternetzpunkt: Le point X se trouve sur le point de la grille :	(____ , ____)
h)	Der Punkt Y befindet sich auf dem Gitternetzpunkt: Le point Y se trouve sur le point de la grille :	(____ , ____)

Aufgabe 3

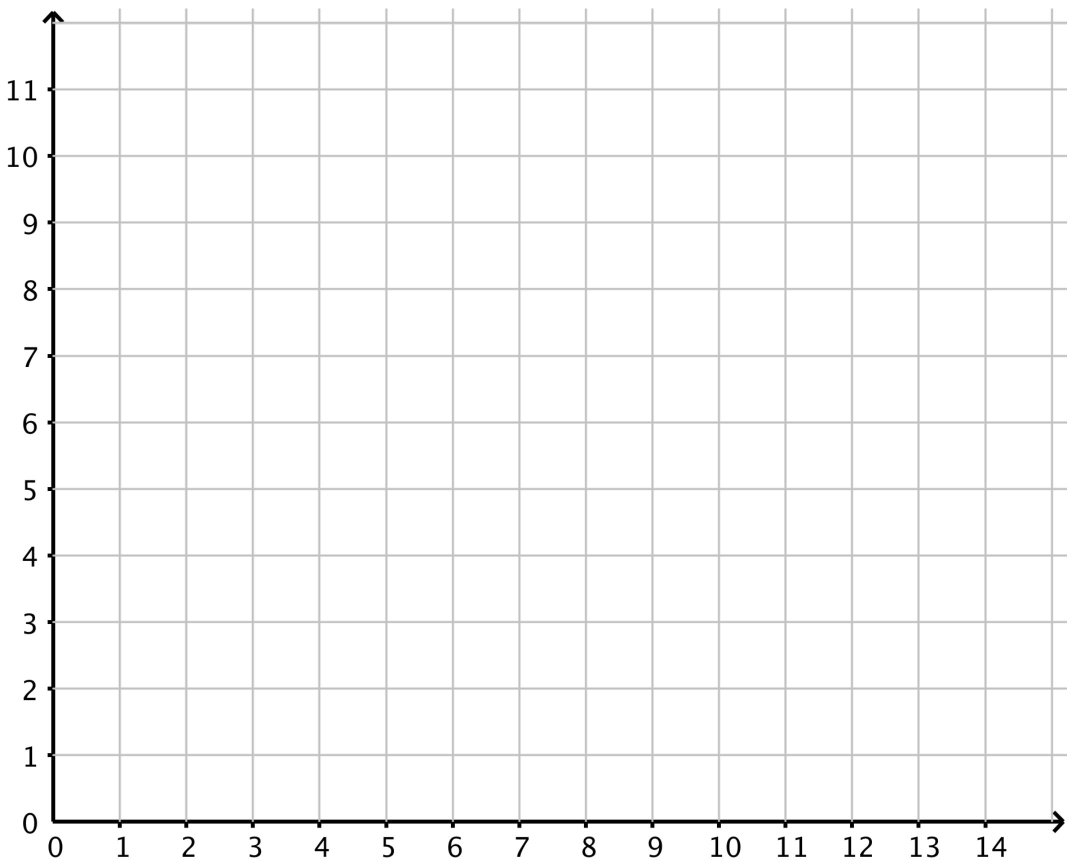
Zeichne folgende Punkte mit ihrem Namen aufs Gitternetz.
Dessine les points suivants, avec leur nom, sur la grille.

$A(3,4)$ $E(13,8)$

$B(6,1)$ $O(0,0)$

$C(1,6)$ $X(2,0)$

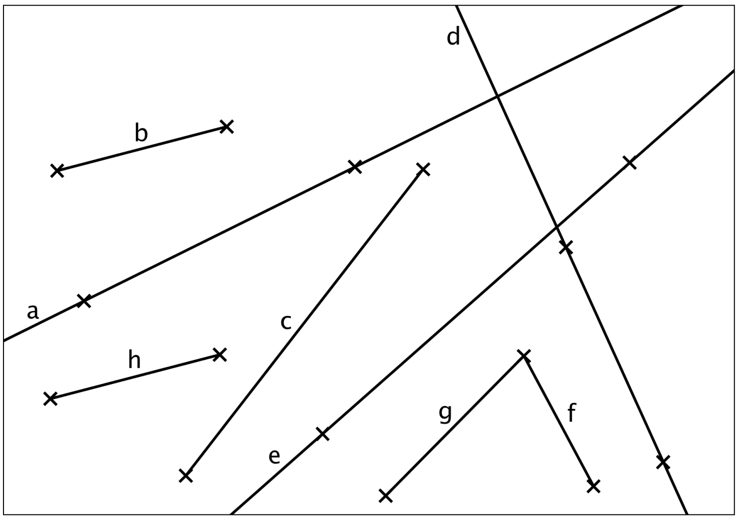
$D(9,9)$ $Y(0,9)$



Aufgabe 4

Gerade oder Strecke?
Kreuze jeweils die richtige Antwort in den untenstehenden Tabellen an.

Droite ou segment ?
Coche à chaque fois la bonne réponse dans les tableaux figurant ci-dessous.



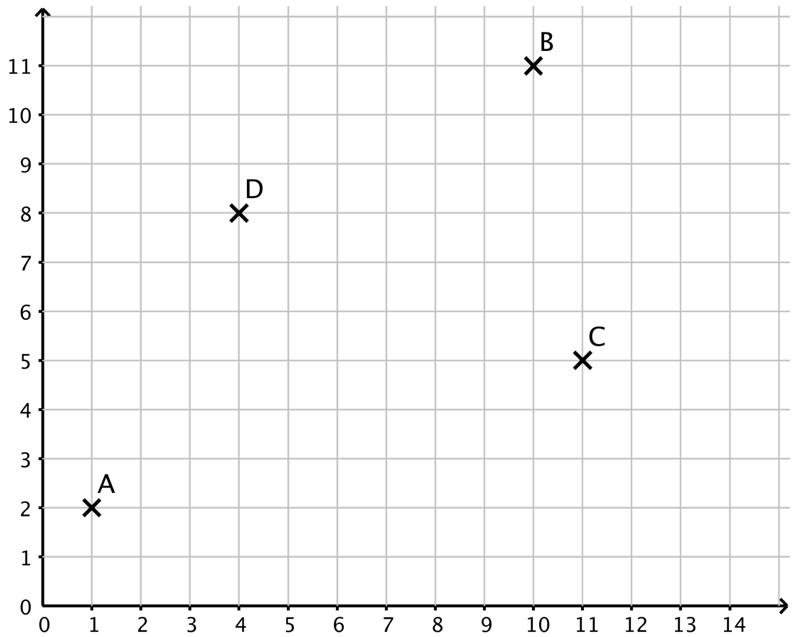
	eine Gerade. / une droite.	eine Strecke. / un segment.
a ist / a est	☐	☐
b ist / b est	☐	☐
c ist / c est	☐	☐
d ist / d est	☐	☐

	eine Gerade. / une droite.	eine Strecke. / un segment.
e ist / e est	☐	☐
f ist / f est	☐	☐
g ist / g est	☐	☐
h ist / h est	☐	☐

Aufgabe 5

Zeichne jeweils die Geraden und die Strecken ins darunterstehende Gitternetz ein.
Dessine à chaque fois les droites et les segments dans la grille située en dessous.

- die Gerade durch den Punkt A und den Punkt D
  la droite passant par le point A et le point D
- die Strecke zwischen A und B
  le segment entre A et B
- die Strecke zwischen C und D
  le segment entre C et D
- die Gerade durch den Punkt B und den Punkt C
  la droite passant par le point B et le point C

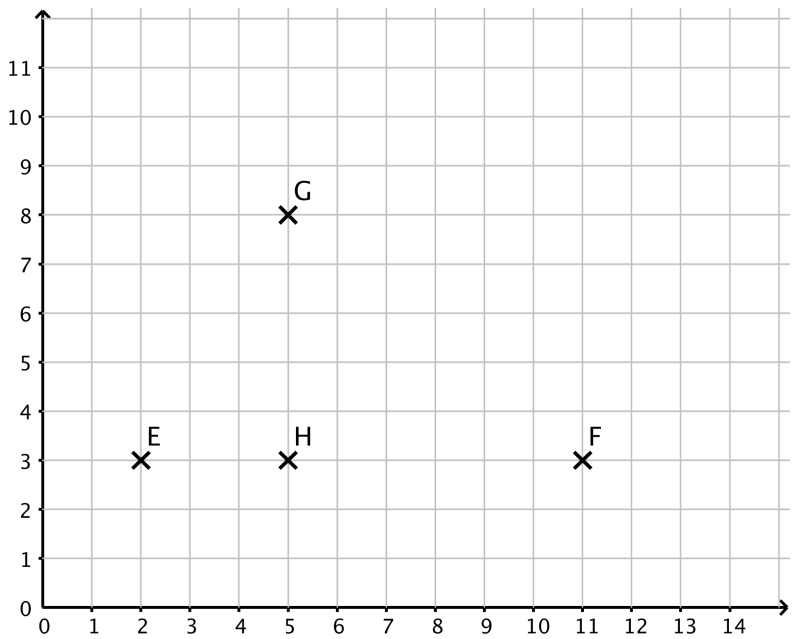


die Strecke $[EG]$
le segment $[EG]$

die Gerade (EF)
la droite (EF)

die Strecke $[GF]$
le segment $[GF]$

die Gerade (GH)
la droite (GH)

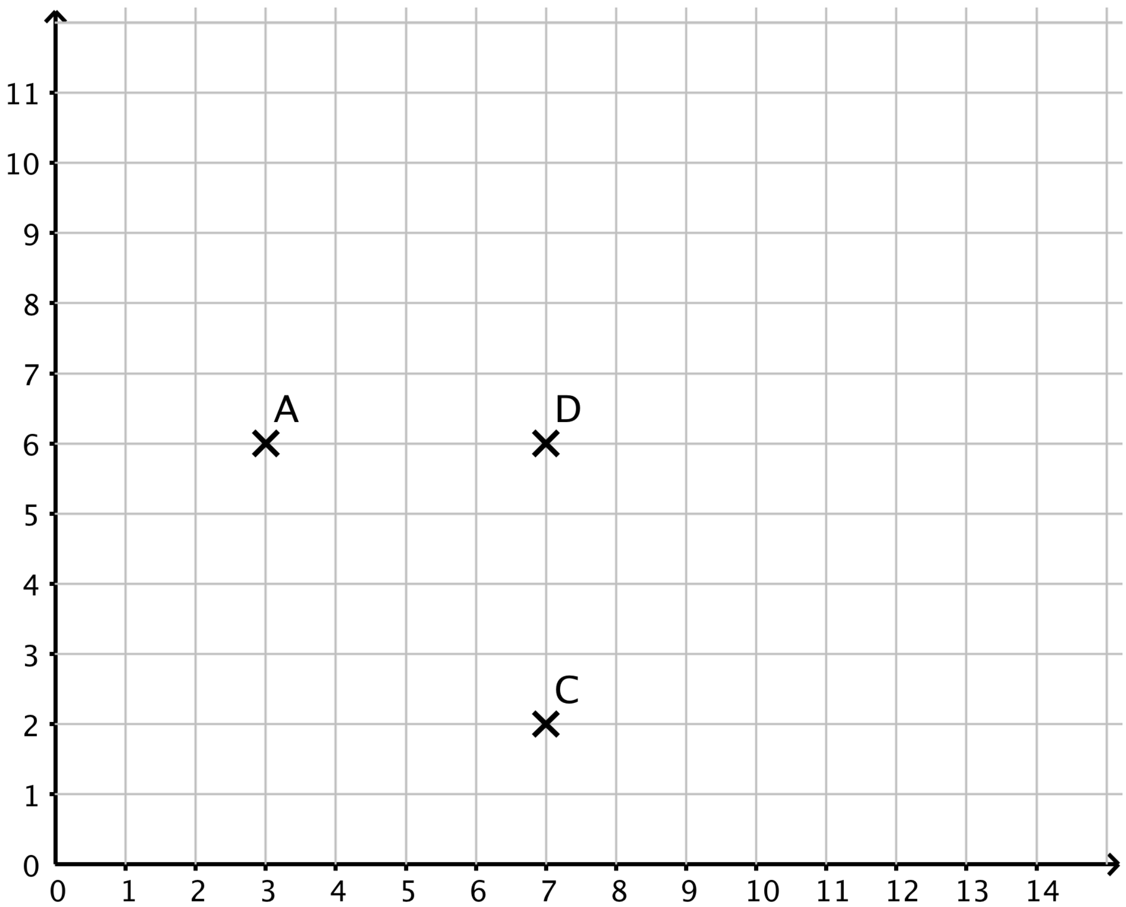


Aufgabe 6

a)

Gegeben sind die Punkte A, C und D.
Das Viereck $ABCD$ ist ein Quadrat.
Zeichne den fehlenden Punkt B **und** das Quadrat ins Gitternetz ein.

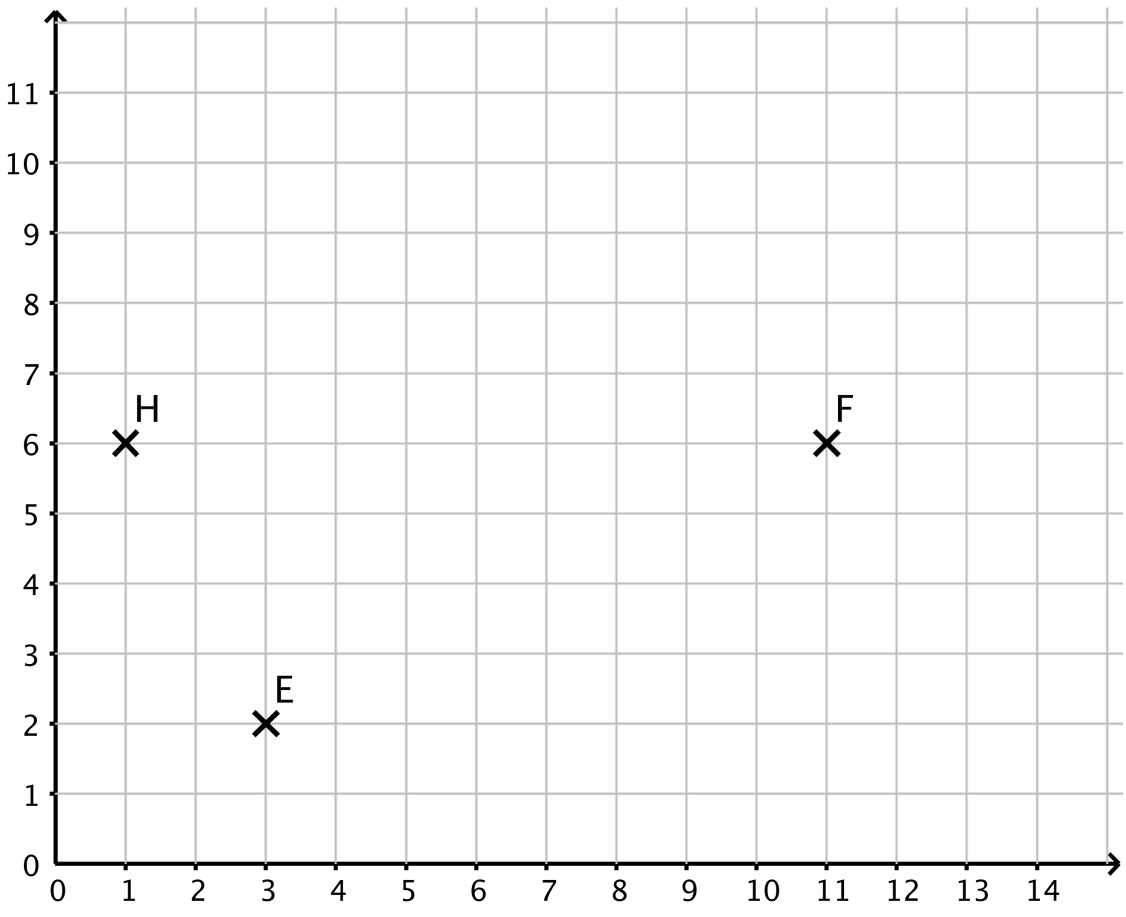
Étant donnés les points A, C et D.
Le quadrilatère $ABCD$ est un carré.
Dessine le point B, qui manque, **et** le carré dans la grille.



b)

Gegeben sind die Punkte E, F und H.
Das Viereck $EFGH$ ist ein Rechteck.
Zeichne den fehlenden Punkt G **und** das Rechteck ins Gitternetz ein.

Étant donnés les points E, F et H.
Le quadrilatère $EFGH$ est un rectangle.
Dessine le point G, qui manque, **et** le rectangle dans la grille.

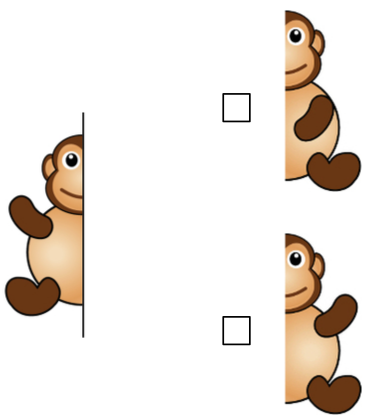


Aufgabe 7

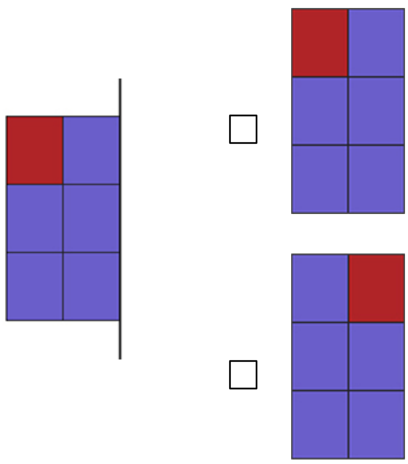
Vervollständige folgende symmetrische Bilder, indem du jeweils die richtige Antwort ankreuzt.

Complète les images symétriques suivantes, en cochant à chaque fois la bonne réponse.

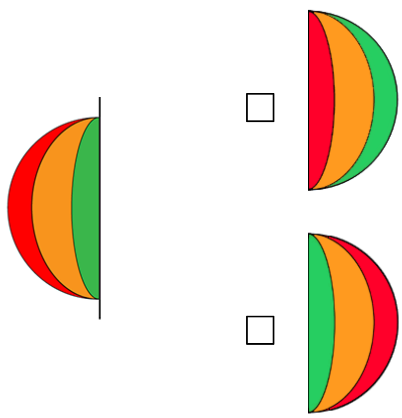
a)



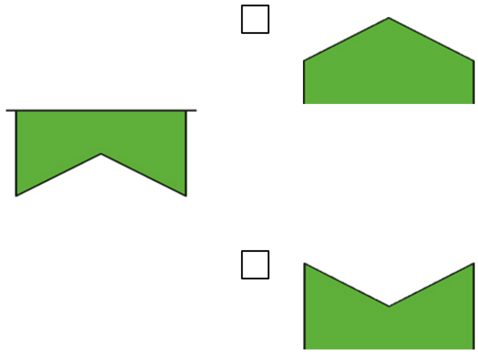
b)



c)



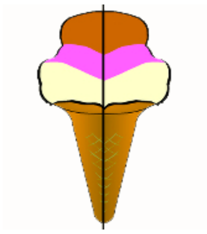
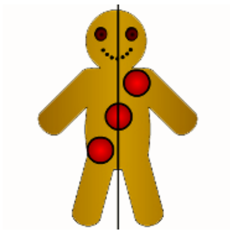
d)

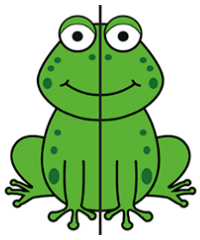


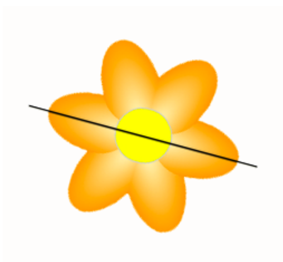
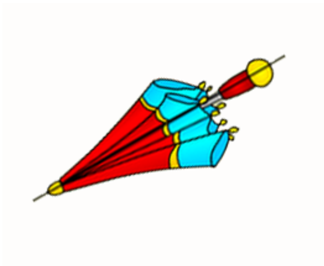
Aufgabe 8

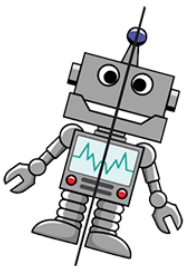
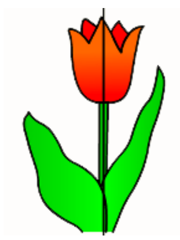
Welche Bilder sind symmetrisch? Kreuze die richtigen Antworten an.

Quelles images sont symétriques ? Coche les bonnes réponses.

☐ 
☐ 
☐ 

☐ 
☐ 
☐ 

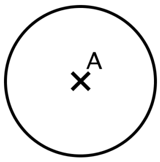
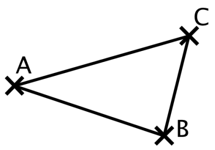
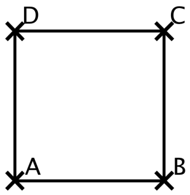
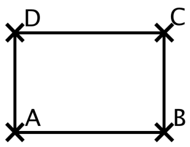
☐ 
☐ 
☐ 

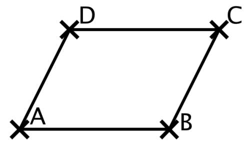
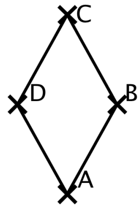
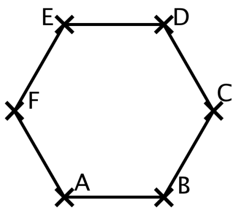
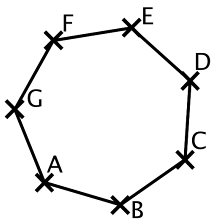
☐ 
☐ 
☐ 

Aufgabe 9

Wie heißen folgende geometrische Figuren **genau**? Sind sie symmetrisch?

Quel est le nom **précis** des figures géométriques suivantes ? Sont-elles symétriques?

	Figur Name Nom de la figure	Symmetrisch? Symétrique ?
	__________	☐ ja / oui ☐ nein / non
	__________	☐ ja / oui ☐ nein / non
	__________	☐ ja / oui ☐ nein / non
	__________	☐ ja / oui ☐ nein / non

	Figur Name Nom de la figure	Symmetrisch? Symétrique ?
	______	☐ ja / oui ☐ nein / non
	______	☐ ja / oui ☐ nein / non
	______	☐ ja / oui ☐ nein / non
	______	☐ ja / oui ☐ nein / non

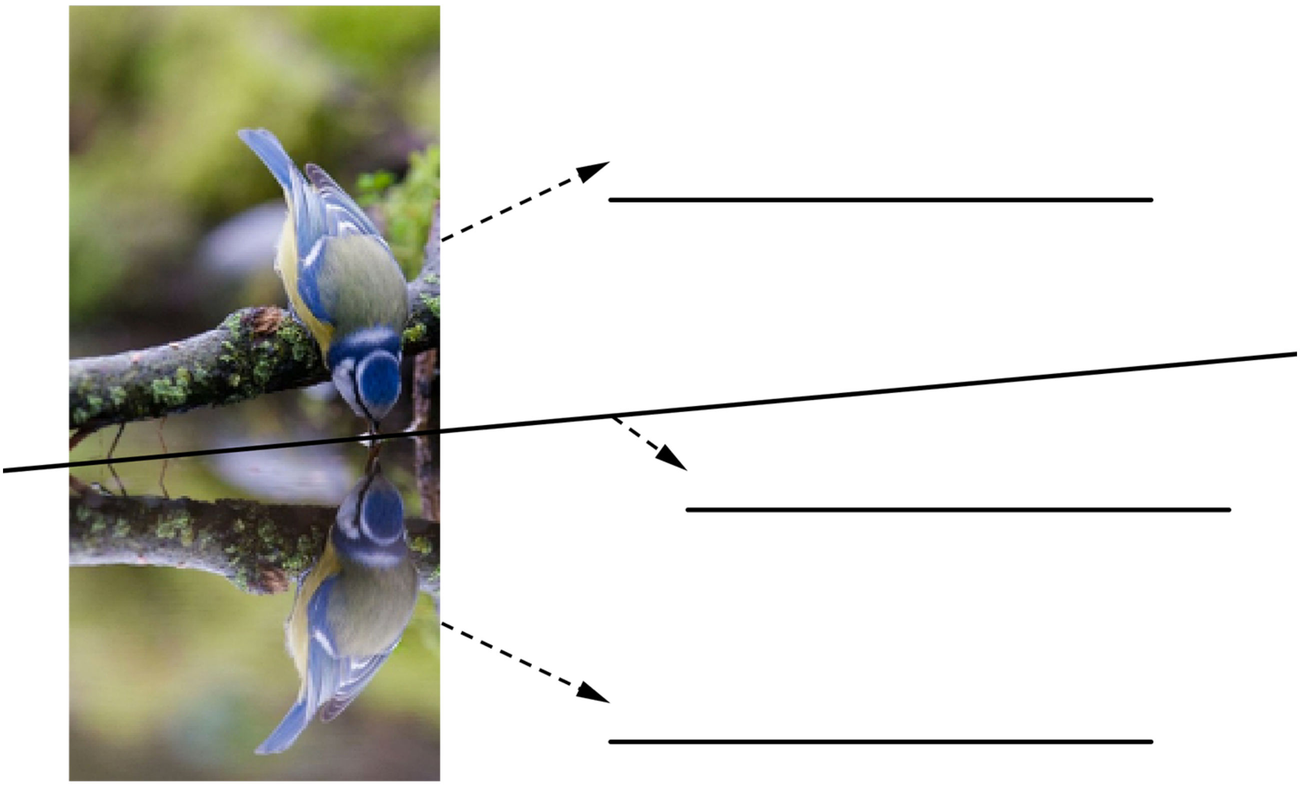
Aufgabe 10

Schreibe folgende Begriffe auf den passenden Strich.

1. Bildfigur
2. Originalfigur
3. Spiegelachse

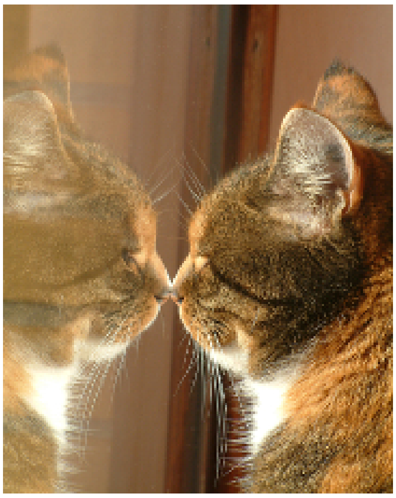
Écris les notions suivantes sur le bon trait.

1. Figure image
2. Figure origine
3. Axe de symétrie



Aufgabe 11

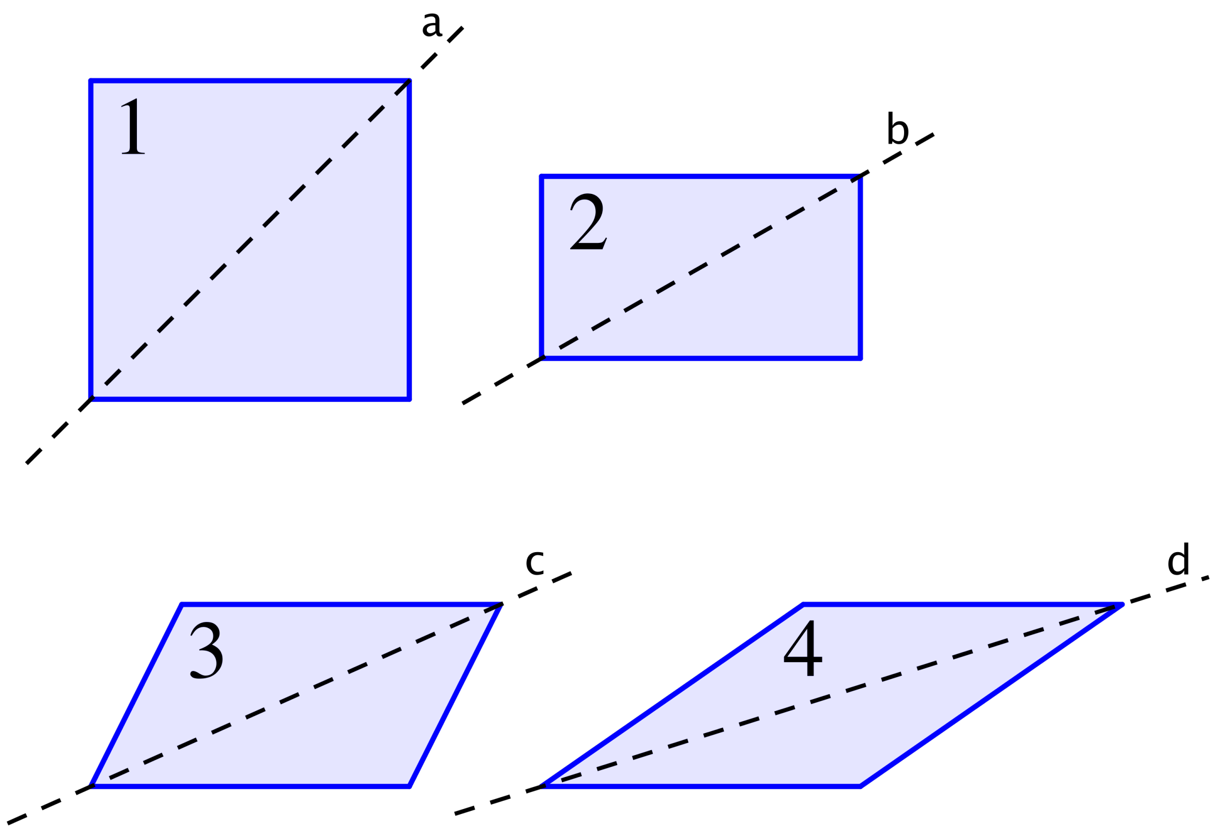
Zeichne jeweils die Spiegelachse ein.
Représente à chaque fois l'axe de symétrie.



Aufgabe 12

Richtig oder falsch?
Kreuze jeweils die richtige Antwort in der nebenstehenden Tabelle an.

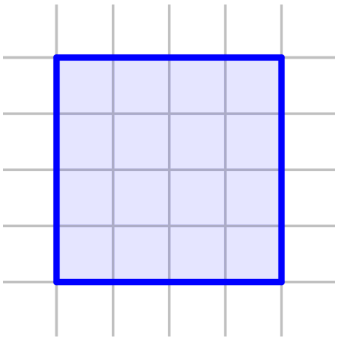
Correct ou incorrect ?
Coche à chaque fois la bonne réponse dans le tableau situé ci-contre.

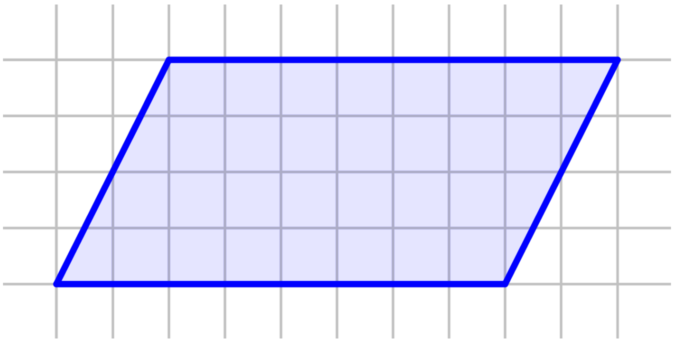


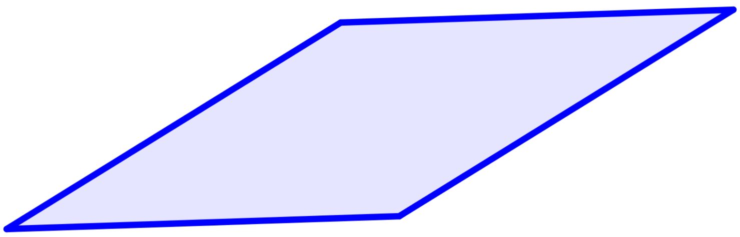
	Richtig? Correct ?	Falsch? Incorrect ?
a ist eine Spiegelachse der Figur 1. a est un axe de symétrie de la figure 1.	☐	☐
b ist eine Spiegelachse der Figur 2. b est un axe de symétrie de la figure 2.	☐	☐
c ist eine Spiegelachse der Figur 3. c est un axe de symétrie de la figure 3.	☐	☐
d ist eine Spiegelachse der Figur 4. d est un axe de symétrie de la figure 4.	☐	☐

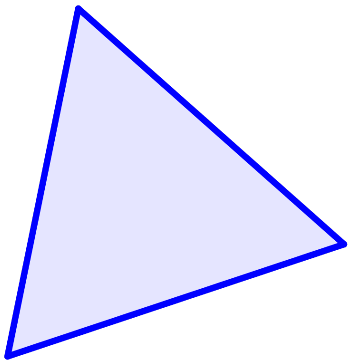
Aufgabe 13

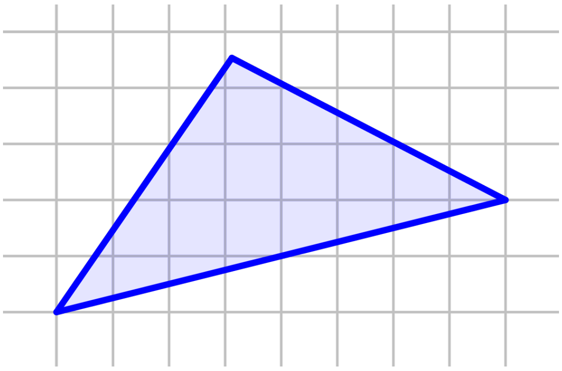
Zeichne **alle** möglichen Spiegelachsen ein.
Dessine **tous** les axes de symétrie possibles.

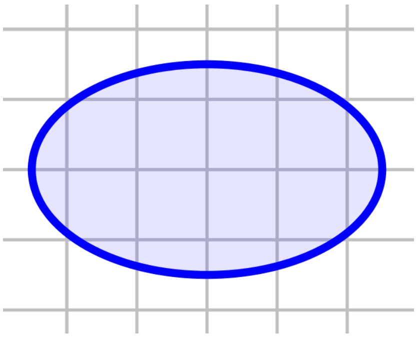


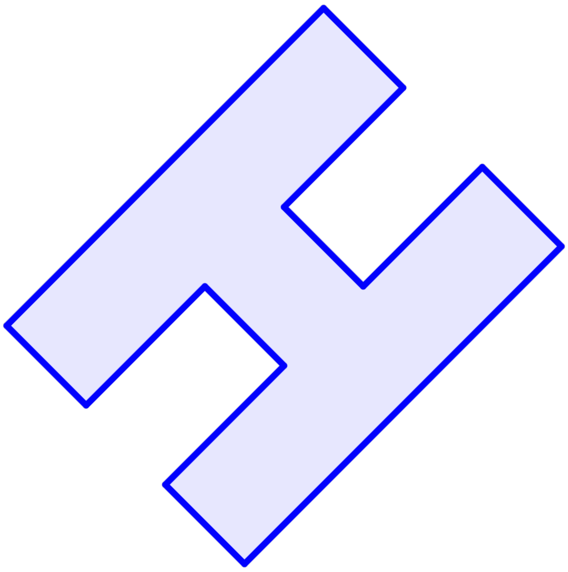








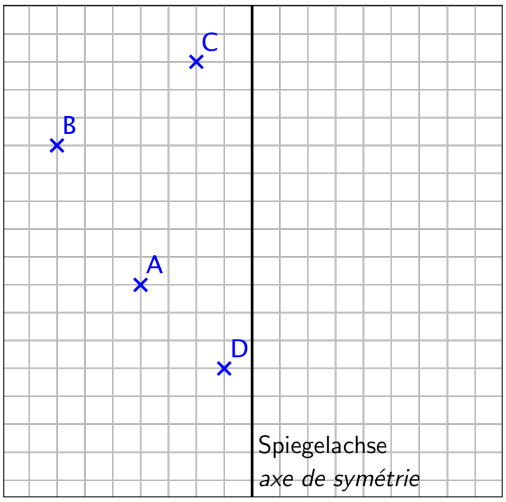


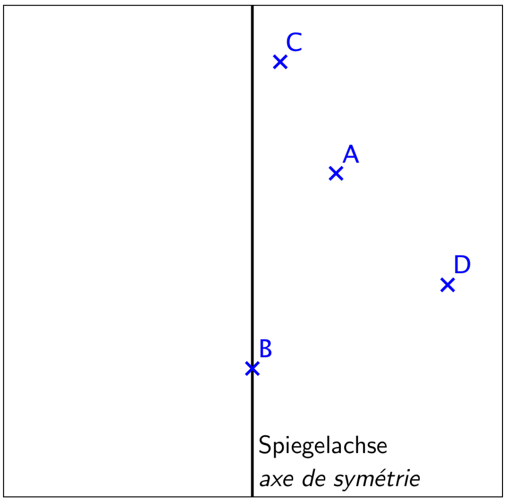


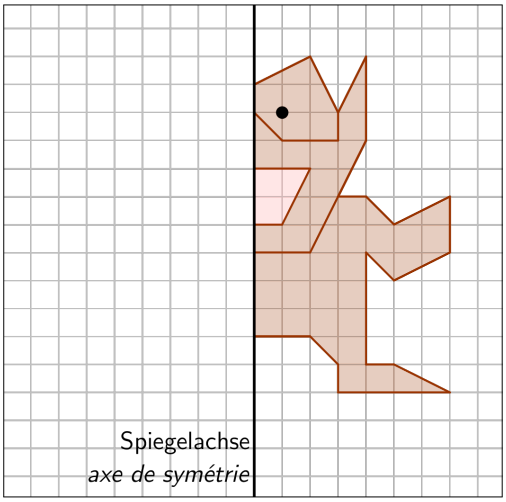
Aufgabe 14

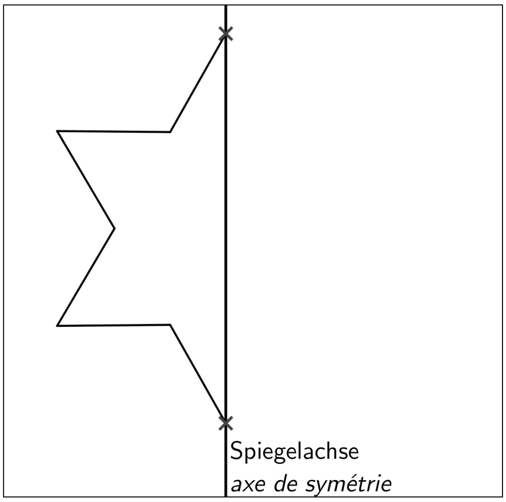
Spiegle folgende Figuren jeweils an der Spiegelachse.

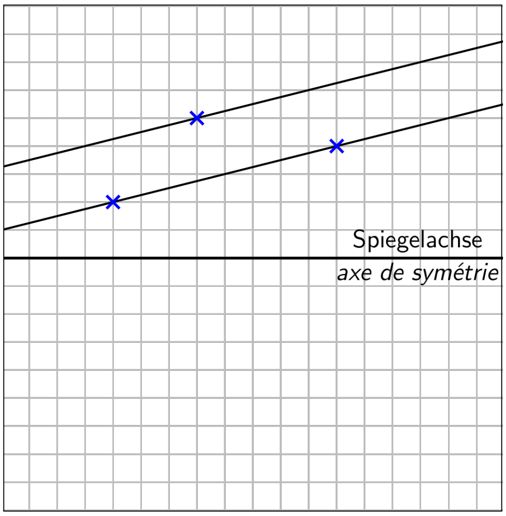
Trace à chaque fois le symétrique des figures suivantes par rapport à l'axe de symétrie.

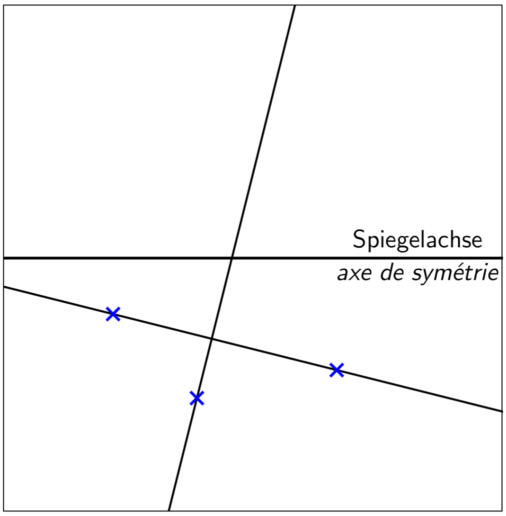


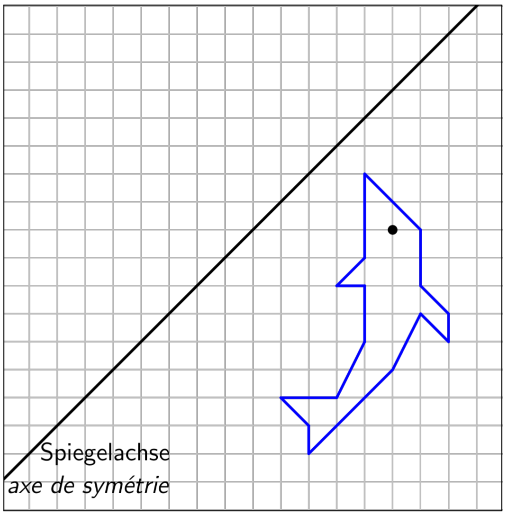


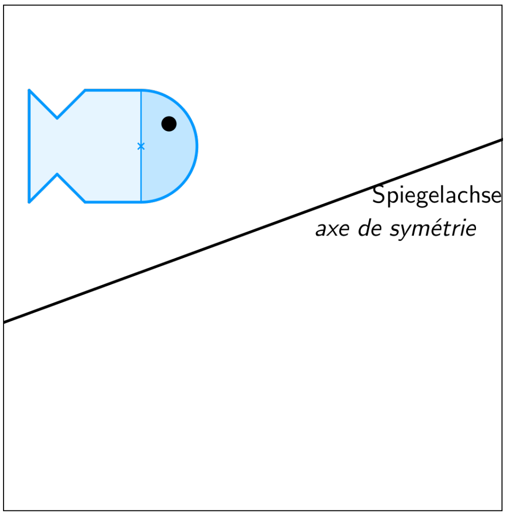





Spiegelachse
axe de symétrie

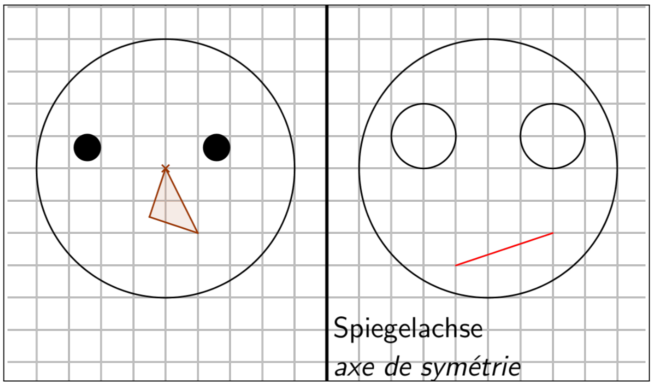

Spiegelachse
axe de symétrie

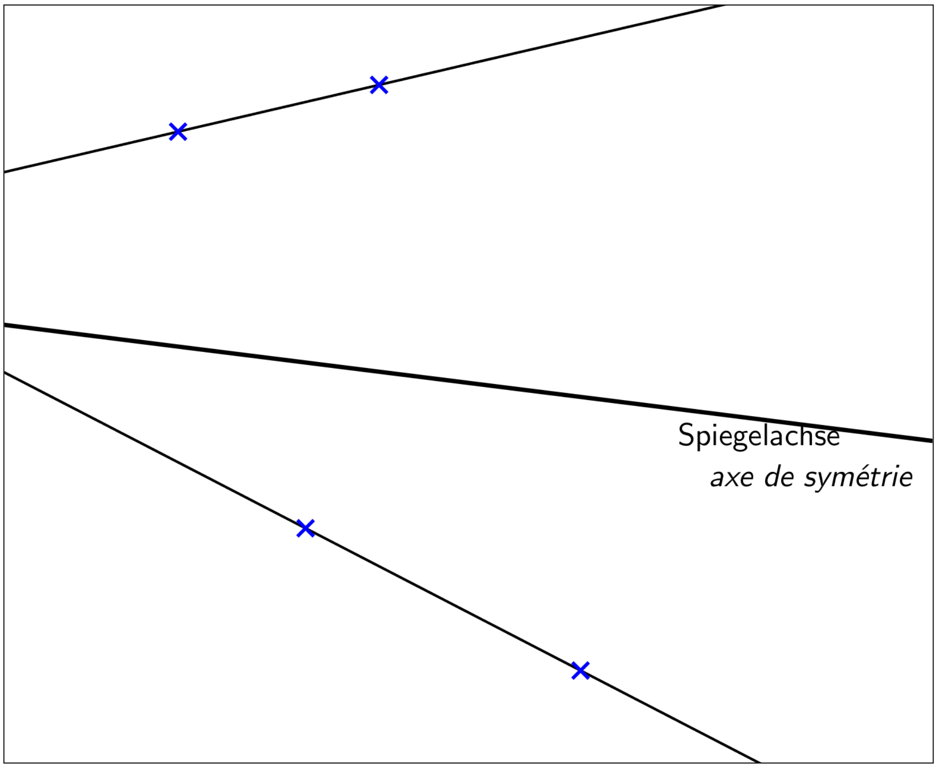

Spiegelachse
axe de symétrie


Spiegelachse
axe de symétrie

Aufgabe 15

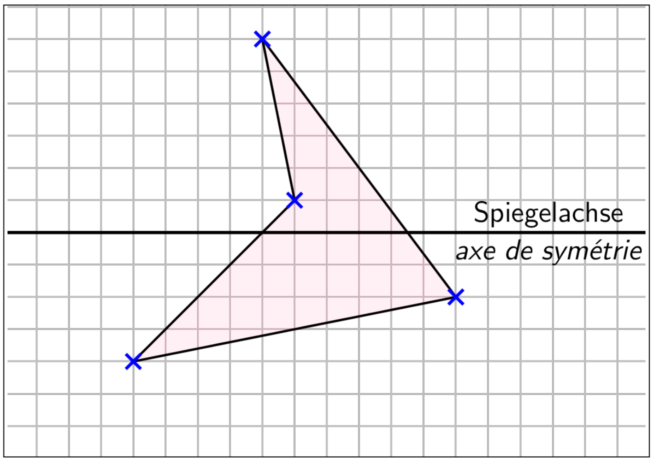
Vervollständige folgende Figuren so, dass ein Spiegelbild entsteht.
Complète les figures suivantes de telle manière à obtenir un reflet.

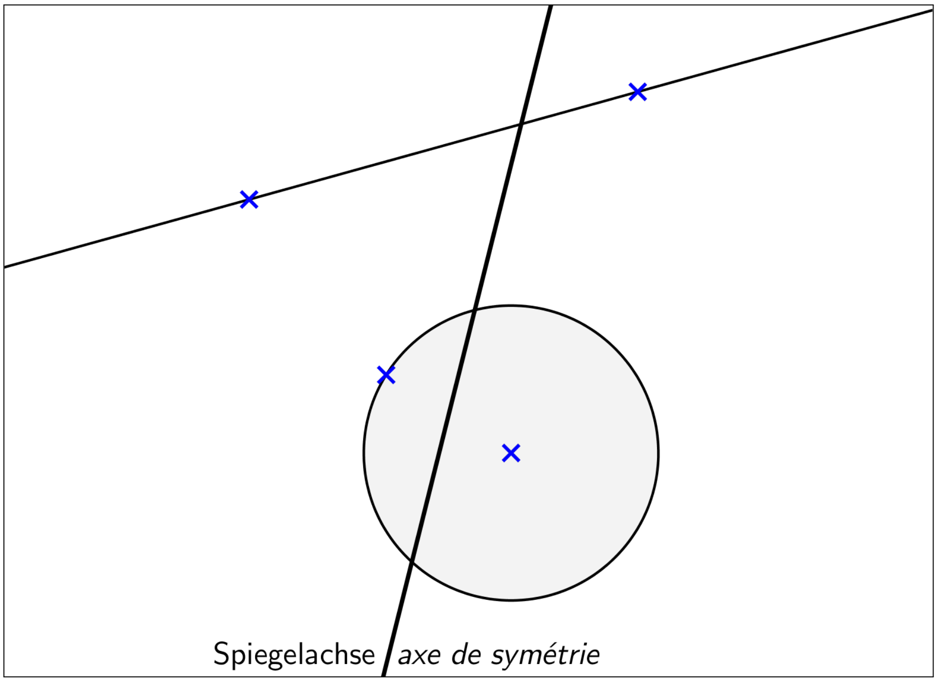




Aufgabe 16

Zeichne noch folgende Spiegelbilder.
Trace encore le symétrique des figures suivantes.

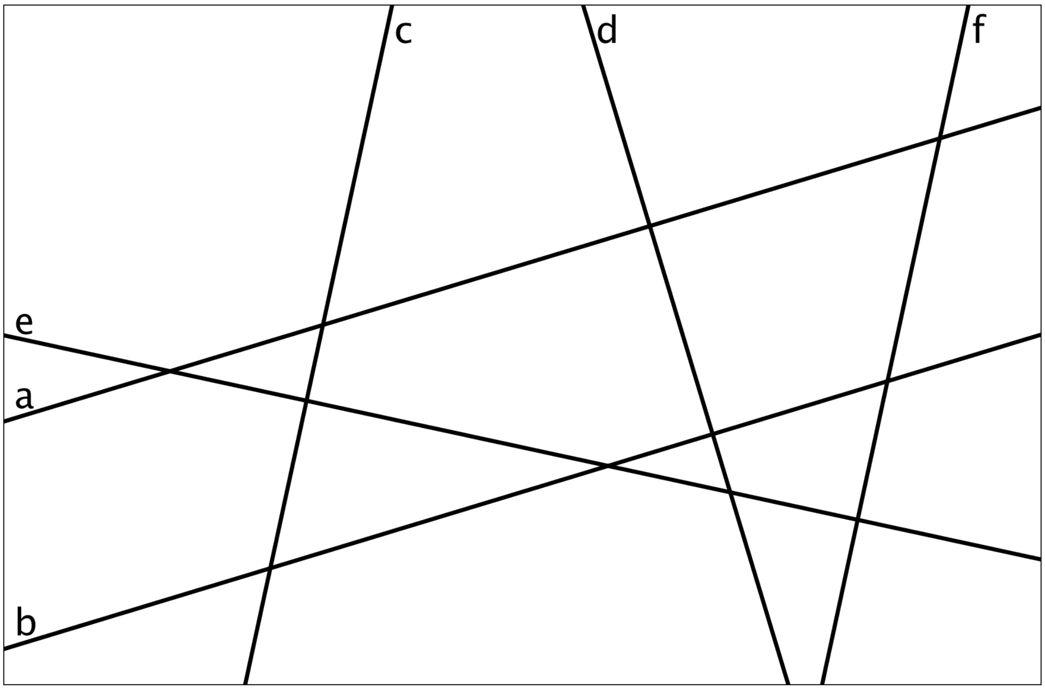




Aufgabe 17

Zueinander parallel, zueinander senkrecht oder weder noch?
Kreuze jeweils die richtige Antwort in der nebenstehenden Tabelle an.

Parallèles, perpendiculaires ou ni l'un ni l'autre ?
Coche à chaque fois la bonne réponse dans le tableau situé ci-contre.



	sind **parallel**. / sont **parallèles**.	sind **senkrecht**. / sont **perpendiculaires**.	sind weder noch. / ne sont ni l'un ni l'autre.
a und b / a et b	☐	☐	☐
a und d / a et d	☐	☐	☐
c und e / c et e	☐	☐	☐
a und c / a et c	☐	☐	☐
b und d / b et d	☐	☐	☐
c und f / c et f	☐	☐	☐
e und f / e et f	☐	☐	☐
c und d / c et d	☐	☐	☐

Aufgabe 18

Richtig oder falsch? Begründe.

Correct ou incorrect ? Justifie.

a) Eine Spiegelung verwandelt parallele Geraden in Geraden die sich schneiden.

Une symétrie transforme des droites parallèles en des droites qui se coupent.

☐ Richtig? / Correct ?   ☐ Falsch? / Incorrect ?

Begründung: / Justification : ____________________

____

b) Eine Spiegelung verwandelt senkrechte Geraden in parallele Geraden.

Une symétrie transforme des droites perpendiculaires en des droites parallèles.

☐ Richtig? / Correct ?   ☐ Falsch? / Incorrect ?

Begründung: / Justification : ____________________

____

c)

Eine Spiegelung verwandelt ein Quadrat wieder in ein Quadrat.
Une symétrie retransforme un carré en un carré.

☐ Richtig? / Correct ? ☐ Falsch? / Incorrect ?

Begründung: / Justification : ______________________

____

d)

Eine Spiegelung verwandelt zwei Kreise die sich *(in einem Punkt)* berühren in zwei Kreise die sich *(in zwei Punkten)* schneiden.
Une symétrie transforme deux cercles qui se touchent *(en un seul point)* en deux cercles qui se coupent *(en deux points)*.

☐ Richtig? / Correct ? ☐ Falsch? / Incorrect ?

Begründung: / Justification : ______________________

____

Appendix P

Teacher's questionnaire

Questionnaire d'évaluation de GeoGebraTAO et de son utilisation en classe par le personnel enseignant

Dans le cadre de la mise en place du projet GeoGebraTAO, nous entendons recueillir votre avis concernant le logiciel GeoGebraTAO lui-même, son utilisation et son intégration dans votre classe.

Votre collaboration est essentielle à l'évaluation de notre projet et éventuellement à l'amélioration de la qualité du logiciel GeoGebraTAO. Les résultats de cette enquête vont donc nous aider à identifier les éléments à développer pour mieux vous soutenir, ainsi que les problèmes rencontrés.

Il est donc très important que vous répondiez à toutes les questions avec autant de soin et de précision que possible. Veuillez répondre aux questions en cochant la case correspondant à votre réponse, et/ou en donnant votre opinion personnelle, vos suggestions, et/ou en formulant un commentaire éventuel.

Ce questionnaire est confidentiel. Seul les chercheurs du projet GeoGebraTAO auront besoin d'avoir accès à vos données, afin de pouvoir vous donner un feedback concernant le logiciel GeoGebraTAO lui-même et éventuellement des conseils sur la façon de l'intégrer en classe.

En vous remerciant à l'avance de votre précieuse collaboration.

Carole DORDING
Doctorante en sciences de l'éducation

LUXEMBOURG INSTITUTE OF SCIENCE AND TECHNOLOGY | LIST

uni.lu UNIVERSITÉ DU LUXEMBOURG

LUCET | LUXEMBOURG CENTRE FOR EDUCATIONAL TESTING

Nom de l'enseignant(e) : ________________________

École : ______________________________________

1. Globalement l'utilisation de GeoGebraTAO en classe a été :
   - ☐ très facile pour vous
   - ☐ facile pour vous
   - ☐ assez difficile pour vous
   - ☐ très difficile pour vous

   Si difficile, alors décrivez brièvement pourquoi ?

   ____
   ____
   ____
   ____

2. Quelle note portez-vous sur la navigation en général, notamment sur la disposition des boutons et sur l'enchaînement des activités ? (1 mauvais à 10 excellent)

☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
1	2	3	4	5	6	7	8	9	10

3. D'une manière générale, comment les élèves se repèrent-ils en travaillant avec GeoGebraTAO ? (1 très facilement à 10 très difficilement)

☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
1	2	3	4	5	6	7	8	9	10

4. GeoGebraTAO remplit-il le rôle que vous en attendez ?
   - ☐ oui, en grande partie
   - ☐ oui, moyennement
   - ☐ non, plutôt non
   - ☐ non, pas du tout

   Vos commentaires éventuels :

   ____
   ____
   ____
   ____

5. Citez (si possible) trois **qualités** de GeoGebraTAO qui vous paraissent importantes ?
   - ______________________________
   - ______________________________
   - ______________________________

6. Citez (si possible) trois **défauts** de GeoGebraTAO qui vous paraissent importants ?
   - ______________________________
   - ______________________________
   - ______________________________

7. Avez-vous rencontré des difficultés lors de l'utilisation de GeoGebraTAO en classe ?

   ☐ oui, souvent

   ☐ oui, de temps en temps

   ☐ oui, mais très rarement

   ☐ non, pas du tout

   Si oui, quels genres de difficultés ?

   ☐ des difficultés au niveau du logiciel GeoGebraTAO lui-même

   ☐ des difficultés au niveau de son intégration en classe

   ☐ les deux genres de difficultés

   Vos commentaires éventuels :

   ____

8. Quelles améliorations seraient intéressantes à apporter au logiciel GeoGebraTAO ?

   ____

9. Avez-vous suivi une formation sur l'intégration d'un logiciel mathématique ou non mathématique en classe auparavant ? Si oui, quel était le titre de cette formation / de ces formations ?

____

____

____

____

10. Dans les années à venir, comptez-vous intégrer GeoGebraTAO dans votre cours de mathématiques ?

☐ sûrement oui

☐ plutôt oui

☐ plutôt non

☐ sûrement non

Si non, qu'est-ce qui s'y oppose ?

____

____

____

____

11. Serait-il intéressant d'avoir des logiciels similaires à GeoGebraTAO en mathématiques ? (1 sûrement oui à 10 surtout pas)

☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
1	2	3	4	5	6	7	8	9	10

12. Serait-il intéressant d'avoir des logiciels similaires à GeoGebraTAO dans d'autres branches ? (1 sûrement oui à 10 surtout pas)

☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
1	2	3	4	5	6	7	8	9	10

13. Pensez-vous que l'utilisation de GeoGebraTAO aura un impact positif sur la qualité de votre travail d'enseignant(e) ?

☐ sûrement oui

☐ plutôt oui

☐ plutôt non

☐ sûrement non

Si vous pensez que l'utilisation de GeoGebraTAO aura un (des) impact(s) positif(s) sur la qualité de votre travail d'enseignant(e), lequel ou lesquels ?

- ☐ meilleure répartition du temps de travail
- ☐ meilleur respect du rythme de travail de chaque élève
- ☐ meilleure façon de susciter la motivation d'un élève indolent
- ☐ meilleure façon de susciter la motivation de chaque élève
- ☐ avoir le sentiment d'être soutenu(e) et soulagé(e) en classe
- ☐ ____
- ☐ ____
- ☐ ____

14. Si vous utilisiez à nouveau GeoGebraTAO en classe, qu'amélioreriez-vous probablement au niveau de son intégration ?

____

____

____

____

Merci d'avoir pris le temps de répondre à ce questionnaire et d'avoir participé avec votre classe au projet.

Carole Dording

Appendix Q

List of the GEOGEBRATAO tool exploratory learning assignments

Table Q.1: Overview of the GeoGebraTAO tool exploratory learning assignments

Assignment (main assignment, resp. '*external prim*' one)	**Domain(s) / GeoGebra related**	**Translated title**	**Number of compulsory items**	**Number of supplementary items** ('*internal prim*' items)	**Device(s) used** [b]	**Form of help available on a voluntary basis** [c]	**possible set back** (cf. subsection 2.2.8)	**Link** [a]
Aktivitéit $0a$	(Introduction activity)	GeoGebraTAO (part 1)	1 item	0 item	device C	no help	no	akt0a/
Aktivitéit $0b$	(Introduction activity)	GeoGebraTAO (part 2)	1 item	0 item	device B	no help	no	akt0b/
Aktivitéit 1	GeoGebra related	Movement Tool	1 item	0 item	device A	1 video clip	no	akt1/
Aktivitéit $1'$	GeoGebra related	Movement Tool	1 item	0 item	device A	1 video clip	no	akt1prim/
Aktivitéit $2a$	GeoGebra related	Movement Tool (part 1)	1 item	0 item	device A	1 video clip	no	akt2a/
Aktivitéit $2a'$	GeoGebra related	Movement Tool (part 1^*)	1 item	0 item	device A	1 video clip	no	akt2aprim/
Aktivitéit $2b$	GeoGebra related	Movement Tool (part 2)	2 items	$2 + 0$ items	device B	1 video clip	no	akt2b/
Aktivitéit $3a$	lines and segments	Lines and Segments (part 1)	2 items	$2 + 1$ items	device B	no help	no	akt3a/
Aktivitéit $3b$	lines and segments	Lines and Segments (part 2)	2 items	$2 + 1$ items	device B	no help	no	akt3b/
Aktivitéit $3c$	lines and segments	Lines and Segments (part 3)	2 items	0 item	device A	2 video clips	yes	akt3c/
Aktivitéit $3c'$	lines and segments	Lines and Segments (part 3^*)	2 items	0 item	device A	2 video clips	no	akt3cprim/
Aktivitéit 5	lines and segments	Lines and Segments	4 items	0 item	device A	2 video clips	yes	akt5/
Aktivitéit $5'$	lines and segments	Lines and Segments	4 items	0 item	device A	2 video clips	no	akt5prim/
Aktivitéit 6	GeoGebra related	Moving alone?	4 items	$1 + 1 + 1 + 1$ items	device B	1 video clip	no	akt6/
Aktivitéit $7a$	coordinates	Determining y-coordinates (part 1)	2 items	$0 + 2$ items	device B	no help	no	akt7a/
Aktivitéit $7b$	coordinates	Determining y-coordinates (part 2)	1 item	1 item	device C	no help	no	akt7b/
Aktivitéit $7c$	coordinates	Determining y-coordinates (part 3)	2 items	0 item	device C	1 short text	yes	akt7c/
Aktivitéit $7c'$	coordinates	Determining y-coordinates (part 3^*)	2 items	0 item	device C	1 short text	no	akt7cprim/
Aktivitéit $8a$	coordinates	Determining x-coordinates (part 1)	2 items	$0 + 2$ items	device B	no help	no	akt8a/
Aktivitéit $8b$	coordinates	Determining x-coordinates (part 2)	1 item	1 item	device C	no help	no	akt8b/
Aktivitéit $8c$	coordinates	Determining x-coordinates (part 3)	2 items	0 item	device C	1 short text	yes	akt8c/
Aktivitéit $8c'$	coordinates	Determining x-coordinates (part 3^*)	2 items	0 item	device C	1 short text	no	akt8cprim/
Aktivitéit 9	coordinates	Determining coordinates	5 items	varying	device C	2 short texts	yes	akt9/

[a] https://geogebratao.list.lu/akts/
[b] device A: actions in the GeoGebra frame; device B: choices on Likert scales; device C: choices in drop-down menus
[c] on the lower half of the left column (design and layout of the user interface)

Table Q.2: Overview of the GeoGebraTAO tool exploratory learning assignments

Assignment (main assignment, resp. '*external prim*' one)	**Domain(s) / GeoGebra related**	**Translated title**	**Number of compulsory items**	**Number of supplementary items** ('*internal prim*' items)	**Device(s) used** [b]	**Form of help available on a voluntary basis** [c]	**possible set back** (cf. subsection 2.2.8)	**Link** [a]
Aktivitéit 9*lux*	coordinates	Locating a point	5 items	varying	device A	2 video clips	yes	akt9lux/
Aktivitéit 10*a*	coordinates	Locating a point (part 1)	5 items	varying	device A	1 video clip	yes	akt10a/
Aktivitéit 10*b*	coordinates	Locating a point (part 2)	5 items	varying	device A	1 video clip	yes	akt10b/
Aktivitéit 11	GeoGebra related	Drawing point Tool	1 item	0 item	device A	1 video clip	no	akt11/
Aktivitéit 11′	GeoGebra related	Drawing point Tool	1 item	0 item	device A	1 video clip	no	akt11prim/
Aktivitéit 12	coordinates	Drawing point Tool	5 items	varying	device A	1 video clip	yes	akt12/
Aktivitéit 13	lines and segments	Drawing segment Tool	6 items	0 item	device A	1 video clip	no	akt13/
Aktivitéit 13′	lines and segments	Drawing segment Tool	6 items	0 item	device A	1 video clip	no	akt13prim/
Aktivitéit 13*bis*	lines and segments	Drawing segment Tool	5 items	varying	device A	1 video clip	yes	akt13bis/
Aktivitéit 14	lines and segments	Drawing line Tool	6 items	0 item	device A	1 video clip	no	akt14/
Aktivitéit 14′	lines and segments	Drawing line Tool	6 items	0 item	device A	1 video clip	no	akt14prim/
Aktivitéit 14*bis*	lines and segments	Drawing line Tool	5 items	varying	device A	1 video clip	yes	akt14bis/
Aktivitéit 15*a*	lines and segments	Building and transforming a figure (part 1)	1 item	0 item	device A	2 video clips	yes	akt15a/
Aktivitéit 15*b*	lines and segments	Building and transforming a figure (part 2)	1 item	0 item	device A	1 video clip	no	akt15b/
Aktivitéit 15*b*′	lines and segments	Building and transforming a figure (part 2*)	1 item	0 item	device A	1 video clip	no	akt15bprim/
Aktivitéit 16	recognizing symmetry	Symmetrical picture	1 item	0 item	device A	1 video clip	no	akt16/
Aktivitéit 16′	recognizing symmetry	Symmetrical picture	1 item	0 item	device A	1 video clip	no	akt16prim/
Aktivitéit 16*drehen*	recognizing symmetry	Natural symmetry	1 item	0 item	device B	1 static picture	no	akt16drehen/
Aktivitéit 17	recognizing symmetry	Symmetrical picture	1 item	0 item	device A	1 video clip	yes	akt17/
Aktivitéit 17′	recognizing symmetry	Symmetrical picture	1 item	0 item	device A	1 video clip	no	akt17prim/
Aktivitéit 18*a*	recognizing symmetry	Mirror image: moving a point (part 1)	1 item	1 item	device B	no help	no	akt18a/
Aktivitéit 18*b*	recognizing symmetry	Mirror image: moving a point (part 2)	4 items	0 item	device C	no help	no	akt18b/
Aktivitéit 19	recognizing symmetry	Mirror image: moving a point	1 item	1 item	device B	no help	no	akt19/

[a] https://geogebratao.list.lu/akts/

[b] device A: actions in the GeoGebra frame; device B: choices on Likert scales; device C: choices in drop-down menus

[c] on the lower half of the left column (design and layout of the user interface)

Table Q.3: Overview of the GeoGebraTAO tool exploratory learning assignments

Assignment (main assignment, resp. '*external prim*' one)	**Domain(s) / GeoGebra related**	**Translated title**	**Number of compulsory items**	**Number of supplementary items** ('*internal prim*' items)	**Device(s) used** [b]	**Form of help available on a voluntary basis** [c]	**possible set back** (cf. subsection 2.2.8)	**Link** [a]
Aktivitéit 20a	recognizing symmetry	Mirror image: moving a point (part 1)	1 item	1 item	device B	no help	no	akt20a/
Aktivitéit 20b	recognizing symmetry	Mirror image: moving a point (part 2)	2 items	0 item	device C	no help	yes	akt20b/
Aktivitéit 21a	recognizing symmetry	Transforming original figure (part 1)	1 item	0 item	device A	no help	no	akt21a/
Aktivitéit 21b	recognizing symmetry	Transforming original figure (part 2)	1 item	0 item	device A	2 video clips	no	akt21b/
Aktivitéit 21c	recognizing symmetry	Transforming original figure (part 3)	1 item	0 item	device A	2 video clips	no	akt21c/
Aktivitéit 21d	recognizing symmetry	Transforming original figure (part 4)	4 items	$1+0+0+0$ item	device B	no help	no	akt21d/
Aktivitéit 21e	recognizing symmetry	Transforming original figure (part 5)	2 items	0 item	device C	no help	yes	akt21e/
Aktivitéit 22	recognizing symmetry	Mirror image: bird's flight	4 items	0 item	device C	no help	no	akt22/
Aktivitéit 23	recognizing symmetry	Original figure	5 items	$1+0+0+0+0$ item	device B	no help	no	akt23/
Aktivitéit 24a	drawing symmetry	Reflecting object in line Tool (p. 1)	3 items	1 item	device A	1 video clip	no	akt24a/
Aktivitéit 24b	drawing symmetry	Reflecting object in line Tool (p. 2)	3 items	1 item	device A	1 video clip	yes	akt24b/
Aktivitéit 25a	drawing symmetry	Reflecting objects in line (part 1)	1 item	0 item	device A	1 video clip	yes	akt25a/
Aktivitéit 25b	drawing symmetry	Reflecting objects in line (part 2)	1 item	0 item	device A	2 video clips	no	akt25b/
Aktivitéit 25c	drawing symmetry	Reflecting objects in line (part 3)	6 items	0 item	device B	no help	no	akt25c/
Aktivitéit 25d	drawing symmetry	Reflecting objects in line (part 4)	1 item	0 item	device A	no help	no	akt25d/
Aktivitéit 25e	drawing symmetry	Reflecting objects in line (part 5)	3 items	$1+0+0$ item	device B	no help	yes	akt25e/
Aktivitéit 26	drawing symmetry	Drawing axis of symmetry	1 item	0 item	device A	1 short text	no	akt26/
Aktivitéit 26$'$	drawing symmetry	Drawing axis of symmetry	1 item	0 item	device A	1 short text	no	akt26prim/
Aktivitéit 27	drawing symmetry	Drawing axis of symmetry	1 item	0 item	device A	1 short text	yes	akt27/
						1 video clip		
Aktivitéit 27$'$	drawing symmetry	Drawing axis of symmetry	1 item	0 item	device A	1 short text	no	akt27prim/
						1 video clip		
Aktivitéit 28	drawing symmetry	Reflecting geometric objects in line	4 items	0 item	device A	1 video clip	no	akt28/
Aktivitéit 28$'$	drawing symmetry	Reflecting geometric objects in line	4 items	0 item	device A	1 video clip	no	akt28prim/
Aktivitéit 29	drawing symmetry	Reflecting geometric objects in line	4 items	0 item	device A	1 video clip	yes	akt29/
Aktivitéit 29$'$	drawing symmetry	Reflecting geometric objects in line	4 items	0 item	device A	1 video clip	no	akt29prim/

[a] https://geogebratao.list.lu/akts/
[b] device A: actions in the GeoGebra frame; device B: choices on Likert scales; device C: choices in drop-down menus
[c] on the lower half of the left column (design and layout of the user interface)

Table Q.4: Overview of the GEOGEBRATAO tool exploratory learning assignments

Assignment (main assignment, resp. '*external prim*' one)	**Domain(s) / GeoGebra related**	**Translated title**	**Number of compulsory items**	**Number of supplementary items** ('*internal prim*' items)	**Device(s) used** [b]	**Form of help available on a voluntary basis** [c]	**possible set back** (cf. subsection 2.2.8)	**Link** [a]
AKTIVITÉIT 30	drawing symmetry	Reflecting obj. with dyn. set square	3 items	1 item	device A	2 video clips	no	akt30/
AKTIVITÉIT 31	drawing symmetry	Reflecting obj. with dyn. set square	7 items	5 items	device A	2 video clips	yes	akt31/
AKTIVITÉIT 32	drawing symmetry	Reflecting obj. with dyn. set square	3 items	1 item	device A	2 video clips	no	akt32/
AKTIVITÉIT 33	drawing symmetry	Reflecting obj. with dyn. set square	3 items	1 item	device A	1 video clip	no	akt33/
AKTIVITÉIT 34	drawing symmetry	Reflecting obj. with dyn. set square	3 items	1 item	device A	1 video clip	yes	akt34/
AKTIVITÉIT 35	drawing symmetry	Reflecting obj. with dyn. set square	3 items	1 item	device A	1 video clip	no	akt35/
AKTIVITÉIT 36a	drawing symmetry	Reflecting parallel lines (part 1)	2 items	0 item	device A	2 video clips	yes	akt36a/
AKTIVITÉIT 36b	drawing symmetry	Reflecting parallel lines (part 2)	4 items	0 item	device A	no help	no	akt36b/
AKTIVITÉIT 37a	lines and segments	PARALLEL LINE Tool (p. 1)	3 items	1 item	device A	1 video clip	no	akt37a/
AKTIVITÉIT 37b	lines and segments	PARALLEL LINE Tool (p. 2)	3 items	1 item	device A	1 video clip	no	akt37b/
AKTIVITÉIT 38a	drawing symmetry	Reflecting perpendicular lines (p. 1)	2 items	0 item	device A	2 video clips	no	akt38a/
AKTIVITÉIT 38b	drawing symmetry	Reflecting perpendicular lines (p. 2)	4 items	0 item	device A	no help	no	akt38b/
AKTIVITÉIT 39a	lines and segments	PERPENDICULAR LINE Tool (p. 1)	3 items	1 item	device A	1 video clip	no	akt39a/
AKTIVITÉIT 39b	lines and segments	PERPENDICULAR LINE Tool (p. 2)	3 items	1 item	device A	1 video clip	no	akt39b/
AKTIVITÉIT 40	lines and segments	Symmetrical image?	3 items	$0+0+1$ item	device B	no help	no	akt40/
AKTIVITÉIT 41a	shapes and recognizing sym.	Determining sym. axis/axes (p. 1)	4 items	$0+0+1+0$ item	device B	2 static pictures	no	akt41a/
AKTIVITÉIT 41b	shapes and recognizing sym.	Determining sym. axis/axes (p. 2)	7 items	$0+0+0+0+1+0+0$ item	device B	3 static pictures	yes	akt41b/
AKTIVITÉIT 42a	shapes and recognizing sym.	Determining sym. axis/axes (p. 3)	3 items	$0+1+0$ item	device B	2 static pictures	no	akt42a/
AKTIVITÉIT 42b	shapes and recognizing sym.	Determining sym. axis/axes (p. 4)	7 items	$0+0+0+0+1+0+0$ item	device B	3 static pictures	yes	akt42b/
AKTIVITÉIT $last$	(final activity)	GEOGEBRATAO	0 item	0 item	no device	no help	no	aktlast/

[a] https://geogebratao.list.lu/akts/
[b] device A: actions in the GeoGebra frame; device B: choices on Likert scales; device C: choices in drop-down menus
[c] on the lower half of the left column (design and layout of the user interface)