# Bilingual student performance in the context of probability and statistics teaching in Danish High schools

Mie Haumann Petersen Kandidatspeciale - Matematik

Vejleder: Carl Winsløw

IND's studenterserie nr. 94, 2020

#### INSTITUT FOR NATURFAGENES DIDAKTIK, <u>www.ind.ku.dk</u> Alle publikationer fra IND er tilgængelige via hjemmesiden.

#### IND's studenterserie

- 55. Laura Mark Jensen: Feedback er noget vi giver til hinanden Udvikling af Praksis for Formativ Feedback på Kurset Almen Mikrobiologi (2017)
- 56. Linn Damsgaard & Lauge Bjørnskov Madsen: Undersøgelsesbaseret naturfagsundervisning på GUX-Nuuk (2017)
- 57. Sara Lehné: Modeling and Measuring Teachers' praxeologies for teaching Mathematics (2017)
- 58. Ida Viola Kalmark Andersen: Interdiciplinarity in the Basic Science Course (2017)
- 59. Niels Andreas Hvitved: Situations for modelling Fermi Problems with multivariate functions (2017)
- 60. Lasse Damgaard Christensen: How many people have ever lived? A study and research path (2018)
- 61. Adonis Anthony Barbaso: Student Difficulties concerning linear functions and linear models (2018)
- 62. Christina Frausing Binau & Dorte Salomonsen: Integreret naturfag i Danmark? (2018)
- 63. Jesper Melchjorsen & Pia Møller Jensen: Klasserumsledelse i naturvidenskabelige fag (2018)
- 64. Jan Boddum Larsen, Den lille ingeniør Motivation i Praktisk arbejdsfællesskab (2018)
- 65. Annemette Vestergaard Witt & Tanja Skrydstrup Kjær, Projekt kollegasparring på Ribe Katedralskole (2018)
- 66. Martin Mejlhede Jensen: Laboratorieforsøgs betydning for elevers læring, set gennem lærernes briller (2018)
- 67. Christian Peter Stolt: The status and potentials of citizen science: A mixed-method evaluation of the Danish citizen science landscape (2018)
- 68. Mathilde Lærke Chrøis: The Construction of Scientific Method (2018)
- 69. Magnus Vinding: The Nature of Mathematics Given Physicalism (2018)
- 70. Jakob Holm: The Implementation of Inquiry-based Teaching (2019)
- 71. Louise Uglebjerg: A Study and Research Path (2019)
- 72. Anders Tørring Kolding & Jonas Tarp Jørgensen: Physical Activity in the PULSE Exhibit (2019)
- 73. Simon Arent Vedel: Teaching the Formula of Centripetal Acceleration (2019)
- 74. Aputsiaq Bent Simonsen: Basic Science Course (NV) (2019)
- 75. Svenning Helth Møller: Peer-feedback (2019)
- 76. Lars Hansen & Lisbeth Birch Jensen: Feedbackformater på Mulernes Legatskole (2019)
- 77. Kirsi Inkeri Pakkanen: Autobiographical narratives with focus on science (2019)
- 78. Niels Jacob Jensen: Engineering i naturen og på naturskolen (2019)
- 79. Yvonne Herguth Nygaard: Diskursanalyse af litteraturen og hos lærer i forbindelse med brugen af eksterne læringsmiljø, med en underviser tilknyttet (2019)
- 80. Trine Jørgensen: Medborgerskab i naturfagsundervisningen på KBHSYD (2019)
- 81. Morten Terp Randrup: Dannelse i Fysik C (2019)
- 82. Thomas Mellergaard Amby: Undersøgelses baseret naturfagsundervisning og science writing heuristic (2019)
- 83. Freja Elbro: Important prerequisites to understanding the definition of limit (2019)
- 84. Mathilde Sexauer Bloch Kloster: Inquiry-Based Science Education (IBSE) (2019)
- 85. Casper Borup Frandsen: Undersøgelsesbaseret undervisning i idrætsundervisningen på gymnasieskolen (2019)
- 86. Vibeke Ankjer Vestermarken: An Inquiry Based Introduction to Binomial Distributions (2019)
- 87. Jesper Jul Jensen: Formativ evaluering og faglige samspil i almen studieforberedelse (2020)
- 88. Karen A. Voigt: Assessing Student Conceptions with Network Theory Investigating Student Conceptions in the Force Concept Inventory Using MAMCR (2020)
- 89. Julie Hougaard Overgaard: Using virtual experiments as a preparation for large scale facility experiments (2020)
- 90. Maria Anagnostou: Trigonometry in upper secondary school context: identities and functions (2020)
- 91. Henry James Evans: How Do Different Framings Of Climate Change Affect Pro-environmental Behaviour? (2020)
- 92. Mette Jensen: Study and Research Paths in Discrete Mathematics (2020)
- 93. Jesper Hansen: Effekten og brugen af narrative læringsspil og simuleringer i gymnasiet (2020)
- 94. Mie Haumann Petersen: Bilingual student performance in the context of probability and statistics teaching in Danish High schools (2020)

**IND's studenterserie** omfatter kandidatspecialer, bachelorprojekter og masterafhandlinger skrevet ved eller i tilknytning til Institut for Naturfagenes Didaktik. Disse drejer sig ofte om uddannelsesfaglige problemstillinger, der har interesse også uden for universitetets mure. De publiceres derfor i elektronisk form, naturligvis under forudsætning af samtykke fra forfatterne. Det er tale om studenterarbejder, og ikke endelige forskningspublikationer. Se hele serien på: <u>www.ind.ku.dk/publikationer/studenterserien/</u>

# UNIVERSITY OF COPENHAGEN DEPARTMENT OF SCIENCE EDUCATION



# MASTER THESIS

MIE HAUMANN PETERSEN

# BILINGUAL STUDENT PERFORMANCE IN THE CONTEXT OF PROBABILITY AND STATISTICS TEACHING IN DANISH HIGH SCHOOLS

SUPERVISOR: CARL WINSLØW SUBMITTED ON: 21/12-2020

# TABLE OF CONTENT

ABSTRACT	4
INTRODUCTION	5
THEORETICAL FRAMEWORK	8
REASONING FOR THE CHOSEN THEORETICAL FRAMEWORK	8
THE THEORY OF DIDACTICAL SITUATIONS	9
Personal and institutional knowledge	9
Milieu, didactical and adidactical situations	10
Didactical contracts	10
Phases of TDS	11
THE USE OF LANGUAGE IN MATHEMATICAL TEACHING	13
Mathematical symbolic representations	13
Semantics, syntax and morphology	14
Academic language, mathematical discourse and the mathematical register(s)	15
Transition between natural and academic language	17
LEARNERS WHOSE FIRST LANGUAGE IS NOT THE LANGUAGE OF INSTRUCTION	18
Definition of the term "bilingual" and related problems	18
The use of more than one language when learning mathematics	19
Strategies to help bilingual students in mathematics teaching	20
RESEARCH ON LEARNING STATISTICS	21
Lexically ambiguous words within statistics	22
Advantage for learners whose first language is not the language of instruction?	24
RESEARCH ON LEARNING PROBABILITY	24
Why is the teaching of probability is difficult for mathematics teachers?	27
Combinatorial reasoning and problem solving	28
Language and lexical ambiguity in the probability register	29
How to improve students understanding of probability: Real data and teacher education	31
Learning probability in a non-native language	32
ANALYSIS	33
STATISTICS	33
Methodology	33
ANALYSIS OF STUDENT ANSWERS TO APPENDIX 1A	36
Answers in relation to "spredning" (spread)	36
The case of bilingual student 1: Semantical difficulties	
The case of bilingual student 2: Semantical and syntactical difficulties	
Final comments on answers	
The case of monolingual student 1 and 2	41

Comparison of bilingual student replies and monolingual student replies	
Findings in relation to "spredning" (spread)	
Answers in relation to "hyppighed" (frequency)	
The case of bilingual student 8: Language related difficulties presented as either syntactical or semantical error	45 47
The case of bilingual student 5: Language related difficulties	
Other mathematical definitions by bilingual students	
The case of monolingual student 1 and 2: Correct definitions and usage of the word in both registers	50
Comparison of bilingual student replies and monolingual student replies	
Findings in relation to "hyppighed" (frequency)	53 5 2
Answers in relation to Jrekvens (relative frequency)	
The case of bilingual student 9 and bilingual student 3	
The case of monolingual student 1 and 2: Semantical difficulties and difficulties in the formulation	
Comparison of bilingual student replies and monolingual student replies	59
Findings in relation to "frekvens" (relative frequency)	60
Answers in relation to "uafhængighed" (independence)	
The meaning of "uafhængig" (independent) in relation to different subjects of mathematics	60 62
A PARTIAL CONCLUSION	
PROBABILITY	
Methodology	
INTRODUCTION TO THE MILIEU	
The use of computer and CAS-tools	
CATEGORIZATION OF THE TYPES OF PROBLEMS	
À PRIORI ANALYSIS OF SELECTED PROBLEMS OF THE WORKSHEETS	68
Problem 1 worksheet 1	68
Problem 7 worksheet 1	
The "sleening problem" from PowerPoint no 3 slide 3 (appendix 7)	
Droblem 1 workshoot no 2	75
Problem 1, worksheet no. 5.	
OBSERVATIONS	
Classroom observation no. 1: Language related difficulties in the formulation phase	
Situation 1: Bilingual students who struggle with word problem but not with straight forward calculations	79 82
Situation 3: Bilingual students who struggles with formulation	
Classroom observation no. 2: The term "random" and equiprobability bias	
Classroom observation no. 3: Language related difficulties in the formulation phase	
Classroom observation no. 4: The use of CAS and related problems	
A PARTIAL CONCLUSION	
DISCUSSION	
What did the data show?	00
WHAT DID THE DATA SHOW (	
Difficutties in the formulation phase	
The aspect of reading	
A combination of semantical and syntactical errors	
Semantical difficulties	100

GENERAL DISCUSSION OF THE RESULTS	101
The results in statistics	102
Suggestions to obtain better answers in relation to independence	
Discussion of results in probability	104
Quantitative vs qualitative data	105
Suggestions to improve bilingual student performance in the context of statistics and	
probability	106
CONCLUSION	108
REFERENCES	109
LIST OF FIGURES	113
LIST OF IMAGES	113
APPENDIX	115
APPENDIX 1A: LEXICAL AMBIGUITY WITHIN STATISTICS	115
APPENDIX 1B: LEXICAL AMBIGUITY WITHIN STATISTICS ENGLISH TRANSLATION	116
Appendix 2A: Worksheet 1 Danish	117
APPENDIX 2B: WORKSHEET 1 ENGLISH TRANSLATION	118
APPENDIX 3A: WORKSHEET 2 DANISH	120
APPENDIX 3B: WORKSHEET 2 ENGLISH TRANSLATION	121
APPENDIX 4A: WORKSHEET 3 DANISH	123
APPENDIX 4B: WORKSHEET 3 ENGLISH TRANSLATION	124
APPENDIX 5: POWERPOINT PRESENTATION FROM MODULE 1	125
APPENDIX 6: POWERPOINT PRESENTATION FROM MODULE 2	130
APPENDIX 7: POWERPOINT PRESENTATION FROM MODULE 3	136
APPENDIX 8: TRANSCRIPTION OF OBSERVATION NO. 1	141
APPENDIX 9: TRANSCRIPTION OF OBSERVATION NO. 2	143
APPENDIX 10: TRANSCRIPTION OF OBSERVATION NO. 3	144
APPENDIX 11: TRANSCRIPTION OF OBSERVATION NO. 4	146

# ABSTRACT

This thesis investigates the performance of bilingual students in the context of probability and statistics based on case studies following the methodology of TDS (the Theory of Didactical Situations), and focuses on the difficulties related to language which, according to the research literature, may arise for learners whose first language is not the language of instruction. The thesis investigates this in relation to critical notions such as spread, frequency, relative frequency and independence. The data has been obtained through a questionnaire answered by students at a Danish High School (H.C. Ørsted Gymnasium, Frederiksberg) and observations made at another Danish High School (Frederiksberg Gymnasium), where the focus of the following analysis has been lexically ambiguous words and the transitions between formal and "everyday" language. The case studies showed several cases of semantical and syntactical difficulties amongst the bilingual students which became evident and problematic particularly in the formulation phases. The thesis concludes that bilingual students experience language related problems which are both semantical and syntactical but as some of these language related problems were also experienced by monolingual students, the thesis does not conclude that language related problems in the context of probability and statistics occur exclusively with bilingual students. However, it would appear that some of the language related difficulties are enforced due to bilingualism.

# INTRODUCTION

The understanding of probability and statistics has become essential in the modern world, as the medias are full of statistical information and interpretation, where the statistical use of data is present in for example weather reports, predictions on election results, and in political decisions concerning economics. In order to judge the correctness of an argument which is supported by seemingly persuasive data, it is crucial to understand the underlying mathematics and the related vocabulary of statistics and probability. It is therefore central that students are introduced to strategies and ways of reasoning such that they are capable of making appropriate decisions in everyday and professional situations where chance is present, and where collection, organization, description, and analyzation of numerical data is happening. If not, the consequence will be citizens making decisions based on misinterpreted data and intuitive feelings about probability rather than on a factual foundation which is scientifically supported.

The need of teaching probability and statistics is evident, but as seen in research literature, the understanding of probability is difficult as "...*counterintuitive results in probability are found even at very elementary levels, whereas in other branches of mathematics counterintuitive results are encountered only when working at a high degree of abstraction.*" (Batanero, Godino & Roa, 2004, p. 2). In addition, the language related to statistics and probability is ambiguous in the sense that the semantics of some of the vocabulary is changing when transitioning from everyday language to a more formal mathematical language. Words such as "uafhængig" (independent), "spredning" (spread), and "hyppighed"(frequency) all have a specific meaning within the register of mathematics and another meaning in natural language. The cognitive challenge for the students therefore also lies with accepting and changing the semantics of a certain word, based on which register they are applying, and the transition from natural language to formal language may cause trouble.

The aspect of the use of language within mathematics thus becomes even more important in relation to statistics and probability and the need for research in this particular area has been acknowledged as research has been conducted in several countries. Several studies in the area of the language related problems in statistics and probability has been conducted for example by Batanero et al (2016), Dunn et al (2016), Batanero and Sanchez (2013), Díaz and Batanero (2009) and Langrall and Mooney (2007). Additionally, in 2016 it was reported that 11.3% of the 150,000 students at danish High Schools, had an ethnic origin other than danish: An increase compared to 2006 where 8.1% of the students had another ethnic origin (Danmark Statistik, 2018). 7.9% of the students in 2016 were

descendants of immigrants and 3.4% were immigrants and as a comparison the numbers in 2006 were 4.0% for descendants of immigrants and 4.2% for immigrants (Danmarks Statistik, 2018). The amount of students whose first language is not the language of instruction therefore seems to be increasing in Danish High Schools, and the linguistic aspect of mathematics thus becomes even more important, as another report from the Danish Evaluation Institute (Danmarks Evalueringsinstitut, 2006) shows that the bilingual students in Danish High Schools generally get lower grades than ethnically Danish students (Danmarks Evalueringsinstitut, 2006, p 21). In the report, it is argued that ethnically Danish students get an average of 8.1 in their grades while bilingual students, who are either born in Denmark or has moved to Denmark before the age of 6 averagely get a 7.8 in their grades (Danmarks Evalueringsinstitut, 2006, p. 21). The same report also points out that it is essential that the teachers develop competences such that they are capable of dealing with bilingual students. This development of competences includes for example different didactical aspects of the Danish linguistic aspects of the subjects being taught, as only few of the teachers have the insight needed to become aware of and meet the academic and social needs of many bilingual students. (Danmarks Evalueringsinstitut, 2006, p. 10).

In relation to mathematics, it has also been found that bi- and monolingual students are doing equally well when solving mathematical problems with straight forward calculations, when the problem is solvable without any interpretation of the provided text of the problem. However, there are significant differences between the bi- and monolingual students when the given problem is to be interpreted before calculating (Andersen, 2004, p. 286). It would thus seem that the bilingual students may experience some language related difficulties which are not experienced by the monolingual students.

This thesis therefore intends to investigate the performance of bilingual students in the context of probability and statistics based on case studies following the methodology of TDS (the Theory of Didactical Situations). The case studies will be used to analyze the difficulties related to language which, according to the research literature, may arise for learners whose first language is not the language of instruction. This will be investigated in relation to critical notions such as spread, frequency, relative frequency and independence and the thesis will focus on the transitions between formal and natural language.

The case studies will be provided by observations of student conversations and through written answers to a questionnaire concerning lexically ambiguous words within statistics. As the scope of the thesis is concerned with the performance of bilingual students, the replies of the questionnaire will be compared to the replies provided by the monolingual students. Based on the case studies, this thesis investigates if bilingual students experience syntactical and semantical difficulties when using and defining lexically ambiguous words used within statistics and probability, which are not experienced by monolingual students. The findings and the design of the investigation will be discussed thoroughly as it could be argued that the findings in the data, due the qualitative design of the investigation, cannot be generalized with certainty to a larger population, as other variables can affect the outcome. However, as this thesis will present, the theoretical framework and the related literature will support some of the findings.

# THEORETICAL FRAMEWORK

In this section the reasoning of the theoretical framework for the thesis and the selected theoretical framework will be presented. First, there will be an introduction to the Theory of Didactical Situations (TDS), where the related terminology will be introduced. As this thesis has a particular focus on the use of language within mathematics, this section will contain a brief introduction to notions such as: Semantics, syntax, natural language, academic language, mathematical classroom discourse, and mathematical register. The thesis is focusing on learners whose first language is not the language of instruction and therefore I present and discuss the definition of the terms mono-, bi- and multilingual and the related issues concerning these definitions. Additionally, there will be an overview of some of the research on learning mathematics while using more than one language. As the research question of this thesis is in the context of statistics and probability, the current research within these areas will also be presented. In particular, there will be an introduction to the notion of lexically ambiguous words.

#### REASONING FOR THE CHOSEN THEORETICAL FRAMEWORK

In this thesis the main focus is on the use of language and transitions between formal and everyday language with a focus on bilingual students. Several other aspects to probability and statistics could have been reviewed as these subjects are generally difficult for students, but as presented in the introduction, the number of learners whose first language is not the language of instruction seems to be increasing in Danish High Schools, thus the focus is highly relevant.

The theoretical framework has been selected through the guidance of the supervisor connected to this thesis and through a comprehensive journals cross reference search. The theoretical framework has been researched and selected by the use of for example Google Scholar through a cross reference search on relevant terms such as: "probability", "statistics", "EAL", "second language learners", "misconceptions", "semantics and syntax" etc. Some of the theoretical framework is based on articles published in mathematics research journals such as Educational Studies in Mathematics, Journal for Research in Mathematics Education, Journal of Mathematics Teacher Education and ZDM: The International Journal on Mathematics. All of these journals have been graded with an A or A\* by the Education Committee of the European Mathematical Society (EMS) and the Executive Committee of the European Society for Research in Mathematics Education (ERME) and supported by the International Commission for Mathematical Instruction (ICMI) (Törner & Arzallo, 2012, p. 52). The

chosen literature contributing to the theoretical framework has thus been carefully selected and critically reviewed before it was applicated.

# THE THEORY OF DIDACTICAL SITUATIONS

The Theory of Didactical Situations (TDS) was originally introduced by Guy Brousseau in the late 1960's and the theory has aided in producing ideas and results helping teachers plan and develop the construction of mathematical knowledge (Jessen & Winsløw, 2017, p. 30). Mathematical knowledge is usually presented through an axiomatic presentation where the objects of study is defined with previously introduced notions. This should be a way of letting a teacher and student order "...their activities and accumulating in the shortest time the maximum number of items of knowledge which are reasonably close to the experts' knowledge" (Brousseau, 2002, p. 21) when combining this with examples and problems which require the usage of the knowledge presented. This type of teaching has one major problem according to Brousseau: It hides and remove all the traces of the history of how the knowledge has been established and hides "... the true functioning of science, which is impossible to communicate and describe faithfully from the outside and replaces it with an imaginary genesis." (Brousseau, 2002, p. 21). The problem is, according to Brousseau, that in order to make teaching easier certain notions and properties has been isolated and taken away from "... the network of activities which provide their origin, meaning, motivation and use." (Brousseau, 2002, p. 21). The idea behind TDS is that the intellectual work of the students must be more like the scientific activity as "... knowing mathematics is not simply learning definitions and theorems in order to recognize when to use and apply them." (Brousseau, 2002, p. 21). The idea behind TDS is thus inquiry based as the process of inquiry is considered essential if the student is supposed to gain knowledge of these results and convince herself/himself of their validity of them.

## PERSONAL AND INSTITUTIONAL KNOWLEDGE

Within TDS there is a distinguishing between two types of knowledge: *Institutional* and *personal*. Institutional knowledge is also referred to as public or official knowledge and it is the knowledge which is presented in "... *textbooks, webpages, research journals and other shared resources*" (Jessen & Winsløw, 2017, p. 30). The mathematical knowledge presented is a fusion of individually made mathematical activities which has been validated by others and the mathematical knowledge is presented in a concise and precise way but the entire inquiry process leading to this form is usually not presented and thus not visible (Jessen & Winsløw, 2017, p. 30). In contrast, personal knowledge is the knowledge constructed "...while interacting with a mathematical problem" (Jessen &

Winsløw, 2017, p. 31) and is often implicitly given and is related to a given context. The personal knowledge is often a bit different from the institutional knowledge, but it can be developed and formalized if students share and discuss their knowledge with others (Jessen & Winsløw, 2017, p. 32). The importance of communication is thus emphasized as this is where the development of the student's initial ideas is happening.

#### MILIEU, DIDACTICAL AND ADIDACTICAL SITUATIONS

In order to create a situation, in which the student can develop their personal knowledge, the teacher needs to plan and create a didactical milieu in which the students are able to investigate and discover the targeted knowledge chosen by the teacher. The didactical milieu consists of "... the student's previous knowledge and artefacts such as pen and paper, ruler, calculator, CAS-tools, a puzzle etc." (Jessen & Winsløw, 2017, p. 32). The milieu should have a high potential for students to construct the intended knowledge without the teacher lecturing it, as it should produce a need amongst the students to create the intended knowledge in order to solve a given problem (Jessen & Winsløw, 2017, p. 32). A properly created milieu will make the students able to interact autonomously without further interference from the teacher, and such situations are referred to as adidactical situations. The adidactical situations are the situations when students interact and engage in a given problem and investigate within the presented milieu without any teacher interference. Didactical situations are in contrast the situations where the teacher "... explicitly interacts with the students in order to further their learning of something specific" (Jessen & Winsløw, 2017, p. 33). The purpose of the didactical situations is to ensure that the knowledge students are developing is shared and validated and it also serve as a way to regulate and moderate the adidactical situations. Thus, the learning potential lies "... in the dialectic between adidactical and didactical situations or between personal and shared knowledge" (Jessen & Winsløw, 2017, p. 33).

#### DIDACTICAL CONTRACTS

Teaching through inquiry-based situations poses a lot of challenges for the teacher as it requires the teacher to design and devolve appropriate mathematical situations in which the students are capable of developing their personal knowledge: A process which can be both difficult and time consuming. It may also pose a challenge for the teacher not to intervene when students are choosing a wrong or less favorable approach to solve a problem. In addition, the students may struggle to adapt to this new way of learning as they are probably more used to another structure where they try to do what they believe the teacher expects and rewards. These expectations of what students and teachers are

supposed to do, and this system of reciprocal obligation resembles a contract, this this is what we call the *didactical contract* of the situation (Jessen & Winsløw, 2017, p. 36).

If students are for example used to the fact that their teacher provides them with answers from the beginning "... a certain amount of frustration can occur when they are given open ended inquiry based activities" (Jessen & Winsløw, 2017, p. 36). The students might expect the teacher to give them the expected strategy and it can be tempting for the teacher to simply explain the students what to do as this is easier and less time consuming. But as this will ruin the learning potential for the students a solution might be to explain to the students in the beginning that they are expected to "... engage in solving problems even if they feel unprepared" (Jessen & Winsløw, 2017, p. 36) and thus explicitly addressing a change in the didactical contract.

#### PHASES OF TDS

The main idea of TDS is an inquiry-based creation of situations which addresses a "... well-known obstacle regarding a piece of mathematical knowledge, which create the need for the students to develop or construct new mathematical knowledge" (Jessen & Winsløw, 2017, p. 36). The teaching situation can be divided into five phases in which the sequencing is not strictly given. The first phase of TDS is called the *devolution phase* where the teacher presents and hands over the problem and the milieu to the students. The teacher will explain the rules which apply to the problem and it is important that the teacher makes sure that the students have understood the problem and the rules such that they are able to start engaging actively with the problem after the devolution. The milieu can be handed over through an example or simply by presenting the artefacts related to the concrete problem and situation (Jessen & Winsløw, 2017, p. 37). Next is the action phase where the teacher stands back, observe and let the adidactical situation begin. The students are working independently and adidactically in order to gain some experience with the mathematical knowledge to be taught. It is now the importance of the milieu becomes evident: A well-designed milieu will allow the students to interact without any teacher inference. This phase also has some similarities with the process of researchers first approach to an open problem. The next phase is the formulation phase where "... students are required to present what they did in the action phase; initial ideas, hypothesis or simply what they have tried to do so far" (Jessen & Winsløw, 2017, p. 38). The main goal of the formulation phase is to force students to articulate the experiences and ideas obtained from the interaction with the problem. They thus begin constructing elements of the mathematical theory to be thought.

What follows next is the *validation phase* where the hypotheses (or simply their ideas or strategies) of the students are tested against the milieu and can therefore adidactically be validated without the teacher telling them if they are right or wrong (Jessen & Winsløw, 2017, p. 39). The fifth and final phase of TDS is the *institutionalization phase* where the teacher will gather ideas, sum up main points and present the institutional knowledge to be learned. This presentation "...*will often be a presentation of mathematical knowledge being concise and accurate as in textbooks*" (Jessen & Winsløw, 2017, p. 39). The mathematical knowledge is thus established from the students experience and reasoning instead of being presented as a matter of fact, and the amount of mathematical detail in this presentation should be aligned with the activities carried out by the students (Jessen & Winsløw, 2017, p. 39). The institutionalization should not end up being a lecture such that the actions of the students become useless, as this will mean that the students are "... *not likely to treasure or engage in mathematical inquiry and autonomous construction of knowledge but will imitate the teacher when doing mathematics*" (Jessen & Winsløw, 2017, p. 39). In order to provide an overview, Jessen and Winsløw (2017, p. 41) has made figure 1 below where the functions and actions of teachers, students, and milieu is summarized:

	Role of teacher	Role of student	Milieu	Situation
Devolution	Introduces, hands	Receive, try to	Is being estab-	Didactical
	over the milieu	take on a prob-	lished	
		lem		
Action	Observes and re-	Act and reflect	Problem is being	Adidactical
	flects		explored	
Formulation	Organizes, if	Formulate as spe-	Open discussion	Adidactical or di-
	needed initiates	cifically as possi-		dactical
	through ques-	ble		
	tions			
Validation	Listens and eval-	Argue, try to fol-	Guided discus-	Often didactical
	uates if needed	low others' argu-	sion	
		ments		
Institutionalisa-	Presents and ex-	Listen and reflect	Institutionalised	Didactical
tion	plains		knowledge	

FIGURE 1: AN OVERVIEW OF THE PHASES OF TDS AND THE ROLES OF THE PARTICIPANTS (JESSEN & WINSLØW, 2017, P. 41)

As a benefit of the inquiry-based approach to teaching there is a lot of evidence that over time students will develop a relationship with mathematics which is more positive, and they will also view mathematics as a whole instead of a meaningless and endless list of given answers. Mathematics will thus become more rational and challenging but also a more satisfying activity in which the students are more likely to engage in (Jessen & Winsløw, 2017, p. 36).

#### THE USE OF LANGUAGE IN MATHEMATICAL TEACHING

In a didactical situation, several factors contribute to the didactical contract in praxis: Does the teacher have a good contact with the students? Do the students work well together? These things are some-what connected to the use of language as this is the primary medium through which students exchange their knowledge. But the formal talk about mathematics is not the only sort of communication going on in a didactical situation. Students will also talk about how to understand a given problem and also about things which have little or no connection to a given mathematical problem (Winsløw, 2006, p. 155). It is easy to view this informal talk as irrelevant talk blocking actual learning but this is not the case: The informal talk actually has a huge role in the learning process of students, and the success of the teaching situation also depends on the entire communicative situation going on in the teaching situation and not just the dialogue in regard to mathematics (Winsløw, 2006, p. 155). It becomes impossible to separate mathematics and language as formal definitions can not exist or be understood without the use of different types of linguistic expressions and especially without naturally spoken and written language (Winsløw, 2006, p. 155).

#### MATHEMATICAL SYMBOLIC REPRESENTATIONS

Symbolical representations which are non-verbal such as mathematical formulas, diagrams, graphs, and so on are no exception: The usage and mastery of these are an inevitable part of mastering mathematics and can not be achieved without the use of a more natural everyday language (Winsløw, 2006, p.156). The word *representation* will be used as something standing instead of a material object. A representation is always expressed through a *medium* such as paper, a computer screen or soundwaves and we can for practical reasons use a distinction between verbal and written medias which will be sufficient. Within mathematics the object to be studied are only available in their representation through algebraic expressions, geometric figures or for example through a graphical representation (Winsløw, 2006, p. 157), and there is therefore an ontological question regarding what these representations actually represent? This ontological question will remain unanswered in this thesis as it is not within the scope of the thesis. Instead, three important skills in the use of

mathematical representations will be noted. The first one is to choose a representation which fits the purpose, the second is to understand different representations of the same object and thirdly to change between different representations.

#### SEMANTICS, SYNTAX AND MORPHOLOGY

When studying a language and the building of language it is a versatile and nuanced system which can be viewed from a lot of different perspectives. It does not matter if we are looking at language embedded in writing or speech, it is impossible to consider all of the aspects at once (Christensen & Christensen, 2016, p. 23) and therefore certain parts of the language are being categorized, and specific disciplines of the study of language has been evolved. Christensen & Christensen (2016) are using the following categorization of the study of language with a division into six aspects to be studied: 1) Phonetics (the sound of the words of the language), 2) Orthography (the visual expression of the written words including how letters are to be combined into words thus spelling is essential here), 3) Morphology (the meaning bearing parts of a word including conjugation of words and word formation), 4) Lexicon (the vocabulary of an individual or an entire language), 5) Syntax (the combination of words into a larger units) and finally 6) Semantics (the mental content or meaning of the linguistic usages).

The relevant categories of linguistic analysis for this thesis are morphology, syntax and semantics. Morphology is a structural analysis of the construction of single standing words and it is the study of how the morphemes, the smallest meaning bearing elements, are organized and combined into words (Christensen & Christensen, 2016, p. 31). Morphology therefore have two parts: Word formation and word conjugation. Syntax is the rules of how words are combined into valid sentences (Christensen & Christensen, 2016, p. 161). In particular, Danish (and English) is a language where the order of the words is crucial to the meaning of the sentence. As an example of this one might consider the two sentences "hunden bed babyen / *the dog bit the baby*" and "babyen bed hunden / *the baby bid the dog*" which has two very different meanings depending on the placement of "babyen / *the baby*" and "hunden / *the dog*". Thus, the transfer of linguistic meaning does not only depend on the words we choose, but also the order of the combination (Christensen & Christensen, 2016, p. 161).

*Semantics* is the study of linguistic meaning, where the *lexical semantics* is the meaning of a word (Christensen & Christensen, 2016, p. 247). The word "lexicon" is a synonym to the word "vocabulary" and within linguistic this is meant to cover three related meanings of the word: 1) the total vocabulary of a language, 2) dictionaries showing the (or several) vocabularies of a language and 3)

the vocabulary of an individual in a linguistic community (Christensen & Christensen, 2016, p. 247). No matter if you are learning a language as a first or second language it is grounded in a social context, where a condition for engaging actively and efficiently in all communication is acquiring the current praxis on the words and their semantics (Christensen & Christensen, 2016, p. 247). This praxis is what is called the lexical norm and if the lexical norm is not followed by someone, it becomes difficult to understand what is being communicated.

We are able to talk about the meaning of a certain word isolated from its context, but in reality, we rarely define and talk about a word as something standing alone - it is grounded in a conversation or a text in a relation with other words. The linguistic settings in which a word occurs is called the *linguistic context* and, in this thesis, it will be defined in a wide sense including entire conversations and texts in accordance with the definition from Christensen and Christensen (2016, p. 252). Beside the linguistic context there is also a pragmatic context which is characterized by the ability of exact localization and timing of the communication containing one or more people (Christensen & Christensen, 2016, p. 252).

Most words do not just have a single meaning but several (sometimes connected) meanings, which brings us to the two central definitions within semantics of *homonymy* and *polysemy*. When a word has several connected meanings, it is called a polysemy and an example of such a word in Danish is the word "simpel" (*simple*) which has the different but connected meanings such as "ukompliceret" (*uncomplicated*) and "*jævn*" (even) (Christensen & Christensen, 2016, p. 265). When a word has two (or more) unconnected meanings it is called a homonymy and such an example is the danish word "lade" which has several meanings depending on the context where two of them is either: "*sb. bygning, der bruges som opbevaringssted for korn og hø / sb. Building used as storage for grain and hay*" (Christensen & Christensen, 2016, p. 265) thus it relates to the English word "*barn*" and also as: "*vb. f. eks. lade et gevær / vb. for example to load a gun*" (Christensen & Christensen, 2016, p. 265) thus it relates to the English word "*context (to) load*".

# ACADEMIC LANGUAGE, MATHEMATICAL DISCOURSE AND THE MATHEMATICAL REGISTER(S)

In order to properly discuss the use of language of students in the context of probability and statistics, the notion of a register, academic language, mathematical discourse and the mathematical register will now be defined. Firstly, a *register* of a language is defined as "... *the constellation of lexical and grammatical features that characterizes particular uses of language*" (Schleppegrell, 2001, p. 431).

According to Bitterlich & Shütte (2018) academic language is "... seen as a language register that is used in the context of school and education in order to impart knowledge, and that orients itself by written language with its higher degree of complexity and explicitness" (Bitterlich & Shütte, 2018, p. 35). In addition, Snow and Ucelli (2009) have identified five central aspects of academic language: Interpersonal stance, information load, organization of information, lexical choices, and representational congruence (Snow & Ucelli, 2009, p. 119). The interpersonal stance in academic texts is authoritative and detached through grammatical and lexical choices, and the information load is dense and concise. The organization of information is a logical unfolding of ideas, with lexical choices having a high lexical diversity and formal expressions with the use of abstract and technical terms. Finally, the representational congruence relates to the fact that some words are often used in a way which is different from their normal usage.

The use of language expected in a mathematics classroom is related to the concept a *mathematical classroom discourse*, which is the verbal and written communication which takes places in a mathematical classroom (Winsløw, 2006, p. 162). Moschkovich (2003) describes that some characteristic of the mathematics classroom discourse which include "... *being precise and explicit, searching for certainty, abstracting, and generalizing,[... and] imagining*" (Moschkovich, 2003, p. 327). The demand of precision and being explicit thus closely relate to the demand of information load and lexical choices in academic language. The students are expected to communicate in a certain and appropriate way during lessons, which for example include that they have to "... *explain and justify mathematical solutions or to answer the teachers' questions in a correct manner*" (Bitterlich & Shütte, 2018, p. 35).

In all communication it is necessary to follow the syntactical rules in order to make other people understand what is being communicated. Within mathematics it is impossible to separate the mathematical content from the mathematical representations and both are a part of the language used within mathematics (Winsløw, 2006, p. 159). Thus, when discussing the formal language of mathematics, it is not a language per say but rather a certain use of the language including the correct use of representations (Winsløw, 2006, p. 159). An important part of learning mathematics in general, is learning the formal language of mathematics as this is a way of accessing both new texts and a new way of communicating with others (Winsløw, 2006, p. 160). The *mathematical register* will therefore be defined as the formal language of mathematics expected in a mathematical classroom discourse.

#### TRANSITION BETWEEN NATURAL AND ACADEMIC LANGUAGE

When learning to properly use the mathematical register, the students are using their *everyday language* which is meant as the language (including syntax, semantics, and morphology) used outside the mathematics classroom. The acquisition of the use of the mathematical register can in some ways be compared to learning a new language: When learning a new language, it is based on or connected to an already known language. New words can be explained by the use of old words and little by little it becomes possible to express the same meaning through the new language - perhaps more clearly than the old language allowed. A main goal is to teach the students the mathematical content but using a mathematical register to express the knowledge should also be a part of teaching (Winsløw, 2016, p. 160). In validation phases, the importance relies mainly on the meaning and content (hence the semantics) but particularly in formulation phases the shape of the statement also becomes important. It becomes a matter of syntax in the sense that students are expected to compose and combine different representations in a way which is formally correct (Winsløw, 2016, p. 160).

The misunderstandings made by students in the presented mathematics is often a combination of syntax and semantics and it can be difficult to separate the semantical misunderstandings from the syntactical ones (Winsløw, 2016, p. 160). Winsløw (2016, p. 160) points out that a deliberate work on formulation where the students are guided to express themselves clearly and concisely in the mathematical register can aide the students' participation in validation situations to help the students' conception of the mathematical content. In addition, it has been shown that "...*regarding the discourse practices explanation and argumentation, that many learners do not gather enough language-based experience in family and peer groups prior to entering school. But often, such language competences in explaining and arguing are assumed*" (Bitterlich & Schütte, 2018, p. 36). The focus on the linguistic competences in the formulation phases is therefore supported.

The use of language within mathematics in relation to implications and logical structures are often implicit or informal but at the same time fundamentally different from the everyday use of language (Winsløw, 2016, p. 161). In standardized isomorphic problems i.e., problems where the students have a clear previously practiced techniques, this might not cause trouble for the students. If the context or formulation of the problem is slightly changed, however, it will become evident that only a small amount of semantics is behind these syntactical skills (Winsløw, 2016, p. 161). Winsløw (2016, p. 161-162) points out that evolvement of the students' sense of practical and theoretical range presumes that they are given the opportunity to rationally relate to them. Once more it relies on the quality of

the action- and formulation phases such that the validation and institutionalization is not simply reduced to an authoritative practice by the teacher (Winsløw, 2016, p. 162). Finally, the quality depends very much on the student's relationship with the mathematical register as a way of communication (Winsløw, 2016, p. 162).

#### LEARNERS WHOSE FIRST LANGUAGE IS NOT THE LANGUAGE OF INSTRUCTION

#### DEFINITION OF THE TERM "BILINGUAL" AND RELATED PROBLEMS

As this thesis focuses on learners whose first language is not the language of instruction it is relevant to discuss the use of terminology in relation to students who are *bi*- or *multilingual*. Both as a phenomena and concept "bilingual" has been subjected to a lot of research and discussion throughout the years (Holm & Laursen, 2019, p. 23) and generally in the older research, the discussion has been affected by the norm of monolingualism and the general conception of bilingualism has been deviating and problematic (Holm & Laursen, 2019, p. 23). There have been examples of older definitions of bilingualism which requires a complete (and perfect) mastery of two entire languages, such as the definition made by Cramer, Henriksen, Kunøe, Larsen, Togeby and Widell (1996), shown translated below:

"Bilingualism is used to describe a complete mastery of two languages. Bilingualism can be obtained by growing up in a bilingual environment where most commonly one language is dominant, or it can be achieved through moving to another linguistic area or (more rarely) by teaching"

- Cramer, Henriksen, Kunøe, Larsen, Togeby and Widell (1996, p. 162)

This definition is problematic for several reasons as it raises questions like how one would evaluate or measure the complete mastery of language and also: What is an <u>entire</u> language? Such definitions depend on a norm of double monolingualism (Holm & Laursen, 2019, p. 24) and from these types of definitions it would, for example, be expected that bilinguals do not interchange language during a conversation. As a contrast to this norm of double monolingualism, another (translated) suggestion of the term bilingual has been:

"People, who master two languages and uses in their use of language their entire linguistic competence adapted to the specific purpose of the use of language and the possibilities of the concerned situation including the linguistic competence of the partner of the conversation"

- Jørgensen (1998, p. 142)

This definition of bilingual thus have no expectation of the mastery of one language on the same level as a monolingual and that changes between the two languages is an integrated part of being bilingual (Holm & Laursen, 2019, p. 25).

Bilingual is thus not an unambiguously defined term and Holm and Laursen (2019, p. 25) point out four conditions which can be used to define the term bilingual. The first one is a competence criterion, which has demands of a high and in reality, unobtainable mastery of the two languages. The second one is a functional definition, where the requirements is merely the use of two languages and not a complete mastery. The third is an attitude criterion where the person would characterize herself/himself as bilingual and is determined from the perception made by the individual. The fourth one is a native criterion where there is a difference between simultaneous and successive bilingualism: simultaneous is considered a parallel mastery of two languages and successive bilingualism is when one language is mastered after another. In this thesis I will be using the fourth criteria when using the term bilingual. In a similar way I am defining and using the term multilingual.

#### THE USE OF MORE THAN ONE LANGUAGE WHEN LEARNING MATHEMATICS

When discussing the inclusion of bilingual students in elementary school in Denmark, the main focus has been on the meeting of different cultures and the teaching of the actual content of mathematics has been less prioritized (Andersen, 2004, p. 283). Most often the teaching of language and mathematics is separated, which means that the students do not obtain the linguistic competences, which are a prerequisite to learn mathematics. As pointed out by Høines (1997) most often it is not the mathematics, which causes trouble, but the language and the communication. Consequently, it becomes problematic to make this distinction between mathematics and language. Michael Wahl Andersen (2004) sums up that:

"The teaching of mathematics is very dependent of the learner's language comprehension, but it often happen that language and communication is separated in the teaching of mathematics. Therefore the learners do not get the opportunities to develop the language and communication skills, which is a necessity for successful learning of mathematics."

- Andersen (2004, p. 283)

It has been argued by Dale and Cuevas (1987) that language is in particular important in relation to open ended word problems. It seems as if there is a high correlation between the reading skills of bilingual students and their advancement in mathematics (Dale & Cuevas, 1987, p. 24) and this

applies in particular to open ended word problems. Thus, the teaching of mathematics should not be separated from the teaching of mathematical language, and the students should be given the opportunity to practice talking and/or writing about mathematics (Andersen, 2004, p. 286).

As reviewed in the introduction there has been an increase in the number of bilingual students in Danish High Schools and another report from the Danish Evaluation Institute (Danmarks Evalueringsinstitut, 2006) shows that the bilingual students in Danish High Schools generally get lower grades than ethnically Danish students (Danmarks Evalueringsinstitut, 2006, p 21). It is argued that ethnically Danish students get an average of 8.1 in their grades while bilingual students who are either born in Denmark or has moved to Denmark before the age of 6 averagely get a 7.8 in their grades (Danmarks Evalueringsinstitut, 2006, p. 21). The same report also points out that it is essential that the teachers develop competences such that they are capable of dealing with bilingual students. This development of competences includes for example different didactical aspects of the Danish linguistic aspects of the subjects being taught. Only few of the teachers have the insight needed to become aware of and meet the academic and social needs of many bilingual students. (Danmarks Evaluering-sinstitut, 2006, p. 10)

In the case of undergraduate students of mathematics in Denmark, there has been a tendency to offer all related texts in English and some of the verbal teaching occurring in English as well (Durand-Gurrier et al, 2016, p. 91), and this approach offers some challenges for the bilingual students as well. In pure mathematics it is estimated that around 10% of the students are second-generation immigrants and this group of students seem to be particularly challenged with English even though they are fluent in Danish (Durand-Gurrier et al, 2016, p. 91). The research therefore shows some challenges for the bilingual students throughout the Danish School System.

#### STRATEGIES TO HELP BILINGUAL STUDENTS IN MATHEMATICS TEACHING

Andersen (2004, p. 286) presents that bilingual students are in particular having difficulties when engaging in classroom discussions or collaboration with other students where it is a perquisition that they are to interpret what others are saying. A lot of the bilingual students are having difficulties acquiring the information and a consequence of this is that the bilingual students often are having difficulties in solving word problems in mathematics (Andersen, 2004, p. 286). In fact, it has been found that the bi- and monolingual students were doing equally well when solving mathematical problems with straight forward calculations where the problem could be solved without any interpretation of the provided text of the problem, but there were significant differences between the bi- and

monolingual students when the given problem had to be interpreted before calculating (Andersen, 2004, p. 286). In order to choose the correct mathematical operations, the students have to be able to master a specific vocabulary and they have to be able to identify special relations between certain words, and it is natural to believe that special training in reading would be a solution to the problem. However, when it comes to mathematics it would seem that this is not enough (Andersen, 2004, p. 286).

Instead, Andersen (2004) points out that the teacher should focus on clarifying if the bilingual student is experiencing linguistic, mathematical or cultural difficulties and this rely highly on the teacher having an understanding of Danish as a second language and intercultural competences (Andersen, 2004, p. 289). There are no tests to clarify the type of difficulties, and it therefore rely on the teacher to judge which type of problem the bilingual student is experiencing through dialog and in order to do this, the teacher needs to plan the lessons such that conversation in the lessons between students and conversations between student and teacher is prioritized

The communication between teacher and student is therefore an essential part of an effective mathematical teaching and in a classroom with a lot of bilingual students the communication can be thought of as even more fragile than in a classroom with a lot of monolingual students. When a new subject is introduced, Andersen (2004, p. 289) also recommends that the teacher makes up a list of the words that may be new to some or all of the students and as the words appear during the lessons, they should be explained by students who have understood them. This explanation can be given either through words or through illustrations, and another strategy could be to ask the students to provide synonyms or sentences to explain a word (Andersen, 2004, p. 289).

# **RESEARCH ON LEARNING STATISTICS**

When learning statistics, it is required to learn the associated vocabulary and the technical terms in order to communicate the complex concepts of statistics. Simply learning the new associated words is not enough, and it is necessary to use the language of the discipline in order to master the topic (Dunn et al, 2016, p. 9). The language of statistics should be used through both listening, reading, writing, and speaking, but often the focus of the teaching within statistics is mostly on quantitative aspects such as arithmetic computations (Dunn et al, 206, p. 9). Undoubtably the quantitative aspects are important, but "… students may not completely understand the concepts in these problems at a deeper conceptual level if they do not understand the language surrounding them" (Dunn et al, 2016,

p.9). Some of the challenges, which students face if they do not fully grasp the concepts and the use of the related vocabulary, is pointed out by Dunn et al (2016, p. 9-10) and include: The content, seeking assistance, and group work. In particular the students may not understand written content or even content presented in an oral form and in addition they may not be able to articulate that they need help. Ultimately this might mean that in group work the students are not able to converse with other students because of the lack of vocabulary (Dunn et al, 2016, p. 10). Some of the challenges of learning statistics is thus linguistic and there is therefore a tendency of viewing the learning of statistics as learning a new language.

### LEXICALLY AMBIGUOUS WORDS WITHIN STATISTICS

When learning how to operate within the mathematical register used in relation to statistics students may experience that some of the words, they have used in their everyday language, now obtain a new meaning. The words used in statistics can be divided into six categories (Dunn et al, 2016) and they include:

"(1) words with the same meaning in mathematical English (ME) and general English (GE; Rangecroft uses the phrase "Ordinary English" but we prefer the term used in linguistics); (2) words which have meaning only in ME; (3) words with a meaning only in statistical English (SE);
(4) words with different meaning in GE and ME; (5) words with a different meaning in GE and SE; and (6) words with a different meaning in ME and SE."

- Dunn et al (2016, p. 10)

As this thesis is focused on Danish High Schools, I will use the terms mathematical Danish (MD), general Danish (GD) and statistical Danish (SD) instead of mathematical English (ME), general English (GE) and statistical English (SE) respectively. The mathematical register related to statistics is important to learn in order to communicate within statistics, but the learning of this could also pose some semantical challenges. A part of the challenge lies in the fact that when communicating statistics, one has to use words from what Dunn et al (2016, p. 11) refers to as "different sources or fields" and additionally a word might belong in more than one field, a notion similar to the previously introduced notion of a register. But the word might have different meanings, and this could pose a cognitive challenge for the student to overcome. Such words are what we call *lexically ambiguous* (Kaplan, Fisher & Rogness, 2010, p. 2). The lexically ambiguous words are hominies and anongst such words are for example: Significant, power, control, random, and confidence (Dunn et al, 2016, p. 11). It has

been shown in studies that students tend to struggle to learn "... the technical definitions of lexically ambiguous words and often retain the GE definitions of these words" (Dunn et al, 2016, p. 11). It has been found that students tend to face statistics with "... strongly-held, but incorrect, intuitions that are highly resistant to change" (Kaplan, Fisher & Rogness, 2010, p. 2). Combined with students attaching new knowledge to previous knowledge there is a possible interference in the conception particularly when the statistic terms are lexically ambiguous (Kaplan, Fisher & Rogness, 2010, p. 2).

Research on elementary school children has provided "... evidence that awareness of linguistic ambiguity is a late developing capacity which progresses through the school years" (Durkin & Shire, 1991, p. 48) and studies on the development of the ability to detect linguistic ambiguity has been conducted and found a steady improvement across students in grade 1, 4, 7 and 10 (Kaplan, Fisher & Rogness 2010, p. 2). Hence, if high school students are made aware of the ambiguities, they "...should be able to correctly process the statistics meaning of the ambiguous words" (Kaplan, Fisher & Rogness, 2010, p. 2). However, Kaplan, Fisher and Rogness (2010) also points out that raising this awareness to overcome the effects of lexical ambiguity, is not a trivial task.

Kaplan, Fisher and Rogness (2010) conducted a study with an aim to clarify the statistical meanings most commonly developed and expressed by students at the end of an undergraduate statistics course in relation to the word's association, average, confidence, random and spread and the study was in particular interesting when it came to the words association and spread. They found that in relation to the word association "... many of the students who have developed a relational understanding of association may not have progressed further than describing a numerical relationship (21% pilot; 16% validation). Finally, only 19% of the pilot study students and 25% of the validation sample were able to express the definition of association explicitly as a relationship between two variables." (Kaplan, Fisher & Rogness, 2010, p. 16). In relation to the word spread Kaplan, Fisher & Rogness (2010, p. 16) conclude that it is indicated there is an equal number of students who uses the term as a synonym to "shape", as in how the data look on a graph or where the data are (14% pilot; 35% validation) and as a synonym for "variability" (15% pilot; 42% validation).

The focus of this thesis is on Danish High Schools and it is therefore relevant to mention some possible lexically ambiguous words in Danish. Some of the words previously defined as lexically ambiguous words in English also qualifies as lexically ambiguous in Danish and such words include: "tilfældig" (random), "uafhængig" (independent), and "spredning" (spread). In addition words such as "frekvens" (relative frequency) and "hyppighed" (frequency) also qualify as lexically ambiguous. The bilingual students performance in relation to the words spread, frequency, relative frequency and independent will all be investigated in the case studies in the section "Statistics".

## ADVANTAGE FOR LEARNERS WHOSE FIRST LANGUAGE IS NOT THE LANGUAGE OF INSTRUCTION?

Dunn et al (2016, p. 10) are comparing the learning of statistics to learning a new language and also emphasize that international students who speaks English as an additional language (EALS) do not learn a new vocabulary in the same way as native English speakers. Dunn et al (2016) point out that students who speak English as an additional language tend to "... *rely more on knowledge of the morphological form of new words to associate them with meaning, rather than on a pre-established network of word-meaning associations*" (Dunn et al, 2016, p.10) and also indicate that it is more likely that these students will more successfully adapt the new language of statistics. The argument is that EAL students will consult a dictionary of ask for guidance when facing a concept within statistic in contrast to native English speakers who will assume that "... *their guess of the meaning from context is correct.*" (Dunn et al, 2016, p. 10).

Barton et al (2005) on the other hand, has conducted a study showing that EAL students do not have an advantage in having the instructional language as a second language when studying mathematics in general. The study was conducted in New Zealand at the University of Auckland and showed that the disadvantages of EAL students are greater at third-year students than at first-year students (Barton et al, 2005, p. 726). International students and new immigrant are partly having more success as first year students as "... *their prior knowledge in this subject generally compares favorably with that of L1 English students*" (Barton et al, 2005, p. 727) and in addition the early stages of mathematics can be mastered successfully with limited English because of the repetitive role played by language (Barton et al, 2005, p. 727). The language requirements of third-year students within mathematics are, however, greater and new compared with those in first year and this becomes a disadvantage for EALs.

# RESEARCH ON LEARNING PROBABILITY

The need for learning probability has been recognized in many countries and probability has therefore been included in the curricula. The arguments for learning probability include an emphasis on "... *the contribution of proper judgmental processes and probabilistic reasoning to people's ability to make effective decisions*" (Gal, 2002, p. 1) and it has also been shown that "... *training in statistics can aid in solving certain types of everyday problems*" (Gal, 2002, p. 1). But what is "probability"?

According to Batanero et al (2016, p. 2) "*The theory of probability is, in essence, a formal encapsulation of intuitive views of chance that lead to the fundamental idea of assigning numbers to uncertain events*". Different views on probability exist, and they include: The classical, the frequentist, the propensity, the logical, the subjective, and the axiomatic. The different views on probability can be found in Batanero et al (2016, p. 3-7) in a more elaborated manner. but a brief overview is given in figure 2 below:

Views of probability	Procedures	Properties	Some related concepts
Classical	<ul> <li>Combinatorics</li> <li>Proportions</li> <li>A priori analysis of the experiment struc- ture</li> </ul>	<ul> <li>Proportion of favorable to all possible cases</li> <li>Equiprobability of elementary events</li> </ul>	<ul><li>Expectation</li><li>Fairness</li></ul>
Frequentist	<ul> <li>A posteriori collection of statistical data</li> <li>Statistical analysis of data</li> <li>Curve fitting</li> </ul>	<ul> <li>"Limit" of relative frequencies in the long run</li> <li>Objective; based on empirical facts</li> <li>Repeatable experiment</li> </ul>	<ul> <li>Relative frequency</li> <li>Data distribution</li> <li>Convergence</li> <li>Independence of trials</li> </ul>
Propensity	• A priori analysis of the experimental set up	<ul> <li>Physical disposition or tendency</li> <li>Applicable to single cases</li> <li>Related to the experimental conditions</li> </ul>	<ul> <li>Propensity</li> <li>Probabilistic causal ten- dency</li> </ul>
Logical	<ul> <li>A priori analysis of the space of possibili- ties</li> <li>Propositional logic</li> <li>Inductive logic</li> </ul>	<ul> <li>Objective degree of belief</li> <li>Revisable with experience</li> <li>Relationships between two statements, generalises implication</li> </ul>	<ul><li>Evidence</li><li>Hypothesis</li><li>Degree of implication</li></ul>
Subjective	<ul> <li>Bayes' theorem</li> <li>Conditional probabil- ity</li> </ul>	<ul><li>Subjective character</li><li>Revisable with experience</li></ul>	<ul> <li>Likelihood</li> <li>Exchangeability</li> <li>A priori probability (or distribution)</li> <li>A posteriori probability (or distribution)</li> </ul>
Axiomatic	<ul><li>Set theory</li><li>Set algebra</li></ul>	Measurable function	<ul><li>Sample space</li><li>Certain event</li><li>Algebra of events</li><li>Measure</li></ul>

#### FIGURE 2: OVERVIEW OF DIFFERENT MEANINGS OF PROBABILITY (BATANERO & DIAZ, 2007, P. 117)

With an increasing interest in statistics and the development of technology, it is the frequentist approach which has been preferred in curriculums. In the *frequentist meaning*, probability is defined as "...*the hypothetical number towards which the relative frequency tends when a random experiment is repeated infinitely many times*." (Batanero et al, 2016, p. 4). This definition allows the range of applications to expand compared to the classical view, as the empirical tendency is visible in a lot of natural phenomena (Batanero et al, 2016, p. 4). However, this definition has been criticized, due to the fact that only an estimate of probabilities, which could variate from one sample to another, is obtained. Additionally, when it is not possible to make a repetition of experiments under the exact same conditions this approach would also be inappropriate (Batanero et al, 2016, p. 4).

This view of probability allows an experimental approach and as pointed out by Batanero et al:

"An experimental introduction of probability as a limit of relative frequencies is suggested in many curricula and standards documents (e.g., the Common Core State Standards in Mathematics [CCSSI] 2010; the Ministerio de Educación, Cultura y Deporte [MECD] 2014; and the National Council of Teachers of Mathematics [NCTM] 2000), and probability is presented as a theoretical tool used to approach problems that arise from statistical experiences."

- Batanero et al (2016, p. 8)

In Denmark probability has recently been added to the curriculum and in the 2019 summer exams at STX and HF it was considered as a new subject. In the finals of mathematics at B level of STX it was concluded in the evaluation rapport from the Ministry of Education that the students had obtained the lowest score in questions related to probability (Undervisningsministeriet, 2019, p. 26). The students who had mathematics at A level of STX also struggled with problems related to probability, and the rapport concluded that probability was very difficult for a lot of the students. In particular, the students had struggled with a problem related to the normal distribution which had been unintentionally difficult (Undervisningsministeriet, 2019, p. 19). The same tendency was shown in the finals of mathematics at B level of HF.

It could indicate that the students generally struggle with calculations within probability, however, the rapport also notes that a part of the explanation to the relatively low scores could be found in the lack of proper teaching materials. It is pointed out that the reform implementation process was relatively short, which meant that the material developers had difficulties in keeping up causing a lack of access to textbook materials, formula collections, task collections, etc. from the beginning of the course (Undervisningsministeriet, 2019, p. 5).

#### WHY IS THE TEACHING OF PROBABILITY IS DIFFICULT FOR MATHEMATICS TEACHERS?

Probability has specific characteristics such as "... multifaceted view of probability or the lack of reversibility of random experiments" (Batanero et al, 2016, p. 1) which are not found in other areas of mathematics and this might create specific challenges both for students and teachers. Some of the challenges teachers face when teaching probability is the fact that "... counterintuitive results in probability are found even at very elementary levels, whereas in other branches of mathematics counterintuitive results are encountered only when working at a high degree of abstraction." (Batanero, Godino & Roa, 2004, p. 2) and in addition "... probabilistic reasoning is different from logical or causal reasoning..." (Batanero & Sanchez, 2013, p. 260). This may be exemplified in an everyday situation such as a coin toss: Having obtained four consecutive heads, does not affect the probability of the following coin toss. This may seem counterintuitive to many students and it represents one of the cognitive challenges which has to be overcome.

The nature of the conceptions and misconceptions in relation to probability by the students has been described by Batanero and Sanchéz (2013) and the heuristic students use to solve the related problems is divided into four groups. The first is representativeness where students tend to solve probability problems by the use of judgmental heuristics. Students will use "... representativeness heuristic tended to estimate the likelihood for an event based on how well it represents some aspects of the parent population." (Batanero & Sanchéz, 2013, p. 265) thereby the students will believe that small samples ought to reflect the population distribution. It would be natural to believe that students who have had more mathematical training would not apply the representativeness however, it has been "... widely reported among university students even after formal instruction in probability" (Batanero & Sanchez, 2013, p. 265). Another tendency is to believe random events are "equiprobable" by their very nature and this is what is described as equiprobability bias. That is, "... people exhibiting this bias judge outcomes to be equally likely when their probabilities are not equal" (Batanero & Sanchez, 2013, p. 265). A third heuristic has been identified as the "outcome approach" which has been applied in the interpretation of frequentist probabilities. In particular, when students are confronted with an uncertain situation, the goal is not specifying probabilities that reflect "... the distribution of occurrences in a series of events, but as predicting the result of a single trial" (Batanero & Sanchez, 2013, p. 267). The students will evaluate the probabilities in terms of how close they appear to the values 0%, 50% or 100% and it is only when the probability is very close to 50% will they consider an event as "random" (Batanero & Sanchez, 2013, p. 267). Finally, a misconception may arise due to a lack of identification of similar mathematical structure in different probabilistic problems.

Another problematic aspect of teaching probability is in the informal ideas assigned to chance and probability by children and adolescents before instruction which may affect subsequent learning (Batanero, 2009, p. 3). In particular, there is evidence that students will maintain probability misconceptions that are hard to eradicate based only on a formal teaching of the topic (Jones, Langrall & Mooney, 2007). As an example of this it has been found that children do not see dices or marbles in urns as random generator with constant properties. They therefore "… consider a random generator has a mind of its own or may be controlled by outside forces" (Batanero, 2009, p. 3).

#### COMBINATORIAL REASONING AND PROBLEM SOLVING

Combinatorics serve as an essential part of probability particularly in some calculations of probabilities. It is possible to classify three models of combinatorial configurations: Selections, distributions, and partition. In the selection model "... *a set of m (usually distinct) objects are considered, from which a sample of n elements must be drawn, according to some specification.*" (Batanero & Sanchez, 2013, p. 261) and usually it is easy for the students to model a counting method for this type of problem. An example of such a problem is found in Batanero and Sanchez, (2013, p. 261) as problem 1:

Problem 1: There are four numbered marbles in a box (with the digits 2, 4, 7, 9). We choose a marble and note down its number. Then we put the marble back into the box. We repeat the process until we form a three-digit number. How many different three-digit numbers is it possible to obtain? For example, we may obtain the number 222.

In the distribution model the problem refers to the distribution of a set of n objects into m cells and it can be exemplified in the following problem from Batanero and Sanchez (2013, p. 261):

Problem 2: Suppose we have three identical letters, we want to place them into four different envelopes: yellow, blue, red and green. It is only possible to introduce one letter into each different envelope. How many ways can the three identical letters be placed into the four different envelopes? For example, we could introduce a letter into the yellow envelope, another into the blue envelope and the last one into the green envelope.

An example of the partition model could be to split a set of n objects into m subsets, as in the following problem from Batanero and Sanchez (2013, p. 261):

Problem 3: "Mary and Cindy have four stamps numbered from 1 to 4. They decide to share out the stamps, two for each of them. In how many ways can they share out the stamps? For example, Mary could keep the stamps numbered 1 and 2 and Cindy the stamps numbered 3 and 4."

These three problems were given to 700 students (age 14-15) by Batanero et al. (1997) both with and without instructions in combinatorics and they found that "... selection, distribution and partition problems were not equivalent in difficulty for the students, even after being taught combinatorics." (Batanero & Sanchez, 2013, p. 262). Problems 2 and 3 were in particular difficult for the students. Some of the types of errors included:

- 1) Error of order
- 2) Error of repetition
- 3) Confusing the type of object
- 4) Excluding some elements to form the configurations
- 5) Nonsystematic listing
- 6) Not remembering the correct formula for a combinatorial operation that has been correctly identified.
- 7) Not remembering the meaning of the values of parameters in the combinatorial formula
- 8) Faulty interpretation of the tree diagram

Combinatorics is a difficult subject and researchers has tried to identify the elements leading to difficulties for students and also highlighting some of the common errors made by students and this include for example " ... over-counting and confusing situations in which order matters or does not matter" (Lockwood, Swinyard & Caughmann, 2015, p. 29). Another difficult aspect of combinatorics is the fact that students may obtain large numerical answers which can be difficult to verify (Lockwood, Swinyard & Caughmann, 2015, p. 29).

# LANGUAGE AND LEXICAL AMBIGUITY IN THE PROBABILITY REGISTER

The notion of lexically ambiguous words was previously introduced in the section on research on statistics, but lexically ambiguous words are not exclusively found in the context of statistics. As pointed out by Molnar (2018, p. 28) a term such as random also has lexical ambiguity as it changes meaning among "... everyday, probability, mathematics, and statistics registers" (Molnar, 2018, p.

28). In the probability register a definition of a random phenomenon is "...*if individual outcomes are uncertain, but there is nonetheless a regular distribution of outcomes in a large number of repetitions*" (Moore et al. 2013, p. 261). In everyday language "random" instead describes a process without method or conscious choice and in everyday language "... *the process that generates outcomes is not known, unlike the common usage in probability and statistics.*" (Molnar, 2018, p. 28). In the statistical register, a random sample is for example defined as consisting of "...*n individuals chosen in such a way that every set of n individuals has an equal chance to be the sample actually selected*" (Moore et al. 2013, p. 204). This definition of a random sample therefore satisfies the general definition in the probability register, "... *but it restricts the known process by mandating that each observation has equal probability.*" (Molnar, 2018, p. 29).

Another example of a lexically ambiguous word within probability and statistics is the term independent. Misconceptions arise in relation to this notion and one potential reason for this trouble is the lexical ambiguity as independence has different meanings in probability, everyday language, algebra and statistics. In probability two events A and B are said to be independent if  $P(A \cap B) = P(A)$ . P(B) i.e. if the occurrence of one event does not change the probability of the other. Within probability the concept is therefore symmetrical. In everyday language independent relates to something "... not depending on the authority of another, not in a position of subordination or subjection; not subject to external control or rule; self-governing, autonomous, free" (Molnar, 2018, p. 31) and in everyday language, independent and dependent are not necessarily symmetric. Within algebra, the notion of independent and dependent are also used to describe variables. The dependent variable is thus an a quantity expressed by another variable: the independent one. This definition of variables is never symmetric "... although closer to the sometimes-symmetric everyday definition than the always-symmetric probability version" (Molnar, 2018, p. 31). Finally, in statistics independence is applied in multiple situations, where for example "... independent variables in regression are predictors that control the value of the dependent response variable" (Molnar, 2018, p. 31) and this definition is "... never symmetric, similar to mathematical variables but opposite of probability events" (Molnar, 2018, p. 31). Lexically ambiguous words therefore also pose a challenge within the probability register as well as some words used in the vocabulary belong to several fields and the meaning changes in relation to the fields.

How to improve students understanding of probability: Real data and teacher education

As the technology has improved and Computer Algebra Systems (CAS-tools) have become easily accessible it has been suggested that the "... *increasing use of communication and information technologies is rapidly changing the teaching of probability*" (Batanero & Sanchez, 2013, p. 277). It has become easier to obtain and work with real data which could aide the students when learning probability as "... working with real data allows students to appreciate the difference between empirical *phenomena and probabilistic models; it shows them the usefulness of these models in explaining, predicting and controlling a variety of real phenomena beyond pure games of chance.* " (Batanero & Sanchez, 2013, p. 277)

In a report from the Conference Board of the Mathematical Sciences entitled "The Mathematical Preparation of Teachers" noted in relation to American teachers that:

"Of all the mathematical topics now appearing in middle grades curricula, teachers are least prepared to teach statistics and probability. Many prospective teachers have not encountered the fundamental ideas of modern statistics in their own K-12 mathematics courses...Even those who have had a statistics course probably have not seen material appropriate for inclusion in middle grades curricula."

- Conference Board of the Mathematical Sciences (2001, p. 114)

Several reports also note that "... many of the current programmes do not yet train teachers adequately for their task to teach statistics and probability." (Batanero 2009, p. 2). This includes reports from Russell (1990), Batanero, Godino & Roa (2004) and Franklin & Mewborn (2006). As teachers serve as a link or mediator between the curriculum and the students, the proper training of teachers is essential. Two main aspects to focus on (according to Batanero(2009)) the epistemological reflection on the meaning of the concepts being taught within probability i.e. it is essential that the teachers reflect on the different meanings of probability. The second aspect is the students' various levels of understand and conceptual difficulties in problem solving i.e., the teachers should be "...organizing and implementing statistics projects, experiencing students' multiple forms of work and understanding experiments, simulations and graphical representations not just as methodological teaching aids, but rather as essential means of knowing and understanding." (Batanero, 2009, p. 7). Students build new knowledge in "... an active way, by solving problems and interacting with their classmates we should use this same approach in training the teachers especially if we want them later use a constructivist and social approach in their teaching" (Batanero, Godino & Roa, 2004, p. 3).

Probability is difficult to teach and it is suggested that "... we should not only present different models and show their applications, but we have to go deeper into wider questions, consisting of how to obtain knowledge from data, why a model is suitable, how to help students develop correct intuitions in this field and deal with controversial ideas, such as randomness or causality." (Batanero, Godino & Roa, 2004, p. 2). A problem is the fact that few teachers have prior experience with conducting probability experiments or simulations. It could thus propose a challenge to implement an experimental approach to teaching probability or teaching through statistical investigation if the teacher has no training in this area (Batanero, 2009, p. 7).

#### LEARNING PROBABILITY IN A NON-NATIVE LANGUAGE

As presented, the meaning of a word may vary in relation to the register, and a word may also vary in everyday language relating on the given language. It therefore becomes relevant to review literature concerning the learning of probability and learners whose first language is not the language of instruction. As an example, in southern Malawi most school children are offered their secondary mathematical instruction in English instead of the native language Chichewa, and Kazima (2006) has surveyed 154 students who had yet to receive formal instruction in probability. The study found, that by the complexity of the translation it became challenging for Chichewa speakers "... to differentiate between scale levels with modifying adjectives, such as equally likely versus not very likely" (Molnar, 2018, p. 34). Additionally, Kazima (2006, p. 169) found that a lot of the students" "... preconceived meanings for probability vocabulary were distant from established conventional meanings" (Kazima, 2006, p. 169) and some of these meanings were rooted in the students' first language. It was recommended by Kazima that the students were offered something more than teacher-spoken definitions, "instances which would help them refine their construction of meanings for the vocabulary" (Kazima, 2006, p. 187) and how words are appropriately used.

# ANALYSIS

In this part I present the justification for the methodologies chosen to investigate the bilingual students' performance in the context of probability and statistics teaching in Danish High Schools. The focus of the analysis will be on the difficulties related to language which (according to the research literature presented earlier) may arise for learners whose first language is not the language of instruction. The analysis is split in two sections: Statistics and probability.

In the first section (statistics), I present and analyze student answers to a questionnaire related to lexically ambiguous words in the context of statistics. The analysis focuses on semantical and syntactical difficulties and difficulties in the transition from natural to formal language. The data was collected at H.C. Ørsted Gymnasiet (HTX, Frederiksberg) where the class had just finished the introduction to statistics. In the second part (probability), I present and analyze observations of student conversations in relation to probability obtained from classroom observations made at Frederiksberg Gymnasium (STX, Frederiksberg), where the observations were made in the middle of the introduction to probability.

# **STATISTICS**

As presented in the theoretical framework, lexically ambiguous words generally pose a challenge to students in the context of statistics as the semantics may cause trouble and the investigation in the context of statistics is therefore in relation to lexically ambiguous words. In order to investigate which language related difficulties bilingual students experience, a questionnaire (see appendix 1A (1B for English)) was created. The questionnaire aimed to clarify how the students thought of and understood some of the basic but very important definitions within statistics: Spredning (spread), hyppighed (frequency), frekvens (relative frequency) and uafhængighed (independence). As the selected concepts all qualify as lexically ambiguous words, it will be analyzed which registers the bilingual students use when asked to define the selected concepts, and if the semantics and syntax is causing difficulties. In order to evaluate if the semantical and syntactical difficulties are due to the students being bilingual, the answers have been compared to the monolingual student's replies.

#### METHODOLOGY

The investigation was conducted at H.C. Ørsted Gymnasiet (Frederiksberg) at a second-year math class where the students had just finished an introduction to statistics. The students had been introduced to concepts within descriptive statistics such as the mean, maximum, minimum, variation,
median, quartiles, population and sample sets, variation, spread, frequency, and relative freequency. The students had also had an introduction on how to visualize data with programs such as Excel (making boxplots, histograms, and tables). The questionnaire (appendix 1A) was given to the students, who had just ended their course in descriptive statistics, and there were n=17 students present the day the questionnaire was handed out. This is a relatively low amount of respondents, as 200 respondents would be more appropriate as this was the amount used in the similar investigation made by Kaplan, Fisher and Rogness (2010). The discussion will further examine this fact. In order to provide a quantitative aspect more respondents would have been required and in that case, it would have been possible to categorize the types of answers in accordance with the categorizations made by Kaplan, Fisher and Rogness (2010).

As the focus of this thesis is on students whose first language is not the language of instruction, the questionnaire included that the students wrote which language they spoke at home. This was done in order to clarify which of the students were bi- or multilingual in accordance with the definition of bi- and multilingual discussed in the section of "Definition of the term "bilingual" and related problems" in the theoretical framework. A student therefore qualifies as bilingual or multilingual if they speak another language than Danish at home.

Of the 17 respondents, 6 replied Danish and some other language (Chinese, English, Urdu, Arabic, French and Hindi), 6 replied another language (Arabic, Turkish, Sinhala, and Iraqi), 2 replied Danish and 3 did not reply to this question and there was therefore a total of 12 bilingual students and 2 monolingual students who replied to the questionnaire. In the analysis, the bilingual students have been denoted  $BS_i$  for i=1..12 and the monolingual students have been denoted  $MS_i$  for i=1 and 2.

As the data was collected at only one Danish High School in a specific class, it is possible that other factors than bilingualism may have contributed to the differences in the answers of the questionnaire. This aspect of representativity of the obtained data, will be further examined in the discussion.

The focus of the questionnaire was in the context of lexically ambiguous words and the students were for example asked to use the word "uafhængig" (independent) in a sentence using its primary meaning in everyday language and afterwards give a definition that maintained the meaning from the created sentence. The students were then asked to write a sentence with the same word while using its primary meaning from High School mathematics and afterwards give a definition of the word as used in the sentence. The students were asked to do similar tasks with the words "spredning" (spread/standard deviation), "hyppighed" (frequency) and "frekvens" (relative frequency).

As previously introduced in the theoretical framework, independent is a lexically ambiguous words having several different meanings depending on the register and it is similar in the Danish translation "uafhængig". "Spredning" (spread) also qualify as a lexically ambiguous word and most students are likely to use it in their natural language or have heard it being used in natural language. In natural language "spredning" (spread) means according to the Danish dictionary that someone or something is expanding over a larger area (ordnet.dk, 2020). In the statistical register, spread is defined at the square root of the variance i.e.,  $\sigma = \sqrt{VAR}$ . In the statistical register, the spread is therefore an estimate of how far away from the mean, the values of the data set are.

In natural language, "hyppighed" (frequency) means the number of times something is repeated or happens in each of a larger number of equal lengths of time (ordnet.dk, 2020 (2)) and this meaning is very similar to the meaning in the mathematical register, where "hyppighed" (frequency) is the number of occurrences in relation to total number of units. However, in natural language "hyppig" (frequent) is a synonym to "ofte" (often) i.e., the meaning changes in the sense that it is no longer a numerical estimate. For this reason the word "hyppighed" (frequency) has also been included as lexically ambiguous as the conjugation of the word changes the meaning in natural language.

"Frekvens" (relative frequency) might not be used as much in natural language by students, but the concept has been chosen anyway as it has a meaning in natural language in relation to radio, in relation to a scientific language (within physics for example) and a meaning within statistics. In addition, "frekvens" (relative frequency) can be expressed by the use of the symbolic register as there is a formula expressing the relative frequency as a quotient of the frequency and the amount of observations in the data set, n, (Marthinus et al, 2020) i.e.:

$$relative frequency = \frac{frequency}{n}$$

The methodology of this investigation is similar to the study conducted by Kaplan, Fisher and Rogness (2010) and asking students to define mathematical concepts was inspired by Carmen Díaz and Carmen Batanero (2009) who made a study within probability asking, amongst other things, the students to define what conditional probability is and provide a correct example (Díaz & Batanero, 2009, p. 142).

## ANALYSIS OF STUDENT ANSWERS TO APPENDIX 1A

The methodology has now been introduced and in this section, I will analyze the replies to appendix 1A, which include a comparative analysis of the monolingual student's replies. Only two students replied that they only spoke danish at home and it provides a relatively small foundation of comparison, but it has still been included. This will be further discussed in the discussion of the results of the analysis.

## ANSWERS IN RELATION TO "SPREDNING" (SPREAD)

In the first section of the questionnaire the students were asked to use and define the word "spredning" (spread) as used in "everyday language" (thus in natural language) and to use and define the word using the meaning they knew from High Schools Mathematics. 11 bilingual students were able to correctly use the word "spredning"(spread) with some meaning of everyday language, where examples of correct usages of the word in everyday language include "spredningen af coronavirusset er blevet større / the spread of the corona virus has become larger" and "spredningen af mennesker i lokalet ser fornuftigt ud / the spread of people in the room looks reasonable".

As this investigation was made during September 2020 and thus during the Covid-19 pandemic, a lot of the bilingual students naturally related their everyday use of "spredning" (spread) to the spread of coronavirus or the spread of people. Thus, most bilingual students have heard or used the word "spredning" (spread) in their natural language and most of the bilingual students, who were able to use the word correctly, also defined it correctly. Nine of the bilingual students were able to use "spredning" (spread) correctly in a mathematical sense and only two students have provided an incorrect example of the usage of the word as used in High School mathematics. Some selected answers will now be presented and analyzed.

THE CASE OF BILINGUAL STUDENT 1: SEMANTICAL DIFFICULTIES

The answer provided by bilingual student 1, BS<sub>1</sub>, is shown in image 1 below:

Hvilket sprog taler I hjemme ved dine forældre? AsabiSK, DanSK 1. Spredning A. Skriv en sætning hvor ordet "spredning" indgår med samme betydning, som du bruger det i til daglig spredning at Corona er blevet for doblet B. Giv en definition af hvad ordet "spredning" betyder, når det bruges som du gjorde i svaret på A flvor mange" C. Skriv en sætning hvor ordet "spredning" indgår med den betydning du kender fra gymnasiets matematik " find spredningen at antal Smittede 1 Danmart D. Given definition af ordet "spredning" sådan som det bruges i svaret på C Spredning er at definere "af Starid" på noget besteme

IMAGE 1: ANSWERS FROM BS1 IN RELATION TO "SPREDNING" (SPREAD)

BS<sub>1</sub> replied: "spredning af coronavirus er blevet fordoblet / the spread of corona virus has doubled" when asked to use the word in everyday language and defines spread as: "Hvor mange / how many". This could indicate that the student understands "spredning" (spread) as a measure on how many people are infected. This is, however, not in accordance with the lexical definition of "spredning" (spread) in the Danish dictionary (Ordnet.dk, 2020) where the closest definition of the word is: "det at smitstof, biologiske eller kemiske partikler el. lign. spredes over et større område eller til flere personer / the fact that infectious substances, biological or chemical particles or the like of it, is spread over a larger area or to several people" (Ordnet.dk, 2020).

There could be several reasons why  $BS_1$  has made an incorrect reply such as: 1) The student is having difficulties finding the right words to describe "spredning"(spread) and therefore simply tries to explain what it meant in this exact situation (i.e. how many people has been infected?), 2) the student has heard the word "spredning"(spread) used in relation to how many people the virus has spread to thus making the misunderstood connection causing a semantical error to occur or 3) the student have misread the question and not noted that he/she has to provide a definition of the word and therefore answers incorrectly.

The second explanation of a semantical error of  $BS_1$  is further indicated in the following use of the word as used in High School Mathematics.  $BS_1$  used the word as: *"find spredningen af antal smittede* 

*i Danmark / find the spread of the amount of infected in Denmark*" which could reflect that the student understands "spredning" (spread) as a numerical estimate for the amount of something but this would not be in accordance with the following definition provided by student: "*spredning er at definere afstand på noget / spread is to define distance on something*". The usage of "spredning" (spread) in 1C is however in accordance with the definition of the word provided in 1B. This could indicate that BS<sub>1</sub> relies on the general danish conception of the word rather than the technical definition of the word provided in the mathematical register. This type of semantical problem was also argued by Dunn et al (2016) and reviewed in the theoretical framework.

#### THE CASE OF BILINGUAL STUDENT 2: SEMANTICAL AND SYNTACTICAL DIFFICULTIES

The second answer for analysis is from bilingual student 2, BS<sub>2</sub>, whose reply is presented in image 2 below:

Hvilket sprog taler I hjemme ved dine forældre? arabis / 1. Spredning A. Skriv en sætning hvor ordet "spredning" indgår med samme betydning, som du bruger det i til daglig his or thebre meget er sproching for halrdelen af befolloning har corone B. Giv en definition af hvad ordet "spredning" betyder, når det bruges som du gjorde i svaret Spred-hiby et hvor Stor sandsynlighed er et bestemt antal harler hogod C. Skriv en sætning hvor ordet "spredning" indgår med den betydning du kender fra gymnasiets matematik Sprechisen for 60% at statisticen er tas D. Giv en definition af ordet "spredning" sådan som det bruges i svaret på C Sprednibg et hror spredt i et dataset et bestemt mål Fi

IMAGE 2: ANSWERS FROM BS2 IN RELATION TO "SPREDNING" (SPREAD)

BS<sub>2</sub> uses the word in 1A as: "hvor meget er spredning for halvdelen af befolkningen har corona / how much is the spread that half of the population have corona" and afterwards defines the word "spredning" (spread) as used in everyday language as: "spredning er hvor stor sandssynlighed er et bestemt antal har/er noget / spread is how big a probability a certain amount has / is something". This answer could indicate that the student has a conception of the notion of "spredning" (spread) as the probability of something and hence connecting "spredning" (spread) to how likely something is to happen. Combined the answers to 1A and 1B support the claim that the student is connecting "spredning" (spread) to how likely something is to happen. This is not in accordance with the

definition provided by the danish dictionary (ordnet.dk, 2020) as presented earlier and the answer from  $BS_2$  would therefore qualify as a semantical error as the word is used semantically different from the lexical meaning of the word. A likely explanation of this could be that  $BS_2$  is not using the word in in his/her natural language.

What is interesting is that  $BS_2$  replies: "spredning er hvor spredt i et datasæt et bestemt mål / spread is how spread in a data set a certain measure" which syntactically is incorrect thus it becomes very difficult to extract a meaning of the answer. This syntactical error would qualify as a language related difficulty as  $BS_2$  does not clearly communicate the intended meaning. As reviewed in the theoretical framework this is in particular problematic in formulation phases as it becomes difficult for other students (and the teacher) to understand the intended. As this communication is provided in a written media  $BS_2$  is expected to be concise and precise in the formulation but the answer lacks this.

#### The case of bilingual student 3: Use of the mathematical register

A third answer in this relation came from bilingual student 3, BS<sub>3</sub>, and poses as a contrast to the first two student answers. It is displayed in image 3 below:

Hvilket sprog taler I hjemme ved dine forældre? Sinhala 1. Spredning A. Skriv en sætning hvor ordet "spredning" indgår med samme betydning, som du bruger det i til daglig Fx: Fondelingen of en populations højde, spreder sig med 5 cm Fra middelvardien. B. Giv en definition af hvad ordet "spredning" betyder, når det bruges som du gjorde i svaret Spredning betyder, hvor megent noget fondeler sig med fra middelverdien. Fx: I en hormalfordeling. C. Skriv en sætning hvor ordet "spredning" indgår med den betydning du kender fra gymnasiets matematik Fondelingen of mangale juice i ml, spredir sig med 4ml fra middelvardion. D. Giv en definition af ordet "spredning" sådan som det bruges i svaret på Ç Spredning er fordelingen at observationer fra Observationssetteds middel word. Spredning her symbolist sigma (o), og lan bestemmes val fra variansen. 2. Hyppigh

IMAGE 3: ANSWERS FROM BS3 IN RELATION TO "SPREDNING" (SPREAD)

In 1A the student replies: "Fordelingen af en populations højde, spreder sig med 5 cm fra middelværdien / the distribution of the height of a population is spreading 5 cm from the mean" and afterwards defines in 1B: "spredning betyder hvor meget noget fordeler sig med fra middelværdien. Fx: I en normalfordeling / spread means how much something is distributed with from the mean. Fx: In a normal distribution". BS<sub>3</sub> thus uses the word in everyday language similar to the way it is used within the mathematical register. This could indicate that  $BS_3$  does not use "spredning" (spread) in natural language and therefore rely exclusively on the mathematical conception of the word. This is particularly implied by the use of "*Fx:*. *i en normalfordeling* / *Fx: In a normal distribution*" as this is probably a formulation from a textbook. Another possible explanation can be found in the milieu and the mathematical classroom discourse. As the questionnaire was presented while the students were at school and with the teacher present in the classroom this could have affected the type of answer provided by BS<sub>3</sub>.

In 1C BS<sub>3</sub> replied: "fordelingen af mængde juice i ml spreder sig med 4 ml fra middelværdien / the distribution of the amount of juice in ml is spreadning 4 ml from the mean" and afterwards defines in 1D: "Spredning er fordelingen af observationer fra observationsættets middelværdi. Spredning har symbolet sigma ( $\sigma$ ), og kan bestemmes ud fra variansen / Spread is the distribution of observations from the mean of the set of observations. The spread has the symbol sigma ( $\sigma$ ), and can be determined from the variance". BS<sub>3</sub> formally defines "spredning" by using the related vocabulary of the register of mathematics by using terms as "middelværdi", "observationssæt", "fordeling", "varians" (mean, set of observations, distribution, variance). BS<sub>3</sub> also uses the symbolic register as by including the usage of  $\sigma$  to represent the spread.

This answer is formulated in a more academic language as it is concise, precise and uses technical terms to describe the concept of spread. This answer to 1D poses a contrast to the answer provided by  $BS_2$  where syntactical errors made it difficult to extract a meaning. The answer from  $BS_3$  would be more accepted in a mathematical classroom discourse as it relies on the mathematical register in the formulation and uses a well-chosen representation.

## FINAL COMMENTS ON ANSWERS

In question 1D, 7 of the bilingual students answered incorrectly and the students provided answers like: "Hvor godt observationerne er spredt / How well the observations are spread" from BS<sub>4</sub> and: "Det betyder at den spreder sig med 0,3453 / It means it spreads with 0,3453" from BS<sub>5</sub>. Some of the incorrect answers are listed as incorrect as they are not a definition: Instead the students rephrase their answers from 1C and they do not explain what "spredning"(spread) means mathematically, they simply use the word again. Another incorrect answer was from BS<sub>6</sub> who wrote: "hvor stor forskel der er fra mindste til største værdi / how big a difference there is from the smallest to the largest value". This student has interchanged the concept of "spredning" (spread) with "variationsbredde" (the width of the data).

The four correct answers to question 1D were informally correct but none of the students had chosen a formal, mathematical definition. BS<sub>7</sub> correctly replied: "*Hvor stor en fordeling/ændring/afvigelse der er på resultaterne ud fra middelværdien / how big a distribution/change/deviation there is on the results from the mean*" which shows that BS<sub>7</sub> is able to describe the spread as a deviation from the mean in an informal way where he/she even writes three words (*fordeling/ændring/afvigelse (distribution/change/deviation*)) to describe what he/she means.

#### The case of monolingual student $1 \mbox{ and } 2$

The reply of monolingual student number 1, MS<sub>1</sub>, is presented in image 4 below:

Hvilket sprog taler I hjemme ved dine forældre? DANSK 1. Spredning A. Skriv en sætning hvor ordet "spredning" indgår med samme betydning, som du bruger det i Spredningen dever nes skabtornelse JOVNE A B. Giv en definition af hvad ordet "spredning" betyder, når det bruges som du gjorde i svaret Valicton C. Skriv en sætning hvor ordet "spredning" indgår med den betydning du kender fra gymnasiets matematik at bedømme Spredning bruges forstallige observader er Mange tioner. D. Giv en definition af ordet "spredning" sådan som det bruges i svaret på C Lighed/V Gridtion

IMAGE 4: ANSWERS FROM MS1 IN RELATION TO "SPREDNING" (SPREAD)

Where  $MS_1$  in 1A wrote: "spredningen af elevernes skostørrelse er jævnt fordelt / the spread of the students' shoe sizes is evenly distributed" and defined the word in 1B as "variation /variation".  $MS_1$  uses the word with a semantical meaning close to the everyday meaning of "variation" (variation) i.e. as a relative description of differences when looking at a collection of items. When asked to use the word with a mathematical meaning  $MS_1$  replies: "spredning bruges til at bedømme om der er mange forskellige observationer / spread is used to assess if there are many different observations" and then defines the word as "Lighed /variation / Similarity/ variation". In a mathematical sense  $MS_1$  has a conception of "spredning" (spread) as a description of differences in the dataset and this conception is thus closely related to the meaning provided in 1A and 1B.  $MS_1$  defines "spredning" (spread) in 1D as both similarity and variation and it is possible that  $MS_1$  by this answer intend to define

"spredning" (spread) as a description of the variations and similarities of a dataset. However,  $MS_1$  simply writes the words and it is unclear if  $MS_1$  wants to define spread as a synonym to either similarities or variation: Two synonyms which are not in accordance with the mathematical meaning of "spredning" (spread). If  $MS_1$  sees "spredning" (spread) as a synonym to "variation" (variation) this would not be an uncommon usage: As reviewed in the theoretical framework, Kaplan, Fisher and Rogness (2010) found that 42% of the tested students used variation as a synonym to spread. Combined, these four answers would indicate that  $MS_1$  have the same conception of "spredning" (spread) mathematically as in everyday language.

The reply from MS<sub>2</sub> is presented in image 5 below:

Hvilket sprog taler I hjemme ved dine forældre? Dansk. 1. Spredning A. Skriv en sætning hvor ordet "spredning" indgår med samme betydning, som du bruger det i til daglig spedning lesse på dag ligda og spice fortoldets mellem pholdes mellen Given definition af hvad ordet "spredning" betyder, når det bruges som du gjorde i sværet B på A Spedning besser in copridd der angiven Doe the eb Seb Verdicits tordense of forthet placening if b. himaudeen C. Skriv en sætning hvor ordet "spredning" indgår med den betydning du kender fra gymnasiets matematik Data Selles bester of 10 vondi observationer. Benegen Spiedning - cu for decoracionene. D. Giv en definition af ordet "spredning" sådan som det bruges i svaret på C speckning bollower a copholdet mellen es ansal When man se har langs of wardiene tog Stin the himander, his 2. Hyppighed

IMAGE 5: ANSWERS FROM MS2 IN RELATION TO "SPREDNING" (SPREAD)

When asked to use the word while maintaining the meaning of everyday language MS<sub>2</sub> reply: "Spredning betyder på dagligdagssprog forholdet mellem en bestemt mængde værdier og deres værdier hvor de ligger ift. hinanden / In everyday language, spread means the relationship between a certain amount of values and their values where they lie in relation to each other" and afterwards defines it as: "spredning er et forhold der angiver et sæt værdiers tendenser og placering ift hinanden / spread is a ratio that indicates the tendencies and location of a set of values in relation to each other". In 1A MS<sub>2</sub> is already providing an explanation of what the word means and in 1B MS<sub>2</sub> is providing a definition of the word which is more in accordance with the mathematical meaning of the word if the mean had been mentioned as well. This could indicate that  $MS_2$  only rely on the mathematical register even when asked to use the word in everyday language. As in the case of  $BS_3$  an explanation could be found in the mathematical classroom discourse.

In 2C MS<sub>2</sub> replied: "Datasættet består af 10 observationer. Beregn spredningen for observationerne / The data set consists of 10 observations. Calculate the spread for the observations " and uses the word as seen in problems presented during the introduction to statistics provided through High School mathematics. MS<sub>2</sub> then defines "spredning" (spread) as: "spredning forklarer forholdet mellem et antal observationer i et datasæt og kan beregnes. Med spredning kan man se hvor langt at værdierne står fra hinanden hvilket siger noget om datasættet / The spread explains the relationship between a number of observations in a data set and can be calculated. With the spread, you can see how far the values are apart, which says something about the data set ". The definition in 1D is thus closely related to the definition provided in 1B and it is interesting that MS<sub>2</sub> have a conception of "spredning" (spread) as an estimate of how far the values or observations are apart as this formulation indicate a spatial conception of the word closely related to the everyday conception of "spredning" (spread) when used in relation to for example the spread of people in a room.

# COMPARISON OF BILINGUAL STUDENT REPLIES AND MONOLINGUAL STUDENT REPLIES

In the past analyses we have seen cases of semantical difficulties in particular in the case of  $BS_1$  and  $BS_2$ .  $BS_1$  had semantical difficulties in the sense that the general Danish conception of "spredning" (spread), as a numerical measure on how many people were infected with corona virus, was not in accordance with the lexical meaning of the word. In the mathematical usage of the word this conception was also present, and it would seem that  $BS_1$  rely on the (wrongful) general Danish conception of the word rather than the mathematical meaning even though  $BS_1$  defined the word with a different meaning in 1D.

The answers provided by  $MS_1$  also showed a similar type of semantical difficulties as the conception of "spredning" (spread) provided in an everyday meaning and mathematical meaning were also very closely related, and it was argued that  $MS_1$  also relied on the general Danish conception of the word rather than the mathematical meaning. The difference in the semantical issues of  $BS_1$  and  $MS_1$  is in particular that  $MS_1$  provided a semantical meaning in natural language which is in accordance with the lexical meaning of the word and  $BS_1$  did not. In the case of  $MS_2$  the provided meanings of "spredning" (spread) were also closely related, as  $MS_2$  used and defined the word in everyday language in a mathematical context and the similarities in the usage of the word can thus be explained by this. The case of BS<sub>2</sub> also showed a semantical conception of the word in natural language which was not in accordance with the lexical meaning of the word and in addition the syntactical errors in the 1D qualified as a language related problem as it became difficult to extract a meaning of the provided definition. MS<sub>2</sub> did not experience any syntactical errors which qualified as language related problems such that the indented meaning became difficult to extract: Only the answer provided in 1D by MS<sub>1</sub> became ambiguous as both similarities and variation was written perhaps as synonyms. The language related problems of BS<sub>2</sub> based on syntactical errors were only present with BS<sub>2</sub>: In fact, BS<sub>3</sub> provided a sentence and definition with a mathematical meaning in an academical language which used both the symbolical register and was concise with usage of technical terms in relation to statistics. Neither MS<sub>1</sub> nor MS<sub>2</sub> provided an answer formulated by the use of the register of mathematics. However, BS<sub>3</sub> did, similar to MS<sub>2</sub>, use the word in everyday language and meaning in a way which was more in accordance with the mathematical meaning of the word. As pointed out, this could be explained by the milieu and the mathematical classroom discourse as the questionnaire was given to the students during a normal lesson and this explanation is valid for both BS<sub>3</sub> and MS<sub>2</sub>.

#### FINDINGS IN RELATION TO "SPREDNING" (SPREAD)

In the data concerning the lexically ambiguous word "spredning" (spread) we saw semantical difficulties in the case of  $BS_1$ , who relied on the general danish conception of the word when providing a mathematical definition. The case of  $BS_2$  presented some semantical and syntactical difficulties as the student regarded "spredning" (spread) as the probability of something occurring and the formulation in the definition of 1D had syntactical errors which made it difficult to extract a meaning. The case of  $BS_3$  showed an answer formulated by the use of the mathematical register and in this case there were no semantical nor syntactical errors occurring. The case of  $BS_6$  also showed a semantical error as the student had interchanged the concept with the width of the data.

In the case of  $MS_1$ , "spredning" (spread) was mathematically defined as a synonym to either "lighed" (similarities) or "variation" (variation): Two words which are not synonyms. The case of MS1 therefore showed some language related difficulties as the formulation in the mathematical definition is ambiguous. In the case of  $MS_2$  only a mathematical context was provided in the answers and there were no no transfer of the meaning in natural language of the word to the mathematical meaning of the word. The comparison of the bi- and monolingual replies showed the differences in the difficulties: In particular, it was noted that the bilingual students seemed to experience semantical difficulties, which were not experienced by the monolingual students in relation to the word "spredning" (spread).

#### ANSWERS IN RELATION TO "HYPPIGHED" (FREQUENCY)

When asked to use the word "hyppighed" eight students were able to do so correctly and only four students were not. Most students used the word "hyppigt" (frequent) instead of "hyppighed" and this has been categorized as a correct answer to 2.A even though the word has been conjugated. Most of the students applied the word as a synonym to the word "ofte" (often) in sentences such as: "*man ser hende hyppigt / you see her frequently*" and "*hvor hyppigt siger du ordet? / how frequently do you say the word?*" and therefore in accordance with the lexical meaning of the word "hyppigt" (frequent) as previously introduced.

# THE CASE OF BILINGUAL STUDENT 8: LANGUAGE RELATED DIFFICULTIES PRESENTED AS EITHER SYNTACTICAL OR SEMANTICAL ERROR

In part 2A and 2B, bilingual student 8, BS<sub>8</sub>, replied as shown in image 6 below:

2. Hyppighed A. Skriv en sætning hvor ordet "hyppighed" indgår med samme betydning, som du bruger det i til daglig Hyppigheden aç den aldersgeuppe efter, der ville VI sour være med Ei1 interviewet fox ex Vav. B. Giv en definition af hvad ordet "hyppighed" betyder, når det bruges som du gjorde i svaret på A ville grad Stemmer petule ١. hvor hoi Hyppighed exsemped hvis roget det J: er For man Corstellige aldersgrupper, hvor mange OF huce AFURPE 110 Spille, Colbold. Sa antal af gersonerse der Jake med ex hyppigheden

#### IMAGE 6: ANSWERS FROM BS8 IN RELATION TO "HYPPIGHED" (FREQUENCY) PART A AND B

BS<sub>8</sub> uses the word in everyday language as: "Hyppigheden af den aldersgruppe vi søger efter, der ville være med til interviewet er for lav / the frequency of the age group we are looking for, who wanted to do the interview, is too low". It is possible that "hyppighed" (frequency) is used to mean "the amount of people" thus it is used as a synonym to the Danish word "mængden" (the amount) or "antallet" (the number of), which would not be in accordance with the definition given by the danish dictionary where "hyppighed" is defined as: "grad af udbredelse; antal forekomster i forhold til samlet antal enheder / degree of prevalence; number of occurrences in relation to the total number of units" (ordnet.dk, 2020 (2)). If this is the case, the answer could be regarded as semantically incorrect.

The formulation of the answer could also be viewed as syntactically wrong if we consider the idea that BS<sub>8</sub> has left out a few words. For example, a similar formulation such as: *"Hyppigheden af dem, der ville være med til interviewet af den aldersgruppe vi søger efter, er for lav / the frequency of those, who wanted to do the interview of the age group we are looking for, is too low"*. The answer would then be more in accordance with the lexical meaning contained in the definition found in the danish dictionary. The language related difficulties of S<sub>8</sub> therefore lies particularly in the formulation as it is not clear what he/she means.

In 2B BS<sub>8</sub> defines "hyppighed"(frequency) as: "Hyppighed ville betyde i hvor høj grad stemmer noget med det vi søger efter. For eksempel hvis man undersøger forskellige aldersgrupper, hvor mange af hver aldersgruppe vil spille fodbold. Så vil antal af personer der vil være med er hyppigheden / Frequency would mean the degree to which something matches what we are looking for. For example, if one examines different age groups, how many of each age group will play football. Then the number of people who want to join is the frequency". The first part of this definition is not in accordance with the provided meaning of the word in part 2A, but the second part of the definition supports the claim that BS<sub>8</sub> has made syntactical errors in the answer to 2A, as this part of the definition now relates to a number of occurrences in relation to the total number of units. Additionally, S<sub>8</sub> does not clearly define the word in 2B but rely on a new example to explain the word.

In question 2C and 2D BS<sub>8</sub> replied as shown in image 7 below:

C. Skriv en sætning hvor ordet "hyppighed" indgår med den betydning du kender fra gymnasiets matematik grupperede observationer. Hypeighod beuges F.'/ D. Giv en definition af ordet "hyppighed" sådan som det bruges i svaret på C grupperede observationer Det betyde 1IV at indenfor hvor dataerae intervaler, brug bler delt 1 90 hyppighed, betyler huor mange 00 gange set bringes Sager efter.

IMAGE 7: ANSWERS FROM BS<sub>8</sub> IN RELATION TO "HYPPIGHED" (FREQUENCY) PART C AND D

Where BS<sub>8</sub> replied: "Hyppighed bruges til de grupperede observationer / Frequency is used to the grouped observations" and defines it as: "Det vil betyde at indenfor grupperede observationer hvor dataerne er blevet delt op i intervaller, bruges der hyppighed, som betyder hvor mange gange bringes

op det vi søger efter / It will mean that within grouped observations where the data has been divided into intervals, frequency is used, which means how many times it is brought up what we are looking for". In 2C BS<sub>8</sub> provided a sentence in which the meaning of the word does not appear per say. Frequency is indeed used when dealing with grouped observations within descriptive statistics, but BS<sub>8</sub> does not construct a sentence where the meaning of the word is evident. In 2D the last part of the definition "… frequency is used, which means how many times it is brought up what we are looking for" does capture the mathematical meaning of "hyppighed" (frequency). The case of BS<sub>8</sub> does show some language related difficulties which could both be semantical or syntactical. In addition, the definition in 2B is formulated ambiguously as the first part of the definition has one meaning, and the second part, which is more of a new example, has another meaning.

THE CASE OF BILINGUAL STUDENT 1: SEMANTICAL ERROR IN THE EVERYDAY USE Bilingual student 1, BS<sub>1</sub>, replied to 2A and 2B as shown in image 8 below:



IMAGE 8: ANSWERS FROM BS1 IN RELATION TO "HYPPIGHED" (FREQUENCY) PART A AND B

In relation to question 2A and 2B, BS<sub>1</sub> replied: "*antal(l)et af unge der ryger er hyppigt / the amount of young people smoking is frequent*" with a following definition as: "*at der er mange der ryger / that there is a lot of young people who smokes*". BS<sub>1</sub> uses the word "hyppigt"(frequent) as an estimate of the amount of young people smoking and it could therefore be viewed as a synonym to the words "stort" (big), "højt" (high). This usage of the word is not in accordance with definition of the word as the synonyms would be "ofte"(often) or "tit"(often). The two answers would indicate a semantical error as the applied meaning by the student does not follow the lexical norm and the sentence thus become syntactically wrong.

In 2C and 2D, BS<sub>1</sub> replied as shown in image 9 below:

C. Skriv en sætning hvor ordet "hyppighed" indgår med den betydning du kender fra gymnasiets matematik HVOC HYPPigt Kommer decimalet

D. Giv en definition af ordet "hyppighed" sådan som det bruges i svaret på C

IMAGE 9: ANSWERS FROM BS1 IN RELATION TO "HYPPIGHED" (FREQUENCY) PART C AND D

In 2C and 2D BS<sub>1</sub> replied: "hvor hyppigt kommer decimalet 14 / how frequent is the decimal 14" and defined it as: "hyppig betyder hvor mange gange / frequent means how many times". BS<sub>1</sub> is thus capable of using the word correctly while maintaining the meaning presented in High School mathematics and the following explanation of the word is not incorrect even though it is unprecise.

THE CASE OF BILINGUAL STUDENT 5: LANGUAGE RELATED DIFFICULTIES

In the case of bilingual student 5, the answers to part C and D are shown in image 10 below:

C. Skriven sætning hvor ordet "hyppighed" indgår med den betydning du kender fra gymnasiets matematik hvor stor er Hyppighedden i stette interfad ] 150; 160] 158, 159, 160, 166, 166 ; 167, 168..... D. Given definition af ordet "hyppighed" sådan som det bruges i svaret på C Pet betyder hvor mange tal er der foa 150 tillbo og der er Hyppigheden = 3

IMAGE 10: ANSWERS FROM BS5 IN RELATION TO "HYPPIGHED" (FREQUENCY) PART C AND D

Where BS<sub>5</sub> replied (translated): "How large is the frequency in this interval ]150;160] 158, 159, 160, 166, 167, 168.... " and afterwards BS<sub>5</sub> defines frequency in 2D as: "Det betyder hvor mange tal er der fra 150 til 160 og der er Hyppigheden = 3 / It means how many numbers are there from 150 to 160 and there the Frequency = 3". The formulation of this answer to 2D is ambiguous as the first part of the definition: "It means how many numbers are there from 150 to 160 to 160" could relate to the length of the interval. It is likely that BS<sub>5</sub> is trying to define it as the amount of observations from the dataset drawn just below, which lies within the interval as the second part of the student's definition:"... and there the Frequency = 3" would imply this. In this case it is interesting that BS<sub>5</sub> does not provide a general definition of the concept of frequency, instead he/she rely on an example by making a data set to explain the intended meaning. The language related difficulties in this case therefore lie particularly in the formulation where it deviates from the mathematical language related to the mathematical classroom discourse as it is formulated in a way which is imprecise causing possible misunderstandings between the intended and perceived meaning.

#### OTHER MATHEMATICAL DEFINITIONS BY BILINGUAL STUDENTS

The bilingual students generally did well when asked to use and define the word in a statistical context. For example, it was used in sentences and defined as shown in figure 3 below:

Student	2C	2D
BS <sub>2</sub>	"Hyppigheden for tallet seks er ni /	"hyppigheden er hvor meget noget bliver
	The frequency for the number six is	indgået / The frequency is how much some-
	nine"	thing is being entered into"
BS <sub>3</sub>	"Hyppigheden for elever med en	"hyppigheden betyder antallet af gange en ob-
	højde på 180 cm er 4 / The frequency	servation i et datasæt fremkommer / The fre-
	for students with a height of 180 cm is	quency means the number of times an observa-
	4"	tion in a data set appears"
BS <sub>7</sub>	"Hyppigheden for str. 42 er 10 / The	"Hvor ofte str. 42 fremkommer i datasættet /
	frequency for size 42 is 10"	How often size 42 appears in the data set"

FIGURE 3: SELECTED CORRECT BILINGUAL STUDENT ANSWERS TO 2C AND 2D

The definition provided by BS<sub>3</sub> is in particular formulated within the academic language used within the mathematical register as it is precise, and the information load is dense and concise. It also contains formal expressions with the use of abstract and technical terms ("observation", "datatsæt" / "observation", "data set" ) and it is actually a definition thus the degree of abstraction is high. This type of answer is also in accordance with what is expected in the mathematical classroom discourse and in particular it is formulated in a way expected in a written media. As a contrast, the answer from BS<sub>7</sub> is an explanation of how the word was used in 2C and thus it relates to the sizes of shoes. It therefore lacks the degree of generality provided by BS<sub>3</sub>, and this answer could therefore be viewed as a less academically formulated answer even though it is clear that BS<sub>7</sub> understands the concept of frequency when using it within the register of mathematics.

The case of monolingual student 1 and 2: Correct definitions and usage of the word in both registers

When asked to use and define the word "hyppighed" (frequency) as used in everyday language, MS<sub>1</sub> replied as shown in image 11 below:

- 2. Hyppighed
  - A. Skriv en sætning hvor ordet "hyppighed" indgår med samme betydning, som du bruger det i til daglig

11 dagng 14 risikom for at få corona, er stor 14 Gygdommen forekommer hype B. Given definition af hvad ordet "hyppighed" betyder, når det bruges som du gjorde i svaret

B. Giv en definition af hvad ordet "hyppighed" betyder, når det bruges som du gjorde i svaret på A offere

IMAGE 11: ANSWERS FROM MS1 IN RELATION TO "HYPPIGHED" (FREQUENCY) PART A AND B

And therefore answers: "*risikoen for at få corona, er større da sygdommen forekommer hyppigere / the risk of getting corona is greater as the disease occurs more frequently*" and defines it as "*oftere / more often*". This usage of the word is lexically correct and the word "*oftere / more often*" can be viewed as a synonym for "hyppigere" (more frequent) even though it can be discussed if it is a definition. In the next part, MS<sub>1</sub> reply as shown in image 12 below:

C. Skriv en sætning hvor ordet "hyppighed" indgår med den betydning du kender fra gymnasiets matematik

on foretunity

D. Giv en definition af ordet "hyppighed" sådan som det bruges i svaret på C Ahtald gf: samme observation.

#### IMAGE 12: ANSWERS FROM MS1 IN RELATION TO "HYPPIGHED" (FREQUENCY) PART C AND D

And answers: "hvor mange gange en observation forekommer / how many times an observation occurs".  $MS_1$  does not actually use the word in the sentence in 2C but instead explains the mathematical meaning of the word. In 2D  $MS_1$  defines the word as: "Antalet af samme observation / the number of the same observation" and thus defines it in accordance with the statistical definition. In 2A and 2B, MS<sub>2</sub> replied as shown in image 13 below:

Hyppighed -ULL Size roges om dasa sattlet.
 A. Skriv en sætning hvor ordet "hyppighed" indgår med samme betydning, som du bruger det i til daglig

 Find hyppigheden og Rettinelsen i lassin med 30 aller
 B. Giv en definition af hvad order "hyppighed" betyder, når det bruges som du gjorde i svaret på A
 Hyppighed fotoeller om et antel of en besternes værde

ud of es Stofin Satasat

IMAGE 13: ANSWERS FROM MS2 IN RELATION TO "HYPPIGHED" (FREQUENCY) PART A AND B

And therefore answers: "Find hyppigheden af røde æbler i kassen med 30 æbler / find the frequency of red apples in the box with 30 apples" and defines it as: "Hyppighed fortæller om et antal af en bestemt værdi ud af et større datasæt / Frequency tells about a number of a certain value out of a larger data set". MS<sub>2</sub> uses and defines the word in a mathematical context which is indicated in both the sentence provided in 2A, which is similar to a formulation of a mathematical problem, but also in the definition provided in 2B, which contain words related to the mathematical register such as "datasæt" (data set).

When asked to use the word in 2C and 2D, MS<sub>2</sub> replies as shown in image 14 below:

C. Skriv en sætning hvor ordet "hyppighed" indgår med den betydning du kender fra gymnasiets matematik

En En fipose internetter talvardice et blandeb Summer, Peter vil genne finde nue af hive mulije gange ob talles 4 de a i Risen, Bestern hyppighteden of 4 i tusatta

D. Giv en definition af ordet "hyppighed" sådan som det bruges i svaret på C

Hypp; chel a burges til at bestemme antelles op es bestemb tal ils større di tat Data sæb hvun es tul kan petremue flare gange. Frekvens

IMAGE 14: ANSWERS FROM MS₂ IN RELATION TO "HYPPIGHED" (FREQUENCY) PART C AND D

And uses the word in the sentence: "En pose med en masse talværdier er blandet sammen. Peter vil gerne finde ud af hvor mange gange af tallet 4 der er i posen. Bestem hyppigheden af 4 i talsættet / A bag with a lot of numerical values is mixed together. Peter would like to find out how many times of the number 4 are in the bag. Determine the frequency of 4 in the set of numbers". MS<sub>2</sub> formulates a sentence as a mathematical problem could be formulated and the defines it as: "Hyppighed bruges til at bestemme antallet af et bestemt tal i et større datasæt hvor et tal kan forekomme flere gange / Frequency is used to determine the number of a particular number in a larger data set where a number can occur multiple times " which is a definition in accordance with the statistical definition of the word and the usage of the word in 2C. The provided definition is also very close to the definition provided in 2B and this is not surprising as MS<sub>2</sub> also provided a mathematical usage of the word in 1A.

#### COMPARISON OF BILINGUAL STUDENT REPLIES AND MONOLINGUAL STUDENT REPLIES

In the previous analysis we saw language related difficulties which were both semantical and syntactical in the bilingual student's replies. In the case of BS<sub>8</sub> there was difficulties in the formulation as the intended meaning became unclear in particular in part 2A and the provided definition in 2B was based on a new example to explain rather than define the meaning of the word. The strategy of using an example to define the concept was also applied by BS<sub>5</sub>. The language related difficulties of BS<sub>8</sub> were not present in the replies provided by neither MS<sub>1</sub> nor MS<sub>2</sub> and the strategy of using examples in the definition was not used either. In fact, MS<sub>1</sub> used and defined the word in accordance with the lexical meaning of the word as a synonym to "ofte" (often) and thus used it in natural language.

In the case of  $BS_1$  there were semantical errors in the everyday meaning, however, this incorrect conception of the word was not present in the mathematical meaning. It would therefore indicate that  $BS_1$  is aware of the fact that the word has different meanings within the different register, and it could indicate that  $BS_1$  does not use the word in his/her everyday language and therefore he/she has an incorrect conception of it. In the case of  $MS_2$ , the provided meaning of the word was similar in the everyday usage of the word and the mathematical usage of the word. However, as  $MS_2$  also provided a mathematical meaning in the everyday usage of the word it could perhaps also indicate that  $MS_2$ , similar to  $BS_1$ , does not use the word in everyday language. Another possible explanation for the usage of mathematical context only by  $MS_2$  can be found within the mathematical classroom discourse as the questionnaire was distributed during a lesson.

The bilingual students did provide a lot of correct replies to 2C and 2D which were also formulated in a concise manner such that the intended meaning was unambiguous. Such an example was the definition provided by  $BS_3$  who used technical terms to define the concept which was more precise than the formulation provided by for example  $MS_1$ .

# FINDINGS IN RELATION TO "HYPPIGHED" (FREQUENCY)

The case of BS<sub>8</sub> showed language related difficulties as it was argued that the answers provided by BS<sub>8</sub> could be regarded as either semantical or syntactical errors. The case of BS<sub>1</sub> showed semantical errors in the general danish conception of the word as the conception was lexically wrong. The provided mathematical meaning of the word was however in accordance with the mathematical definition and there was no semantical transfer of the general danish conception to the mathematical definition. The case of BS<sub>5</sub> showed language related difficulties in the formulation as the answer was formulated in natural language rather than the mathematical register, which made the mathematical sentence and definition imprecise. The cases of BS<sub>2</sub>, BS<sub>3</sub> and BS<sub>7</sub> all showed correct usage and definitions of the word, where particularly the formulation of BS<sub>3</sub> in 2D was formulated in the mathematical register with a higher degree of abstraction than for example BS<sub>7</sub>. In relation to "hyppighed" (frequency), no language related problems occurred with the monolingual students.

# ANSWERS IN RELATION TO "FREKVENS" (RELATIVE FREQUENCY)

The third word in the questionnaire was "frekvens" (relative frequency) and before using this word in the questionnaire, the fact that it is probably not a word commonly used by students in their natural language was considered. This concern was somewhat confirmed by the fact that several of the bilingual students (4) chose not to answer part A and B of question 3.

The bilingual and monolingual students generally struggled with both using and defining the word "frekvens" (relative frequency) when using it in natural language. This can be a consequence of the students not having heard or used the word in natural language outside a classroom thus it becomes difficult to both use and define in this way.

The case of bilingual student 6: semantical transfer from (wrongful) general danish definition to mathematical definition

Bilingual student 6, BS<sub>6</sub>, replied as shown in image 15 below:

3. Frekvens A. Skriv en sætning hvor ordet "frekvens" indgår med samme betydning, som du bruger det i til daglig er lot faerorelis Hickvensen på vejrudsillten B. Giv en definition af hvad ordet "frekvens" betyder, når det bruges som du gjorde i svaret på noget Ondriagon, conf 14 C. Skriv en sætning hvor ordet "frekvens" indgår med den betydning du kender fra gymnasiets find Frehvensen skostorclan out becomes D. Giv en definition af ordet "frekvens" sådan som det bruges i svaret på C Man regner frequension ut via \$ form finder Procent at conditions

IMAGE 15: ANSWERS FROM BS6 IN RELATION TO "FREKVENS" (RELATIVE FREQUENCY)

Where BS<sub>6</sub> wrote: "Frekvensen på vejrudsigten er forfærdelig / the relative frequency of the weather forecast is terrible" and defined it afterwards as "ændringen af noget / the change of something". In everyday language BS<sub>6</sub> has a semantical view of "frekvens" (relative frequency) as a synonym to "ændring" (change) which is lexically different from the meaning of "frekvens" (relative frequency). When asked to use and define the word within a mathematical context, BS<sub>6</sub> wrote: "Find frekvensen af lærerens skostørrelser / Find the relative frequency of the teachers shoe sizes" and "man regner frekvensen ud via en formel som finder procent af ændringen / you calculate the frequence by the use of a formula which finds percent of the change". The interesting part of BS<sub>6</sub>'s answer is in 3B and 3D thus in the definitions. In part 3B, BS<sub>6</sub> views relative frequency as the change of something and this conception is also present in the definition of 3D especially in the last part "… which finds percent of the change".

In the theoretical framework it was presented that students face statistics with strongly held and incorrect intuitions which are highly resistant to change and it was also argued that students often rely on the general danish definition of the word: The case of BS<sub>6</sub> could exemplify this, as the wrongful perception from general danish is the one the student maintains when asked to define the concept mathematically even after having ended a course in statistics.  $BS_6$  thus relies on the general danish conception (misunderstood conception by  $BS_6$ ) of the word rather than the statistical conception of the word.

THE CASE OF BILINGUAL STUDENT 9 AND BILINGUAL STUDENT 3 Bilingual student 9, BS<sub>9</sub>, replied as shown in image 16 below:

3. Frekvens A. Skriv en sætning hvor ordet "frekvens" indgår med samme betydning, som du bruger det i til daglig Bolger har hope frekvens. B. Giv en definition af hvad ordet "frekvens" betyder, når det bruges som du gjorde i svaret på Bolge svingninger er høje med stor afstand mellem toppunterne. C. Skriv en sætning hvor ordet "frekvens" indgår med den betydning du kender fra gymnasiets Hvad er fretiensen for talkt " ".... ?. D. Giv en definition af ordet "frekvens" sådan som det bruges i svaret på C Frekvens athanger at hyppighed og artallet af doseration

IMAGE 16: ANSWERS FROM BS9 IN RELATION TO "FREKVENS" (RELATIE FREQUENCY)

When asked to use the word in everyday language BS<sub>9</sub> wrote: "*Bølger har høj frekvens / waves have a high relative frequency*" and afterwards defined it as "*Bølge svingninger er høje med stor afstand mellem toppunkterne / Wave oscillations are high with great distance between the peaks*". BS<sub>9</sub> defines word as used in physics, and it is debatable if this is considered as "everyday language". This can be caused by several things. As mentioned, the students have perhaps not heard the word used in natural language outside a classroom thus it becomes difficult to both use and define in this way.

Afterwards BS<sub>9</sub> is capable of correctly using the word in a mathematical context with the sentence: "Hvad er frekvensen for tallet ""……? / what is the relative frequency for the number ""……?" and defines it as "frekvens afhænger af hyppighed og antallet af observationer / relative frequency depends on the frequency and the number of observations". This is not incorrect but is not precise either as the exact relation is not provided. One student, who did capture the relation between the concepts was BS<sub>3</sub>, who chose to use and define the word as seen in image 17 below:

3. Frekvens A. Skriv en sætning hvor ordet "frekvens" indgår med samme betydning, som du bruger det i til daglig Frekvenson at røde æbler spist i min hendag er hyppigheler > of nøde æbler divideret med andal æbler spist på én dag. B. Giv en definition af hvad ordet "frekvens" betyder, når det bruges som du gjorde i svaret på A Frekvensen for nøde obber or  $\frac{2}{5} = 0.14 = 40\%$  pa én dag C. Skriv en sætning hvor ordet "frekvens" indgår med den betydning du kender fra gymnasiets Frekensen for observationer "180 cm" er 4 = 0,2 = 20% D. Giv en definition af ordet "frekvens" sådan som det bruges i svaret på C Fuchensen for observation "180 cm" in observationens hyppighted divident med antallet of observationer i alt. (Drs. antal elever).

IMAGE 17: ANSWERS FROM BS3 IN RELATION TO "FREKVENS" (RELATIVE FREQUENCY)

And wrote: "frekvensen for observationen "180 cm" er 4/20=0,2=20% / the relative frequency of the observation "180 cm" is 4/20=0,2=20%" and when asked to define it BS<sub>3</sub> wrote: "Frekvensen for observationen "180 cm" er observationens hypppighed divideret med antallet af observationer i alt (dvs antal elever) / The relative frequency of the observation "180 cm" is the frequency of the observation divided by the number of observations in total (i.e. number of students)". BS<sub>3</sub> does not use the symbolic register to provide a formula of the concept but linguistically explain the formula by the use of the language within the mathematical register.

THE CASE OF MONOLINGUAL STUDENT 1 AND 2: SEMANTICAL DIFFICULTIES AND DIFFICULTIES IN THE FORMULATION

The reply from monolingual student 1, MS<sub>1</sub>, is shown in image 18 below:

3. Frekvens med samme betydning, som du bruger det i A. Skriv en sætning hvor ordet huppitchederne plusset Gammen B. Giv en definition af hvad ordet "frekvens" betyder, når det bruges som du gjorde i svaret på at observationerh3 antallet C. Skriv en sætning hvor ordet "frekvens" indgår med den betydning du kender fra gymnasiets matematik Frekvins fil atsomle bringes alle huppighaderne D. Giv en definition af ordet "frekvens" sådan som det bruges i svaret på C Opsummering

IMAGE 18: ANSWERS FROM MS1 IN RELATION TO "FREKVENS" (RELATIVE FREQUENCY)

Where MS<sub>1</sub> replied: "Alle hyppighederne plusset sammen giver frekvensen / All the frequencies added gives the relative frequency" in 3A and "Frekvens er antallet af observationernes / Relative frequency is the amount of the observations" in 3B and have provided an arrow to indicate and exchange in the answers. However, neither of the sentences uses the word as used in everyday language. From the answers to both 1A and 1B it is indicated that MS<sub>1</sub> uses "frekvens" (relative frequency) as having the same semantical meaning as the total number of observations. This semantically incorrect conception is also present in the usage of the word in 1C: "Frekvens bruges til at samle alle hyppighederne / Relative frequency is used to collect all the frequencies". Finally, as a definition in 3D, MS<sub>1</sub> writes: "Opsummering / summary" perhaps to indicate that this could be used as a synonym to relative frequency. This is not a synonym in accordance with the mathematical definition of the concept, even though it could be argued that the relative frequency is used as a descriptive tool (amongst other concepts) in order to provide a description and summary of a given set of data. However, relative frequency is not a synonym to summary neither in a general danish conception of the word nor in a statistical definition of the word.

The answer provided by monolingual student 2, MS<sub>2</sub>, is shown in image 19 below:

to tul kan forkenner flere gange. 3. Frekvens A. Skriv en sætning hvor ordet "frekvens" indgår med samme betydning, som du bruger det i til daglig here aggine med. How bit gener beingen fellenson be dasserties B. Giv en definition af hvad ordet "frekvens" betyder, når det bruges som du gjorde i svaret på Felevens et en verdi som phaller nget om Stepheten of es datasat. C. Skriv en sætning hvor ordet "frekvens" indgår med den betydning du kender fra gymnasiets matematik Et balsas Encour med so Vardeer, hvor Staticher, of dapasottes stal besterned. Bostern fellversen for dabustetteb D. Giv en definition af ordet "frekvens" sådan som det bruges i svaret på C Fickless pangian and Stephelsen of a data sab 46. dets observationer

IMAGE 19: ANSWERS FROM MS2 IN RELATION TO "FREKVENS" (RELATIVE FREQUENCY)

MS<sub>2</sub> uses the word in 3A as: "John vil gerne finde ud af, hvor stort et datasæt han har at gøre med. Han vil gerne beregne frekvensen for datasættet / John wants to find out how big a dataset he is dealing with. He wants to calculate the relative frequency of the data set" and then defines the word as: "Frekvens er en værdi som fortæller noget om størrelsen af et datasæt / Relative frequency is a value that tells something about the size of a data set". MS<sub>2</sub> uses the word in a mathematical context rather than in everyday language, and more interesting is the formulation in the definition where MS<sub>2</sub> reply that the relative frequency is stating "...something about the size of a dataset". This formulation could indicate that MS<sub>2</sub>, similar to MS<sub>1</sub>, views the relative frequency as a measure of the size of a dataset, i.e., the total amount of observations. This conception is further indicated in the usage of the word in 3A, in particular in the first part where MS<sub>2</sub> writes: "John wants to find out how big a dataset he is dealing with...".

In 3C MS<sub>2</sub> uses the word as: "En opgave med 50 værdier, hvor størrelsen af datasættet skal bestemmes. Bestem frekvensen for datasættet / A problem with 50 values, where the size of the data set must be determined. Determine the relative frequency of the data set" and the indications of MS<sub>2</sub> having a conception of "frekvens" (relative frequency) as an estimate of the total amount of elements in a dataset is also present in this usage of the word. MS<sub>2</sub> thus uses the word semantically incorrect perhaps due to a misconception and this semantical mistake is present in both 3A and 3C. In the definition provided in 3D, MS<sub>2</sub> reply: "Frekvens angiver størrelsen af et datasæt ift. dets observationer / Relative frequency provides the size of a data set in relation to its observations". It could be argued by the last part of the definition "... størrelsen af et datasæt ift. dets observationer/ ... the size of a data set in relation to its observations" in particaular by the usage of the word "ift / in relation to" MS<sub>2</sub> intends to state a quotient between two amounts and now it becomes ambiguous what MS<sub>2</sub> intend as the quotient. Firstly, as MS<sub>2</sub> writes "... størrelsen af et datasæt ift. dets observationer / ... the size of a data set in relation to its observations" the syntax would imply that quotient is:  $\frac{\text{the size of data set}}{\text{its observations}} = \frac{n}{n}$ . The formulation "... dets observationer / ... its observations" could be regarded as the amount of observations in the dataset, such that m=n, and the quotient would thus become:  $\frac{\text{the size of data set}}{\text{its observations}} = \frac{n}{n} = 1$  for all possible data sets.

Secondly, if "... dets observationer / ... its observations" is interpreted as the number of observations for a particular element in the data set, k, such that m=k, then the quotient would be:  $\frac{the \ size \ of \ data \ set}{its \ observations} = \frac{n}{k}$  which is also not in accordance with the mathematical definition of the concept. Hence, the lack of precision in the definition combined with the possible syntactical error makes it ambiguous what the intended meaning provided by MS<sub>2</sub> actually is.

#### COMPARISON OF BILINGUAL STUDENT REPLIES AND MONOLINGUAL STUDENT REPLIES

In the case of BS<sub>6</sub> we saw that the wrongful semantical meaning of the word in everyday language was also present in the sentence and definition provided with a mathematical meaning. This transfer of the everyday meaning of the word was also present with MS<sub>1</sub> who viewed relative frequency as having the same semantical meaning as the total amount of the dataset both in everyday usage and with a mathematical meaning. Interestingly enough, the conception of relative frequency as a measure of the size of the data set was also present in the everyday usage of the word provided by MS<sub>2</sub>. The case of BS<sub>9</sub> exemplified a formulation of an unprecise mathematical definition where it was not clear what the exact relation between the frequency and the size of the data set was. However, this imprecision was also present in the case of MS<sub>2</sub> who provided a mathematical definition and as a result of this and syntactical errors the intended meaning became ambiguous. This formulation can be viewed as a contrast to the definition provided by BS<sub>3</sub> who formulated the mathematical definition in a precise and unambiguous manner with the use of technical terms related to statistics. However, both BS<sub>3</sub> and MS<sub>2</sub> provided only sentences in a mathematical context and this could be explained by the fact that "frekvens" (relative frequency) is not a word commonly used in everyday language.

#### FINDINGS IN RELATION TO "FREKVENS" (RELATIVE FREQUENCY)

The data in relation to "frekvens" (relative frequency) showed incidents of a semantical transfer from a wrongful general danish conception of the word to the statistical meaning of the word in the case of BS<sub>6</sub>. In particular, "frekvens"(relative frequency) was viewed as a synonym to "ændring" (change) and this conception was present in the provided meaning in natural language and in the mathematical context. The cases of BS<sub>3</sub> and BS<sub>9</sub> showed two examples of correct usage of the word in a mathematical context but the two definitions were slightly different as the mathematical definition provided by BS<sub>3</sub> was more explicit on the relationship between the frequency and the amount of observations, than BS<sub>9</sub> was.

In relation to "frekvens" (relative frequency) both  $MS_1$  and  $MS_2$  experienced semantical difficulties as both of the monolingual students viewed relative frequency as an estimate of the size of the data set. In particular  $MS_1$  defined it mathematically as "oppsummering / *summary*" and  $MS_2$  defined it mathematically as "*Frekvens angiver størrelsen af et datasæt ift. dets observationer / Relative frequency provides the size of a data set in relation to its observations*". The quotient implied in the answer of  $MS_2$  was discussed and neither of the interpretations provided a concept in accordance with the mathematical definition of relative frequency.

## ANSWERS IN RELATION TO "UAFHÆNGIGHED" (INDEPENDENCE)

The final word of the questionnaire was the word "uafhængighed" (independence) which the students had to use and define as used in natural language and with the meaning known from High School mathematics. As previously introduced in the theoretical framework, independent has several meanings depending on the register.

The students generally did well when using and defining the word as used in natural language. For example  $BS_{10}$  wrote: "*jeg er uafhængig af dig / I am independent of you*" and defined it as "*jeg kan leve fint uden dig / I can live fine without you*". The definition is therefore more an explanation of the usage of the word in the sentence, than an actual definition and several students had similar answers to the first to questions.

# THE MEANING OF "UAFHÆNGIG" (INDEPENDENT) IN RELATION TO DIFFERENT SUBJECTS OF MATHE-MATICS

When the word "uafhængig" (independent) was chosen for the questionnaire it was done so as this is a lexically ambiguous word in the sense that it has a meaning in general danish where it is related to "frihed" (freedom), "autonomi" (autonomy) and "selvstændighed" (independence). It also has a mathematical meaning where two events A and B are statistical independent if and only if their joint probability can be factorized into their marginal probabilities, i.e.,  $P(A \cap B) = P(A) \cdot P(B)$ . If the two events A and B are statistical independent, then the conditional probability equals the marginal probability i.e. P(A|B) = P(A) and P(B|A) = P(B).

However, none of the bilingual students chose to use this meaning of independence when asked to use it with the meaning they knew from high school mathematics. Instead, the bilingual students used the algebraic meaning used when discussing functions where for example x denotes an independent variable and f(x) is a dependent variable. The table below (figure 4) shows examples of such answers provided by the bilingual students.

Student	4C	4D
BS3	"x er uafhængig af F(x) / x is independent of F(x)"	"Uafhængighed betyder at den ene variabel ikke er afhængig af den anden variabel / Independence means that one variable is not dependent on the other vari- able"
BS <sub>5</sub>	"Hvad er dens uafhængig va- riabel / What is its indepen- dent variable"	"Altså noget som ikke afhænger af den anden variabel / So something that does not depend on the other va- riable"
BS9	"x er uafhængig af y / a is in- dependent of y"	"værdien af x ændres ikke i forbindelse med æn- dringen af y-værdien / the value of x does not change in relation to the change of the y value"
BS <sub>11</sub>	"I en graf er alle x-værdier uafhængig af y-værdier / In a graph, all x-values are inde- pendent of y-values"	"Uafhængig i matematisk sammenhæng betyder at en variabel ikke ændrer ens værdier på baggrund af en anden variabel / "Independent in a mathematical con- text means that a variable does not change one's val- ues on the basis of another variable"

FIGURE 4: BILINGUAL STUDENT REPLIES TO 4C AND 4D

As this is outside the scope of this thesis, the answers related to "uafhængighed"(independence) thus did not create any empirical evidence in relation to the research questions of this thesis and the results will be reviewed in the discussion.

#### MONOLINGUAL REPLIES

Monolingual student 1, MS<sub>1</sub>, also chose to define "uafhængighed" (independence) with the algebraic meaning, but MS<sub>2</sub> answered as shown in image 20 below:

- 4. Uafhængig
  - A. Skriv en sætning hvor ordet "uafhængig" indgår med samme betydning, som du bruger det i til daglig

tallan 3 cg 4 a udphenging of himmen, don the star op i 4. 3 ca estid tal & racoline to)

B. Giv en definition af hvad ordet "uafhængig" betyder, når det bruges som du gjorde i svaret på A

C. Skriv en sætning hvor ordet "uafhængig" indgår med den betydning du kender fra gymnasiets matematik

D. Giv en definition af ordet "uafhængig" sådan som det bruges i svaret på C

IMAGE 20: ANSWERS FROM MS2 IN RELATION TO "UAFHÆNGIGHED" (INDEPENDENCE)

Where MS<sub>2</sub> used the word in natural language as: "tallen 3 og 4 er uafhængige af hinanden, da 3 ikke går op i 4. 3 er et ulige tal og 4 er et lige tal / the numbers 3 and 4 are independent of each other as 3 does not divide 4. 3 is an odd number and 4 is an even number". In 4A, MS<sub>2</sub> uses the word in a mathematical context and describes "uafhængig" (independent) as the quotient  $\frac{4}{3} = n$  for some  $n \in$ N, but the following definition in 4B is not clear. MS<sub>2</sub> writes: "uafhængighed angiver hver tal af et sæt værdier ift hinanden / independence indicates each number of a set of values in relation to each other". This formulation is of this definition does contain syntactical errors, and if we assume that MS<sub>2</sub> have left out words, and wants to maintain the meaning provided in 4A, then the intended sentence could be similar to: "uafhængighed angiver hvert tal ift hinanden i et sæt værdier / independence indicates each number in relation to each other in a set of values. This formulation would relate to the usage of the word in 4A, if "... *ift hinanden /* ... *in relation to each other*" is to be interpreted as the quotient of the numbers. The answer from  $MS_2$  could thus indicate semantical misconceptions in 4A and the syntactical errors of 4B making the formulation and intended meaning unclear, could therefore indicate some language related problems.

In 4C MS<sub>2</sub> reply: "Bestem uafhængigheden for datasættet med de 50 værdier / Determine the independence of the dataset of 50 values" and defines the word in 4D as: "Viser hvor uafhængige værdier er ift hinanden i et datasæt / shows how independent values are in relation to each other in a data set". The two provided definitions in 4B and 4D are thus very similar expect for the notion in 4D of "… hvor uafhængige / … how independent" the values are and, in this definition, it is not clear, how MS<sub>2</sub> measures the independence.

In 4C, it is implied in the usage of the word "bestem" (determine) that MS<sub>2</sub> has a conception of "uafhængighed" (independence) as something to be calculated. It can be argued that in the notion of independence as  $P(A \cap B) = P(A) \cdot P(B)$  some calculations can be provided in order to check if two stochastic variables of a data set are independent, but combined with the provided definition in 4D it is more likely that MS<sub>2</sub> has a conception of independence as a specific value to be determined from a single calculation as oppose to checking if the statement  $P(A \cap B) = P(A) \cdot P(B)$  is true.

#### A PARTIAL CONCLUSION

In the context of statistics and lexically ambiguous words we have seen several cases of difficulties in the formulation. This was for example the case with BS<sub>2</sub> (in relation to "spredning"(spread)), BS<sub>5</sub> (in relation to "hyppighed"(frequency)), BS<sub>8</sub> (in relation to "hyppighed"(frequency)). However, this was not exclusively seen with the bilingual students as it was also the case of MS<sub>2</sub> (in relation to "frekvens"(relative frequency)). The difficulties in the formulation of the bilingual students and the monolingual student were different, which can be exemplified in the case of BS<sub>2</sub> (in relation to "spredning"(spread)) who had a formulation with syntactical errors which made it difficult to extract a meaning of the provided sentence. In the case of MS<sub>2</sub> (in relation to "frekvens"(relative frequency)) it was possible to extract a meaning in the definition, but the formulation was ambiguous and it thus became possible to extract different meanings of the statement.

The data also showed several cases of semantical difficulties amongst bilingual students. This was for example the case with  $BS_1$  (in relation to "spredning"(spread)),  $BS_2$  (in relation to "spredning"(spread)),  $BS_8$  (in relation to "hyppighed" (frequency)),  $BS_1$  (in relation to "hyppighed"

(frequency)) and BS<sub>6</sub> (in relation to "frekvens"(relative frequency)). However, some semantical difficulties also occurred with the monolingual students in particular in the case of  $MS_1$  (in relation to "frekvens"(relative frequency)) and  $MS_2$  (in relation to "frekvens"(relative frequency)). The data also showed several difficulties in the formulation

# PROBABILITY

#### METHODOLOGY

One of the main foci of this thesis is language related difficulties that may arise for learners whose first language is not the language of instruction in the context of statistics and probability teaching. In order to analyze these difficulties in relation to probability, I have conducted observations made at Frederiksberg Gymnasium in a second-year math class (B-level, STX). In the class there were 26 students where 8 of them were bilingual, and the observations were made over 3 separate days, where the teaching was divided into so called "modules" on each day which lasted approximately 1,5 hours (from 10.05 to 11.45 including a 10 minutes break in between). The students were in the beginning of learning probability and had previously been introduced to basic combinatorics and had been working with this for a while.

During observations the students were introduced to the binomial coefficient, symmetric and asymmetric probability fields, the sum of probabilities, calculations of the probability of two independent events, A and B, where they were looking at both  $P(A \cap B)$  and  $P(A \cup B)$ , a definition of stochastic variables, and calculations of the mean of stochastic variables. In addition, they were also introduced to binomially distributed stochastic variables and how to do calculations by the use of CAS-tools.

The observations had three main foci, where the first one was on the use of language used by the students, thus the observations focused on how the students communicated in formulation phases i.e. if they were expressing their ideas in the language of mathematics or in natural language both in adidactical and didactical situations. In this relation there was also a focus on how bilingual students interpreted what was being communicated by other students and the teacher. The second focus of the observations were the language related problems arising in understanding of the presented mathematical concepts in particular if there were any semantical or syntactical difficulties. The third focus was on how bilingual students worked with different types of problems within probability, in particular how they worked on 1) problems which were similar to what the teacher had calculated in an example i.e. isomorphic problems, 2) non isomorphic problems, 3) mathematical problems and 4) word problems. In the section "Categorization of the types of problems" the categorization of the types of problems will be more explicit.

While observing notes were made and audio recorded on a Dictaphone. Selected passages of the conversations have been transcribed and translated into English in the following line and is marked

in *italic*. Only selected, relevant passages have been included in the analysis, but a full transcription of the selected episodes can be found in the appendix (appendix 8, 9, 10 and 11). In the transcriptions, I will write T to symbolize the teacher,  $BS_i$  to symbolize the bilingual students, and  $MS_i$  to symbolize monolingual students. The numbering of the students in the different observations starts from 1 in each observation, thus  $BS_1$  in observation no. 1 is not the same as  $BS_1$  in observation no. 2.

## INTRODUCTION TO THE MILIEU

The students were not following a textbook and it was therefore not possible to make a textbook analysis, which could have been used in the analysis of the student conversation. Instead of a textbook, the class relied on the official collection of formulas from the Ministry of Education (EMU, 2020) as a frame of reference. During the observations the teacher referred to the collection of formulas several times and used the number listed in it. For each module the students were electronically handed out so-called "worksheets" containing preselected problems in relation to the topic of the day and these worksheets have been added in the appendix (appendix 2a, 2b, 3a, 3b, 4a and 4b).

An à priori analysis of selected problems of the worksheets has been conducted in order to analyze the observations, where correct and incorrect solution techniques are presented. These problems have been selected to the à priori analysis, as these were the ones involved in the selected passages of the observations. In addition to the worksheets, the teacher had prepared a PowerPoint presentation for each module containing both selected formulas, examples using the formulas, and a few small problems. The slides of these PowerPoint presentations have been added to the appendix as well (appendix 5, 6 and 7).

# THE USE OF COMPUTER AND CAS-TOOLS

The use of Computer Algebra Systems (CAS-tools) consist as a part of the milieu and the students were using WordMat as an aide in the making of calculations and illustrations. WordMat is an addon to Microsoft Word functioning as a CAS-tool and has several standardized calculations in relation to both statistics and probability and provides several standardized Excel sheets depending on which distribution, the students are working on. As seen in image 21 just below the students can choose between a binomial distribution, a normal distribution, a  $\chi^2$ -distribution and a t-distribution.



IMAGE 21: SCREENSHOT OF HOW TO ACCES THE EXCEL SHEET IN RELATION TO THE BINOMIAL DISTRIBUTION IN WORD MAT

After choosing for example the binomial distribution an Excel sheet as pictured in image 22 below will open:



IMAGE 22: SCREENSHOT OF THE EXCEL SHEET PROVIDED BY WORDMAT

Where the students can change the probability parameter, p, and the number parameter, n, (in the read circle) and both the mean,  $\mu$ , and the spread,  $\sigma$  is the calculated. It is also possible to calculate the probability of stochastic variables by entering the values (in the green circle). The use of WordMat as a CAS-tool to assist in the calculations and illustrations within probability therefore has several advantages: 1) It is easy to use, 2) it is fast at calculating probabilities, mean, and spread for stochastic variables and 3) it easily provides graphical illustrations which can aide in visualizing the data. However, the correct use of CAS-tool relies on an understanding of the symbolics related to probability,

as the students need to know what the symbols represents in order use it as an effective aide. The use of CAS-tools has limitations as it aides the students in making calculations easily but they still have to interpret and formulate their own answers.

During the observations the students took notes and wrote down their solutions in a digital worksheet, which made it difficult to obtain images of students' written material from their computer screen without interfering the in the situations and at the same time obtain an image in such a quality that it could be used in the analysis. I therefore chose not to take images of the students' screens nor have them recording their work at their computers. Instead, the data was collected as images of the blackboard, the digital versions of the worksheets, digital versions of the PowerPoint presentations, audio, and in relation to some of the observations made, a reconstruction of the students' solutions when it was related to the use of WordMat as a CAS-tool. In particular, this was relevant in observation no. 4.

# CATEGORIZATION OF THE TYPES OF PROBLEMS

There will be a distinction between isomorphic and non-isomorphic problems, where *isomorphic problems* will be defined as mathematical exercises with the same structure distinguished by the context and/or the numbers included. For example, the following problems:

P1: "Determine the number of ways one can choose 4 element out of a total of 10"

P<sub>2</sub>: "Determine in how many ways one can choose 3 elements out of a total of 15"

P<sub>3</sub>: "In how many ways can one choose 4 students in a class of 10 students?"

Are all isomorphic as they can be solved by using the same technique. *Non-isomorphic* problems will be defined as the problems, which are not isomorphic.

# À PRIORI ANALYSIS OF SELECTED PROBLEMS OF THE WORKSHEETS

#### PROBLEM 1, WORKSHEET 1

As combinatorics is an essential part of probability the students had to work on problems related to the binomial formula. Problem 1 is presented in figure 5 below in a translated version.

# Opgave 1 Problem 1

På et volleyballhold er der tilmeldt syv spillere. Der er altid seks på banen af gangen. *At a volleyball team there are seven players signed up. There is always six people at the field at once.* 

- a) På hvor mange forskellige måder kan man vælge seks spillere til at starte på banen? In how many ways can one choose six players to start at the field?
- b) Hvor mange forskellige måder kan man sætte et starthold på, hvis der havde været tilmeldt otte spillere?
   *In how many different ways can one make a starting team if there had been eight players?*

# FIGURE 5: TRANSLATION OF PROBLEM 1 OF WORKSHEET 1

This problem will be the subject of the first a priori analysis. The problem can be categorized as an isomorphic problem as the students had previously been working with problems which required the same technique to solve it. In particular, it can be solved by the use of the students' formula no. 169 (EMU.dk, 2020):  $K(n,r) = \frac{n!}{r!(n-r)!}$  which provides the number of ways to choose an (unordered) subset of r elements from a fixed set of n elements. Before the students were to work on problem 1 of worksheet 1 (figure 5), the teacher and the students had reviewed and explained the assignment from the related PowerPoint slide no. 3 (appendix 5) which is displayed in figure 6 just below in a translated version.

1. Fire elever skal til mundtlig årsprøve i engelsk. Hvor mange forskellige rækkefølger kan eleverne komme op i? (Hjælpemidler tilladt)

Four students have to take an oral exam in English. In how many different orders can the students take the exam? (Aids allowed)

2. Bestem 26! (Hjælpemidler tilladt) og kom med et eksempel på, hvad tallet siger noget om. *Determine 26! (Aids allowed) and give an example of what the number says something about.* 

3. Bestem  $\frac{6!}{4!}$  (Uden hjælpemidler) Determine (Without aids)

4. Ud af 26 elever i 2.b skal fire vælges til et trivselsudvalg. Hvor mange forskellige kombinationer af elever kan vælges? (Hjælpemidler tilladt)

Out of 26 students in 2.b, four must be elected to a well-being committee. How many different combinations of students can be chosen? (Aids allowed)

FIGURE 6: THE TRANSLATED PROBLEM FROM POWERPOINT SLIDE NO. 3 (APPENDIX 5)
Both part a) and b) of problem 1 (figure 5) is thus very similar to part 4 of the assignment which had just been reviewed by the teacher, however, there is a slight difference in the way it is formulated. In particular, there is a difference between: "*Out of 26 students in 2.b, four must be elected* ..." where it is quite clear that you are to choose 4(=r) out of 26(=n) and it is also intuitive why the order of the selection does not matter. However, in the formulation: "*In how many ways can one choose six players to start at the field?*" and in this formulation there might occur some language related problems, as there is a possible misinterpretation in thinking that the order of the players matters when selecting players to start at the field. If this is the case, the students might apply the following technique, t<sub>1</sub>, to solve the problem and use formula no. 168 (EMU.dk, 2020) stating that:  $P(n,r) = \frac{n!}{(n-r)!}$  Which provides the number of options for selecting r elements from n elements when the order matters and obtain:

$$P(7,6) = \frac{n!}{(n-r)!} = \frac{7!}{(7-6)!} = \frac{7!}{1!} = 7! = 5040$$

Meaning that there should be 5040 different way to select a starting team when having 7 players. The students could calculate this number either by the use of CAS or by hand do the calculation. If the students apply this technique to solve the problem, it could indicate that they believe it is isomorphic to part 1. in figure 6.  $T_1$  is thus not a wrong technique mathematically: It has just been wrongfully applied perhaps due to a misconception caused by language related problems in the interpretation of the problem as it could indicate that the students believe that "forskellige rækkefølger" (differents orders) and "forskellige måder" (different ways) are semantically indifferent. However, they are not in this context.

Some students might correctly read the formulation of part a) of problem 1: "*In how many ways can one choose six players to start at the field*?" and define r=6 and n=7 and then choose to apply technique 2, t<sub>2</sub>, by using formula 169 (EMU.dk, 2020) and obtain that:

$$K(7,6) = \frac{n!}{r!(n-r)!} = \frac{7!}{6!(7-6)!} = \frac{7!}{6!} = 7$$

And obtaining a correct answer. Finally, a third technique,  $t_3$ , to solve the problem is to view the problem the other way around and solve: In how many ways can one player be put outside the field if there are 7 players in total? By doing this, the student will be able to successfully apply both  $t_1$  and  $t_2$  will provide the correct answer as:

$$P(7,1) = \frac{n!}{(n-r)!} = \frac{7!}{(7-1)!} = \frac{7!}{6!} = 7$$
$$K(7,1) = \frac{n!}{r!(n-r)!} = \frac{7!}{1!(7-1)!} = \frac{7!}{6!} = 7$$

T<sub>3</sub> therefore uses the fact that K(n, r) = K(n, n - r) by symmetry of Pascal's triangle. However, as the students has not been introduced to Pascal's triangle it is not likely that they will be able to explain the use of t<sub>3</sub> by arguments relation to Pascal's triangle.

The students could apply three techniques to obtain an answer to problem 1 a) and there is didactical potential in both the application of the techniques but also in a potential following discussion between students who has applied different techniques (and thus obtained different answers). The possible language related problems now lie within the transition from natural language to the formal language of mathematics, as the students have to argue by the use of the register of mathematics in order to explain why heir solution is correct. As previously introduced in the theoretical framework, it is impossible to separate the use of language and mathematics and it is particular in the formulation phases that this becomes evident. Language related problems that might arise could be that the students struggle to find the appropriate words to describe their arguments where indications of this could be repetitions of the same argument or half-finished sentences. The language related problems may also arise within the mathematical register as the students have been handed out a collection of relevant formulas formulated both linguistically and mathematically, an argument to the usage of both  $t_1$  and  $t_2$  could be found in this collection (EMU.dk, 2020).

Part b) of problem 1 is isomorphic to part a) as the structure of the problems are identical thus the techniques to solve the problem from part a) remains the same and the students will probably use the same technique in b) as they did in a) due to the isomorphy. Some of the students who successfully applied  $t_3$  and then  $t_1$  in part a) will perhaps choose the same strategy, however, this time it will not be a successful technique as:

$$P(8,2) = \frac{n!}{(n-r)!} = \frac{8!}{(8-2)!} = \frac{8!}{5!} = 336$$

As oppose to applying t<sub>3</sub> and then t<sub>2</sub>:

$$K(8,2) = \frac{n!}{r!(n-r)!} = \frac{8!}{2!(8-2)!} = \frac{8!}{2! \cdot 6!} = 28$$

Which will still provide a correct answer. The didactical potential is again in the following discussion between students who have applied different techniques, obtained different answers and therefore needs to argue why the selected solution strategy is the correct one.

### PROBLEM 2, WORKSHEET 1

The second problem of worksheet has fewer potential language related problems as it illustrates a different type of problem. Problem 2 can be viewed in figure 4 in a translated version.

It is an isomorphic problem, as the students had previously worked on similar problems and the context is 1:1 isomorphic. Thus, only the numbers to be calculated has changed. In problem 2 a), b) and c) the correct technique is more obvious as there is no need to interpret or model a technique as it is already given. The students are to use formula 169 (EMU.dk, 2020) stating that

$$K(n,r) = \frac{n!}{r! (n-r)!}$$

# Opgave 2 Problem 2

- a) Bestem *K*(3, 2) *Calculate K*(3, 2)
- b) Bestem *K*(5,3) *Calculate K*(5,3)
- c) Bestem *K*(7,2) *Calculate K*(7,2)
- d) Bestem *K*(7,5)*Calculate K*(7,5)

Sammenlign dine svar i Opgave 2c og 2d.

Compare your answers in problem 2c and 2d.

FIGURE 7: PROBLEM 2 FROM WORKSHEET 1 (APPENDIX 2A/2B)

Problems that may arise in the usage of this technique could be within the symbolic register i.e., if the students do not know what "!" symbolizes. If the students choose to solve problem 2 by hand, they might encounter some problems in this relation as they then cannot calculate for example:

$$K(3,2) = \frac{3!}{2! \cdot (3-2)!} = \frac{3!}{2! \cdot 1!}$$

If they do not now the symbolical meaning of "!". As the observed class relied on digital worksheets and the use of WordMat as a CAS-tool most of the students will most likely chose to do the calculations by the use of WordMat and the integrated calculator. The students can thereby enter: K(3, 2) =  $\frac{3!}{2!\cdot(3-2)!}$  and then by the calculation command (CTRL + b / command + b), the CAS-tool calculates the value. By the use of CAS the students are therefore able to obtain an answer to this type of problem without possibly fully understanding the involved symbolic of the applied technique.

As parts a), b), c) and d) are all isomorphic they can be solved by the use of the same technique and by the use of CAS it can be done without fully understanding the applied symbolics involved. The didactical potential of these four assignments is therefore relatively low, however, there lies some potential in the last part of the problem when the students has to compare the answers of 2c and 2d. By usage of technique 2 the students will obtain that:

$$K(7,2) = \frac{n!}{r!(n-r)!} = \frac{7!}{2! \cdot (7-2)!} = 21$$
$$K(7,5) = \frac{n!}{r!(n-r)!} = \frac{7!}{5! \cdot (7-5)!} = 21$$

The students are therefore expected to explain the fact that K(n,r) = K(n,n-r) by symmetry of Pascal's triangle. As the students have not been introduced to Pascal's triangle it is not likely that they will be able to use arguments in relation to Pascal's triangle. The students might argue in relation to problem 1 of worksheet 1 if they have applied  $t_3$  in that assignment. The students will therefore be able to explain why the two numbers become the same by relying on a previous example. Another answer could be formulated in relation to the commutative law of multiplication thus explaining that  $2! \cdot 5! = 5! \cdot 2!$ . This formulation would not be as precise as an argument related to Pascal's triangle, but it would still be formulated in a mathematical register and a more academic language as it is denser and uses formal expressions with the use of abstract and technical terms in relation to our definition in the theoretical framework. The language related problems which could arise in this last part of problem 2 is similar to the ones of problem 1 and involved difficulties finding the right words both within natural language and within the mathematical register.

THE "SLEEPING PROBLEM" FROM POWERPOINT NO. 3, SLIDE 3 (APPENDIX 7)

In the third module the students had been introduced to binomial distribution and they worked on the problem from PowerPoint presentation no. 3, slide 3 (appendix 7) which I will refer to as the "sleeping problem". It can be viewed in a translated version in figure 8 just below.

I en klasse går der 20 elever, der hver har 5 % sandsynlighed for at sove over sig. Den binomialfordelte, stokastiske variabel X betegner antallet af elever, der sover over sig på en given dag. *In a class there are 20 students, each with a 5% probability of sleeping over. The binomially distributed, stochastic variable X denotes the number of students sleeping over on a given day.* 

- *a)* Bestem antalsparameteren og sandsynlighedsparameteren. Determine the number parameter and the probability parameter.
- b) Bestem middelværdien, og forklar, hvad middelværdien siger om eleverne.
  Determine the mean and explain what the mean says about the students.
- c) Det oplyses, at P(X = 0) = 0,36. Hvad betyder dette? It is stated that P (X = 0) = 0.36. What does this mean?

FIGURE 8: THE SLEEPING PROBLEM FROM POWERPOINT PRESENTATION NO. 3, SLIDE 3 (APPENDIX 7)

Part a) and the first part of b) are both isomorphic problems as the structure is identical to previous encountered assignments and the students will most likely fare well in part a. In part b) of the problem a correct technique to solve the problem,  $t_5$ , would be to apply formula 189 (EMU.dk, 2020) stating that  $\mu = n \cdot p = 20 \cdot 0.05 = 1$ . The use of this technique requires that the students understand the symbolics involved, i.e., that  $\mu$  symbolizes the mean, n symbolizes the number parameter, and that p symbolizes the probability parameter. As this formula was also displayed on the slide of the Power-Point (appendix 7) stating that the mean is  $\mu$ , it is very likely that most students will fare well in this part of the problem. However, in the second part of b) when the students have to "... *explain what the mean says about the students*" there might occur some language related problems. A precise answer could be similar to: "Gennemsnitligt er der en elev der sover over sig hver dag / On average one student will oversleep per day" This formulation is precise enough to be considered a correct answer as it captures the definition of the mean,  $\mu$ , of a binomially distributed stochastic variable as in particular being the average value. Some of the students might struggle to formulate the answer in a precise and concise manner, and such answers could for example be similar to the following:

1. Der er chance for at en elev sover over sig hver dag / *There is a chance that one student will oversleep every day* 

- 2. Sandsynligheden er at en elev sover over sig / *The probability is that one student will oversleep*
- 3. En elev vil sove over sig hver dag / One student will oversleep every day

An answer similar to the first one, would not be considered precise as "chance" (chance) is not a defined concept within probability and the word "chance" is also not related to "middelværdi" (mean) or "gennemsnit" (average) neither in the general danish definition no within the register of mathematics. According to the danish dictionary (ordnet.dk, 2020 (4)) the word "chance (chance)" means: "mulighed for et positivt resultat eller en gunstig udvikling fx sejr, fremgang eller succes / opportunity for a positive result or a favorable development eg victory, progress or success" (ordnet.dk, 2020(4)). This formulation would therefore not capture the definition of the mean as an average value and hence it does not explain the precise meaning of the calculated value. An answer similar to this would therefore indicate language related problems as the lack of precision in the formulation effects what being communicated.

The second type of answer is also not a precise description. In this answer the student uses the word "sandsynligheden" (the probability) as a way to describe that is very likely that a student will oversleep. This would indicate a usage of the word "sandsynlighed" (probability) as done so in general Danish in the expression: "Efter al sandsynlighed / in all likelihood" which is used to express that something is very or most likely (ordnet.dk, 2020(3)). This type of answer could thus indicate that the student is using the general danish definition of "sandsynlighed" (probability) rather than the definition provided within the mathematical register. A formulation similar to the second one could therefore indicate language related problems as the lack of precision changes the meaning of what is being communicated. Finally, an answer similar to the third one, is very close to being precise and concise, however it is missing an essential part of the definition of the mean as the mean is stating an average value. This formulation indicates that the calculated value of the mean is always occurring, however, this is not true. Other language related difficulties in this part of the problem could be that the students are not capable of finding the right words to provide a sufficient answer. Indications of this could for example be half-finished sentences followed by either a new half-finished sentence or a finished sentence. Part c) of the problem is very similar to part b) and the language related difficulties therefore remains the same in this part of the problem.

As stated so far in the a priori analysis a lot of the didactical potential lies in part b) and c) where the students have to explain the meaning of calculated values in specific situations. This formulation

phase requires, in particular in part b), that the students are being precise. Otherwise, there is a possibility that what they are communicating can be misunderstood by other students or simply being incorrect.

PROBLEM 1, WORKSHEET NO. 3

In module 3 the students had to work on worksheet 3 (appendix 4a and 4b) where the (correct) use of CAS became essential. Problem 1 of worksheet 3 is presented in figure 9 just below in a translated version.

# Opgave 1 (med hjælpemidler) Problem 1 (aids allowed)

*X* er en binomialfordelt stokastisk variabel med antalsparameter 80 og sandsynlighedsparameter p = 0,35. *X is a binomially distributed stochastic variable with number parameter 80 and probability parameter p = 0,35*.

- a) Tegn et søjlediagram for sandsynlighedsfordelingen for *X*. *Make a bar chart of the probability distribution for X*.
- *b)* Bestem middelværdien og spredningen for *X*. *Determine the mean and spread/standard deviation of X*.
- c) Bestem P(X = 28). Dertermine P(X = 28).

FIGURE 9: PROBLEM 1 FROM WORKSHEET NO. 3 (TRANSLATED) (APPENDIX 4A / 4B)

Before the students were to work adidactically on this problem, they had been introduced to the usage of WordMat to solve isomorphic problems, where the teacher institutionalized the technique in the beginning of the module. The student solutions for part a) could therefore be by the use of CAS as illustrated in image 23, where the student would apply the technique,  $t_5$ , which would be to enter the given values of p and n into the relevant places in the ExcelSheet such that a bar charge is automatically created. This solution can be viewed in image 23 below:



#### IMAGE 23: CAS-SOLUTION OF PROBLEM 1, WORKSHEET 3

Using CAS as a technique to solve this problem mainly requires that the students understand the symbolics involved. As seen in image 23, the use of language within this CAS-tool is within the symbolic register and the usage of this technique requires that the students relate the word "antalsparameter 80 / number parameter 80" with the symbolic representation "n" used by the CAS-tool. In addition, the correct usage of CAS also requires that the students choose the correct distribution as four different distributions are available. However, as it is quite clearly formulated in the problem that the variable is binomially distributed the students are likely to choose the correct distribution.

Some students might choose to use CAS for part b) as well and therefore use the technique previously institutionalized by the teacher. This would involve the students reading the automatically calculated value of the mean and the spread from the calculations in a) as the CAS-tool has automatically calculated  $\mu = 28$  and  $\sigma = 4,266$ . Possible language related problems within the symbolic register could occur as the use of this technique rely on the understanding of the involved symbolics  $\mu$  and  $\sigma$  which represents the mean and the spread. Some students might choose to calculate the values of the mean and spread by the usage of formula 189 (EMU.dk, 2020) and applying t<sub>6</sub> and calculate:

$$\mu = n \cdot p = 80 \cdot 0,35 = 28$$

The use of this technique could indicate language related problems particular in relation to the symbolics involved by the CAS-tool. However, several other explanations are possible: Perhaps, the student prefers to use the mathematical formulas, the student wants an explanation to the number provided by the CAS-tool, or perhaps the student has simply missed that the CAS-tool does the calculation automatically. The same applies in the calculation of the spread,  $\sigma$ .

In part c) of the problem a student solution could also be by the use of CAS as illustrated in image 24 just below:



IMAGE 24: CAS-SOLUTION OF PROBLEM 1, WORKSHEET 3

In this solution X=28 has been entered and once more this technique requires that the student understands the language within the symbolic register.

The didactical potential in this problem is relatively low as the problem is an isomorphic problem which can be solved with the use of a CAS-tool provided that the student understands the symbolics concerning the number parameter, the probability parameter, the mean, and the spread. Obviously, the students will also need to know how to use WordMat (or at least simple functions) in order to

access the standardized Excel Sheet. The problem does not imply any obvious need for applying mathematical arguments why an applied technique is correct or incorrect and the formulation phase is therefore merely in relation to the use of CAS and not the applied mathematics nor the meaning of the calculated values.

# **OBSERVATIONS**

A total of four observations have been selected for analysis, and these are now presented. Observation no. 1 is split into three situations as observation no. 1 was rather long, but each situation showed variations of language related problems. Similarly, observation no. 3 has been split into two parts.

# CLASSROOM OBSERVATION NO. 1: LANGUAGE RELATED DIFFICULTIES IN THE FORMULATION PHASE

During the observations in the first module (1/10-2020) the students had previously been introduced to the binomial coefficient and had seen it stated as formula 169 (EMU, 2020, p. 19) where it is stated as: "Antal muligheder for udvælgelse af r elementer blandt n elementer, når rækkefølgen ikke har betydning / the number of obtains for selection of r elements among n elements, when the order is not significant" and then the formula:  $K(n,r) = \frac{n!}{r!(n-r)!}$ . Prior to this episode the students had been working on the problem presented in the a priori analysis (figure 6).

# SITUATION 1: BILINGUAL STUDENTS WHO STRUGGLE WITH WORD PROBLEM BUT NOT WITH STRAIGHT FORWARD CALCULATIONS

The students were working adidactically in small groups on problem 1 of worksheet 1 (appendix 2a/2b). There was a group of three bilingual students (BS<sub>1</sub>, BS<sub>3</sub> and BS<sub>4</sub>), a single bilingual student BS<sub>2</sub> who worked alone, and another group of bilingual students BS<sub>5</sub> and BS<sub>6</sub>. The transcript of the entire conversation can be seen in appendix 8. BS<sub>1</sub>, BS<sub>3</sub> and BS<sub>4</sub> had applied technique 1 and calculated that:

$$P(7,6) = \frac{n!}{(n-r)!} = \frac{7!}{(7-6)!} = \frac{7!}{1!} = 7! = 5040$$

Meaning there should be 5040 different ways to choose six players to start at the field. The students realize that it can't be right in the following conversation with BS<sub>2</sub>:

BS<sub>2</sub>: Har I lavet opgave et? / Did you finish problem one?

BS<sub>3</sub>: Jeg tror det er sådan her.. syv parentes gange fem gange fire gange tre gange to gange en. Jeg tror, jeg ved det ikke, det er sådan det er mig der har svært ved at sige det. / *I think it is like this.. Seven parenthesis times five times four times three times two times one. I think, I don't know, it is like this, it's me who is having difficulties saying it.* 

• • •

BS<sub>2</sub>: Der kan ikke være fem tusind forskellige muligheder for syv spillere. Det passer ikke. / *There can not be five thousand different options for seven players. That is not true.* 

•••

BS<sub>4</sub>: Ej det her det kan ikke være rigtigt der er noget galt. / *This can not be true something is wrong* 

BS<sub>2</sub>: Ja ikke? Der er et eller andet galt / Yeah right? Something is wrong

BS3: Altså vi kan lave den næste opgave / Well we can do the next problem

FIGURE 10: PART OF THE TRANSCRIPTION OF OBSERVATION NO. 1 (APPENDIX 8)

 $BS_1$ ,  $BS_3$  and  $BS_4$  then continues to work on problem 2 instead which is a mathematical problem isomorphic to previously encountered problems. On this assignments  $BS_1$ ,  $BS_3$  and  $BS_4$  fare well and solve the problems using technique 4.

The episode is interesting as the students all realize that ".. something is wrong" as 5040 seems to be too many options. As reviewed in the theoretical framework some of the common errors for students within combinatorics is "... overcounting and confusing situations in which order matters or does not matter" (Lockwood, Swinyard & Caughmann, 2015, p. 29) and this situation can be regarded as an example of this. In addition, neither of the bilingual students choose to review the linguistic formulation of the applied formula which is stated as: "Antal muligheder for udvælgelse af r elementer blandt n elementer, når rækkefølgen har betydning / Amount of options for selecting r elements from n elements when the order matters" (EMU.dk, 2020) which could have explained why technique 4 provides an unrealistic answer. However, reviewing the statement of the formula would depend on a realization that the order does not matter in this case.

In addition, none of the students argue why the applied technique does not provide the expected answer and neither of the students use arguments formulated in the mathematical register which is expected in a mathematical classroom discourse. The adidactical potential of this problem is relatively high as it provides an opportunity for the students to use mathematical arguments in order to explain their results in this formulation phase. This is particularly evident when BS<sub>3</sub> states that: "…. *I think, I don't know, it is like this, it's me who is having difficulties saying it.*" Which shows some of the language related problems: BS<sub>3</sub> believes this is the correct technique but struggles to find the right words to explain why this is the case. As a consequence of the fact that BS<sub>3</sub> struggles to formulate why the applied technique is correct the adidactical potential of the problem is not being fulfilled: BS<sub>1</sub>, BS<sub>3</sub> and BS<sub>4</sub> continues to work on problem 2 instead. This could reflect some terms of the didactical contract where the students expect the teacher to validate the applied techniques of the action phase.

Another interesting aspect of this observation is the fact that  $BS_1$ ,  $BS_3$  and  $BS_4$  work on problem 2 instead and do this quite successfully as they obtain the correct answers by applying correct techniques. The explanation of this could be found in the fact that problem 2 was categorized as an isomorphic problem in the a priori analysis and it does not contain the same amount of possible language related difficulties as problem 1 as there is not interpretation of the problem.

SITUATION 2: TWO BILINGUAL STUDENTS WHO STRUGGLES WITH FORMULATION

The students kept working adidactically and after a while BS<sub>2</sub> turned to BS<sub>3</sub> again and the following conversation occurred:

BS<sub>2</sub>: Det giver ikke mening med det her. Svaret kan ikke give fem tusind. Der kan ikke være fem tusind forskellige måder at komme syv spillere ind på en bane / *It does not make sense with this. The answer can not be five thousand. There can not be five thousand differint ways to put seven players into a field.* 

BS<sub>3</sub>: Jo / Yes

BS<sub>2</sub>: Nej, tror I der er fire hundrede spillere eller hvad? / *No do you think there are four hundred players or what*?

BS3: Nej, det er der ikke. Nej fordi at .. / No there is not. No because ..

BS<sub>2</sub>: Ja så giver det syv / Yes so it is seven

BS<sub>3</sub>: Nej fordi ... Nej.. Det kan ikke være syv. Der er seks.. / *No because ... No. It can not be seven. There are six ..* 

BS<sub>2</sub>: Ja så giver det syv. Der er en der er ude ad gangen / Yes so it is seven. There is one person who is out at the time.

BS<sub>3</sub>: Nej, fordi enhver af de der.. Nej, fordi på forskellige måder.. / *No because each of these … No because in different ways*...

#### FIGURE 11: PART OF THE TRANSCRIPTION OF OBSERVATION NO. 1 (APPENDIX 8)

Then  $BS_1$ ,  $BS_3$  and  $BS_4$  look at the assignment again, point to the text and mumbles. In this conversation some language related problems in the formulation phase occurs. In particular,  $BS_3$  is struggling to find the right words to describe, why the answer is not seven as suggested by  $BS_2$ . Particularly in the three phrases:

- 1) "Nej, det er der ikke. Nej fordi at .. / No there is not. No because .. "
- "Nej fordi ... Nej.. Det kan ikke være syv. Der er seks.. / No because ... No. It can not be seven. There are six .. "
- "Nej, fordi enhver af de der.. Nej, fordi på forskellige måder.. / No because each of these ... No because in different ways..."

In these three sentences BS<sub>3</sub> stops talking in the middle of it, and this could be an indication that BS<sub>3</sub> struggles to find the right words in order to properly argue why the application of technique 1 is correct. In the last sentence, the use of the phrase "... *på forskellige måder / ... in different ways*"

could indicate that this is the part of the formulation in problem 1, which is causing trouble for BS<sub>3</sub>. As the students had previously worked on the problem from the related PowerPoint slide no. 3 (appendix 5) which was displayed in figure 6 in the à priori analysis, it is possible that BS<sub>3</sub> have read problem 1 as isomorphic to part 4 of that problem, namely:

4. Ud af 26 elever i 2.b skal fire vælges til et trivselsudvalg. Hvor mange forskellige kombinationer af elever kan vælges? (Hjælpemidler tilladt) Out of 26 students in 2.b, four must be elected to a well-being committee. How many different combinations of students can be chosen? (Aids allowed)

Semantically, there is a difference between the formulations of the two problems in the usage of *"forskellige kombinationer / different combinations"* and "... *forskellige måder / ... different ways"*. The misconception made by BS<sub>3</sub> is thus based on a syntactical and semantical error in the sense that BS<sub>3</sub> views *"forskellige måder /different ways"* as a description of the placement of the players and therefore as different combinations and not as a description of the team in total.

Another aspect of the language related problems in this conversation, can be found with BS<sub>2</sub>, who is repeatedly stating that the answer is seven. In this formulation phase, there is no argumentation to why the answer is seven until the statement: "*Ja så giver det syv. Der er en der er ude ad gangen /Yes so it is seven. There is one person who is out at the time.*". It then becomes evident that BS<sub>2</sub> has applied technique 3 (as presented in the à priori analysis), but no mathematical arguments are being used in the conversation with BS<sub>2</sub>. The repetition by BS<sub>3</sub> stating that it is 7 could also indicate some language related difficulties in the formulation. BS<sub>3</sub> is capable of reading problem 1 as non-isomorphic to part 4 of the PowerPoint problem but is not capable of explaining why the answer is correct, and this was also the case in the conversation from situation 1 of the observation. BS<sub>3</sub> therefore has some different language related problems than BS<sub>2</sub> in the sense that BS<sub>3</sub> have read the problem without the semantical and syntactical error which could have occurred with BS<sub>2</sub>. However, they both struggle in the formulation phase and as a result neither of them are capable of understanding why the opposite person believe they are right.

## SITUATION 3: BILINGUAL STUDENT WHO STRUGGLES WITH FORMULATION

After having discussed the problem with several students, BS<sub>2</sub> chooses to validate the answer by teacher conformation. A didactical situation occurs, and the following conversation takes place:

BS<sub>2</sub>: T er der ikke kun syv forskellige måder? / T isn't there only seven different ways?

T: Hvorfor det?/ Why is that?

BS<sub>2</sub>: Fordi øh (griner). Du kan kun have nummer et ude én gang. Du kan kun have nummer to ude én gang. Du kan kun.. / *Because eh (laughs). You can only have number one out one time. You can only have number two out one time. You can only have ...* 

T: Vi kan jo godt vælge en vi sætter ude. Det giver glimrende mening. / We can choose one whom we set outside. That makes excellent sense.

## FIGURE 12: PART OF THE TRANSCRIPTION OF OBSERVATION NO. 1 (APPENDIX 8)

As BS<sub>2</sub> chooses to validate the answer by the use of the teacher could reflect aspects of the didactical contract. In particular, it could indicate that the students are not used to find strategies in order to validate the formulated hypothesis of the action phase. In this didactical situation, it is confirmed that BS<sub>2</sub> has applied technique 3 and the provided explanation by BS<sub>3</sub> is informal as it is a logical explanation as to why it makes sense that the answer should be seven. However, the answer is still somewhat repetitive and BS<sub>2</sub> does not formulate mathematically why the answer is correct. The answer is formulated in natural language and there is no use of the mathematical register as no mathematical representation has been chosen. Additionally, no exact statement is formulated. This situation have didactical potential in relation to the formulation in particular when the teacher asks: *"Hvorfor det?/ Why is that?"* thus forcing BS<sub>2</sub> to argue why the answer is correct. However, the potential is not being fulfilled as the teacher accepts the informal, repetitive formulation provided by BS<sub>2</sub>.

# CLASSROOM OBSERVATION NO. 2: THE TERM "RANDOM" AND EQUIPROBABILITY BIAS

The second observation selected for analysis happened in the second module (5/10-2020) where the teacher introduced the concept of a stochastic variable. While doing this, the teacher had written on the blackboard as seen in image 25 below:



IMAGE 25: WRITTEN EXPLANATION BY THE TEACHER OF "STOCHASTIC VARIABLE" ON THE BLACKBOARD And had thus written: "*stochastic variable"*  $\rightarrow$  "*random variable"*". The introduction to stochastic variables was made by the use of an example of the delay of a bus and how one could describe this as a stochastic variable. The students and the teacher created a probability table as displayed in image 26 and just above the table the teacher wrote: "X is the stochastic variable describing the number of minutes the bus is delayed"



IMAGE 26: A PROBABILITY TABLE DESCRIBING THE NUMBER OF MINUTES THE BUS IS DELAYED

After the table was completed a bilingual student and the teacher engaged in the following didactical situation which can be seen fully transcribed in appendix 9:

BS: Hvis det er tilfældigt hvorfor er det så ikke 25% chance for at .../ If it is random why isn't there a 25% chance to..

T: Det er fordi det ikke er et symmetrisk sandsynlighedsfelt / *That is because this is not a symmetrical probability field* 

BS: Jeg forstår bare ikke hvordan der kan være større chance for at der sker noget som er tilfældigt / *I just don't understand how there can be a greater chance of something happening which is random* 

T: Det er en meget meget øhm.. Det er en meget mærkelig terning vi er ude i. Det var også ligesom før, sandsynligheden for at læreren var syg og læreren var rask var ikke det samme. Selvom noget er tilfældigt kan det godt ske, så der er noget der er mere sandsynligt end noget andet. Sandsynligheden for at vinde i lotteri og for ikke at vinde i lotteri vil også sjældent være det samme / *It is a very uhm.. It 's a very weird dice we have. It is like before, the probability of the teacher being sick and the teacher being healthy was not the same. Even though something is random it can happen that something is more likely that something else. The probability of winning the lottery and not winning the lottery will rarely be the same.* 

# BS: Nåh ja ja / Yeah right

T: Så det er egentlig det at der er nogle forskellige sandsynligheder og hvis vi skulle gætte, hvis bare vi skulle komme med et bud på, hvad tror vi at den her variabel bliver. Så vil vi jo gætte på det er 11 for det er der jo trods alt 50% sandsynlighed for. / So, it is really the fact that there are different probabilities and if we had to guess, just give an estimate on what we believe this variable will be. Then we will guess on 11 because after all there is a 50% probability of this.

BS: Hvorfor er der det? Det er det jeg ikke forstår / Why is that? That is what I don't understand

T: Og det er det.. Det er det der er lidt svært ved det her for det er der ikke noget forklaring på. / *And this is it.. This is what is a little difficult about this as there is no explanation to this.* 

# FIGURE 13: TRANSCRIPTION OF OBSERVATION NO. 2 (APPENDIX 9)

In this didactical situation the bilingual student struggles to understand why  $P(X = x_i) \neq 0,25$  for all i if the outcome is "tilfældigt" (random) and this is particularly implied in the first sentence of the student: "*Hvis det er tilfældigt hvorfor er det så ikke 25% chance for at …*/ *If it is random why isn't there a 25% chance to*..". As reviewed in the theoretical framework, this type of misconception is not uncommon as it was pointed out by Batanero and Sanchez (2013). One of the tendencies within student errors in probability is to believe random events are "equiprobable" by their very nature. This observation is thus an example of equiprobability bias, i.e. "… people exhibiting this bias judge outcomes to be equally likely when their probabilities are not equal" (Batanero & Sanchez, 2013, p. 265).

The equiprobability bias could be based on the teacher's previous introduction to the concept of stochastic variables where the term stochastic variable was "translated" to random variable. It is possible that by giving this translation of stochastic variable, the student has made a connection between "tilfældig" (random) and the meaning of random as a description of the probability of each occurrence. However, in the translation random is intended to describe that the variable is randomly chosen in the sense that it could have been any variable. In this case, the equiprobability bias could therefore be based on syntactical error which is causing semantical problems as it changes the meaning of what is being communicated and what "tilfældig"(random) is meant to describe.

In the second line the teacher tries to explain mathematically why the probability is not evenly distributed by stating: "Det er fordi det ikke er et symmetrisk sandsynlighedsfelt / That is because this is not a symmetrical probability field". This formulation is stated in the mathematical register as it uses technical terms (asymmetric probability field), however, the bilingual student does not seem to accept (or understand) this mathematical explanation as he/she replies: "Jeg forstår bare ikke hvordan der kan være større chance for at der sker noget som er tilfældigt / *I just don't understand how there can be a greater chance of something happening which is random*". This could be regarded as an exemplification of the difficulties in the transition from natural language to formal language.

This could also indicate two things: 1) The bilingual student has not understood the concept of asymmetric probability fields, which in the mathematical register refers to the notion:  $P(X = x_i) \neq \frac{1}{n}$  for all i and n observations. Secondly, it could indicate that the student rely on the general danish conception of the word "tilfældig"(random) and this is what is causing trouble. In general danish the word "tilfældig" (random) can be used to describe different thing as it can both mean that something happens by a coincident and the chances of each occurrence is equal, however something can also be "tilfældigt"(random) with uneven probabilities: The semantics of the word can thus potentially cause some trouble.

This situation illustrated syntactical and semantical errors in the use of the word "tilfældigt"(random) and it could be caused by the translation made by the teacher in the beginning of the lesson. However, it is important to notice that this situation is not necessarily due to bilingualism: As pointed out by Batanero and Sanchez (2013), the equiprobability bias is a common mistake amongst students thus

not occurring exclusively with bilingual students. This aspect will be discussed further in the final discussion of the results.

# CLASSROOM OBSERVATION NO. 3: LANGUAGE RELATED DIFFICULTIES IN THE FORMULATION PHASE

This observation was made in the third module (21/10-2020) where the students had been introduced to the binomial distribution. The focus of this module was on the binomial distribution and how to make calculations such as the mean and the spread and they were introduced to formula (189) and (190) stating that:

$$\mu = n \cdot p$$

And

$$\sigma = \sqrt{n \cdot p \cdot (1 - p)}$$

Where n is the "antalsparameter" (number parameter) and p is the "sandsynlighedsparameter" (probability parameter). As a small exercise in the beginning of the modules the students had discussed the following assignment (the sleeping problem) in small groups before the teacher and the class discussed it:

I en klasse går der 20 elever, der hver har 5 % sandsynlighed for at sove over sig. Den binomialfordelte, stokastiske variabel *X* betegner antallet af elever, der sover over sig på en given dag.

In a class there are 20 students, who each has 5% probability of over sleeping. The binomially distributed, stochastic variable X denotes the number of students sleeping over on a given day.

- *a)* Bestem antalsparameteren og sandsynlighedsparameteren. Determine the number parameter and the probability parameter.
- *b)* Bestem middelværdien, og forklar, hvad middelværdien siger om eleverne. Determine the mean and explain what the mean says about the students
- c) Det oplyses, at P(X = 0) = 0,36. Hvad betyder dette? It is given that P(X = 0) = 0,36. What does this mean?

#### FIGURE 14: THE SLEEPING PROBLEM (APPENDIX 7)

In the following didactical situation, the teacher and the students of the class had just defined n and p in part a) and then continued to discuss part b). In the presented conversations,  $BS_1$ ,  $BS_2$ ,  $BS_3$  and  $BS_4$  were all bilingual students. The entire transcription of the following didactical situation can be found in appendix 10:

T: Ja det giver en. Lige præcis. Og hvordan skal det her tal så fortolkes, det er det næste. Nu ved vi det her, men hvad fortæller det os så om? / Yes it is one. Exactly. And how is this number to be interpreted, that is the next part. Now we know this, but what does it tell us?

BS<sub>3</sub>: Er det ikke bare at der en der har mulighed ... At der er en der kommer til at sove over sig. / *Is it not just that there is one who has the possibility* .. *That there is one who is going to oversleep* 

T: Der er en der kommer til at sove over sig? Ved vi det? At der er en der kommer til at sove over sig? / *There is one who is going to oversleep? Do we know this? That there is one who is going to oversleep?* 

BS<sub>3</sub>: Nej men sandsynligheden er (stopper med at tale) / No but the probability is (stops talking)

T: Sandsynligheden er at der er en der kommer til at sove over sig? Vi nærmer os. Vi er tæt på. / *The probability is that there is one who is going to oversleep? We are getting closer. We are close.* 

#### FIGURE 15: PART OF THE TRANSCRIPTION OF OBSERVATION NO. 3 (APPENDIX 10)

In this conversation, the bilingual student  $BS_3$  is struggling to find the right words to describe the intended meaning which is indicated by half-finished sentences and the fact that  $BS_3$  stops talking in the middle of a sentence. As presented in the a priori analysis this could be an indication of language related problems in the formulation phase. In the first sentence by  $BS_3$ : *"Er det ikke bare at der en der har mulighed ... At der er en der kommer til at sove over sig. / Is it not just that there is one who has the possibility .. That there is one who is going to oversleep"* some language related problems also occur. In the first part by the use of the word "mulighed" (possibility) indicates that there is a chance for a student to oversleep, but the second part of the sentence is formulated such that is a certain event. BS<sub>3</sub> therefore starts by formulating the mean as a description of a possible event and then changes it to be a certain event. The answer is not accepted by the teacher as it is not precise enough, in the sense that the calculated mean does not state that the event will happen. This was also reviewed in the a priori analysis and in particular this would be an answer similar to the third type of answer as this formulation indicates that the calculated value of the mean is always occurring,

In the second sentence by BS<sub>3</sub>: "*Nej men sandsynligheden er (stopper med at tale) / No but the probability is (stops talking)*" BS<sub>3</sub> experiences some language related problems again in the sense that he/she stops talking in the middle of the sentence which could indicate a struggle to find the right words to describe the intended meaning. By the use of the word "sandsynligheden" (the probability) and the previous sentence by BS<sub>3</sub>, it could be indicated that BS<sub>3</sub> is using the word "sandsynligheden" (the probability) to describe that something is very likely. The word is thus used as in the general danish definition of the word where sentences such as: "sandsynligheden er at det sker / the probability is that it will happen" means that something is very likely to happen. BS<sub>3</sub> therefore formulates the answer in natural language where the semantics is causing trouble in the sense that "sandsynligheden" (the probability) does not have the same meaning in the general danish conception as in the mathematical register. The semantics in the transition from natural language to formal language is therefore causing difficulties, as the answer stated by BS<sub>3</sub> is not precise enough to be considered as a correct answer.

The didactical situation continues as another bilingual student BS<sub>4</sub> answers as seen in the figure below:

BS<sub>4</sub>: Jeg er ikke sikker men det er en ud af 20, der sover over sig, er det ikke? / *I am not sure but it is one out of twenty who oversleeps is it not*?

T: Jamen problemet er at det kan vi jo ikke vide. Det her er tilfældigt. Der er bare hver af dem der har en sandsynlighed på 0,05 for at sove over sig. Denne her variable den kan ende i alle mulige [] det eneste vi ved er at den lander et sted mellem 0 og 20. Der er et sted mellem nul og tyve elever der sover over sig. / Well problem is that we can not know this. This is random. It is simply just each of them who has a probability of 0,05 to overslepp. This variable it can end in all sorts of [] the only thing we know is that it will end somewhere between zero and twenty. There is somewhere between zero and twenty students who oversleeps.

 $S_5$ : Middelværdien siger at der gennemsnitligt vil være en der sover over sig / *The mean is saying that averagely one will oversleep* 

T: Der vil gennemsnitligt være en der sover over sig. Vi kan ikke vide om der er en der kommer til at sove over sig, men der vil gennemsnitligt være en der sover over sig / Averagely there will be one who oversleeps. We can not know if there will be one who oversleeps but averagely there will be one who oversleeps.

#### FIGURE 16: PART OF THE TRANSCRIPTION OF OBSERVATION NO. 3 (APPENDIX 10)

In this didactical situation,  $BS_4$  states: "Jeg er ikke sikker men det er en ud af 20, der sover over sig, er det ikke? / I am not sure but it is one out of twenty who oversleeps is it not?" and this formulation is very close to the first formulation provided by  $BS_3$  and there could be several reasons to this. Firstly, it is possible that  $BS_4$  was not listening when  $BS_3$  stated the previous answer and therefore BS<sub>4</sub> does not know the answer is the same. Secondly, it could also be a language related problem in the sense that BS<sub>4</sub> does not realize that there is not essential difference in the answer "At der er en der kommer til at sove over sig / That there is one who is going to oversleep" and the answer: "... det er en ud af 20, der sover over sig, er det ikke? / ... it is one out of twenty who oversleeps is it not?". The language related problems which BS<sub>4</sub> experiences are thus different from the language related problems with BS<sub>3</sub> in the sense that BS<sub>3</sub> displayed indications of struggling to find the right words and BS<sub>4</sub> did not. It is interesting that both BS<sub>3</sub> and BS<sub>4</sub> are very close to providing a correct answer as they have both understood that in the calculation  $\mu = n \cdot p = 20 \cdot 0.05 = 1$  the value does describe that one student will be late.

The lack of precision in their formulations is the reason why the answer is not accepted by the teacher and this could therefore be seen as a language related problem. Interestingly, neither  $BS_3$  nor  $BS_4$ formulates that the calculated value is an average as neither of them uses the word "gennemsnitlig" (averagely) in their replies. In particular,  $BS_4$  does not even though he had just heard the reply by  $BS_3$ . The bilingual students all did very well when asked to calculate the mean and were capable of doing so correctly, but the troubles occurred in the formulation phases when they had to explain what the calculated value meant in this situation. CLASSROOM OBSERVATION NO. 4: THE USE OF CAS AND RELATED PROBLEMS

This observation (21/10-2020) was made in the third module and was in relation to the use of CAStools as an aide in the calculations regarding the binomial distribution. As introduced in the section "The use of CAS and computers" the students were using WordMat as a CAS-tool. The teacher started by showing the class how to use WordMat to calculate the mean, the spread and probabilities of binomially distributed stochastic variables by entering the number parameter and the probability parameter. Afterwards the students worked adidactically on the problems on worksheet 3 (appendix 4a/4b). This observation was in relation to problem 1 which is seen in shown in figure 18 just below.

# Opgave 1 (med hjælpemidler) Problem 1 (aids allowed)

X er en binomialfordelt stokastisk variabel med antalsparameter 80 og sandsynlighedsparameter p = 0,35.

*X* is a binomially distributed stochastic variable with number parameter 80 and probability parameter p = 0.35.

- a) Tegn et søjlediagram for sandsynlighedsfordelingen for *X*. *Make a bar chart of the probability distribution for X*.
- b) Bestem middelværdien og spredningen for *X*. Determine the mean and spread/standard deviation of *X*.
- c) Bestem P(X = 28). Determine P(X = 28).

# FIGURE 17: PROBLEM 1 OF WORKSHEET 3 (APEENDIX 4A/4B)

As argued in the a priori analysis, all three parts of the problems were solvable by the use of WordMat. One bilingual student,  $BS_1$ , was having trouble getting started with the problem and as a strategy to get started,  $BS_1$  had help from another bilingual student,  $BS_3$ . The conversation can be seen in figure 18 below: BS<sub>1</sub>: Hallo kan vi ikke lave noget, wallah, jeg forstår det ikke / *Hallo can we not get started, wallah, I don't understand* 

BS<sub>2</sub>: Hvad er den? / What is it?

BS<sub>1</sub>: Jeg kan ikke finde ud af (stopper med at tale) / I don't know how to (stops talking)

Eleverne taler om andre ting / The students discuss other things

BS<sub>2</sub>: Hvordan gør vi det? Okay vent. Skal man have wordmat? / How do we do it? Okay wait. Do I need wordmat?

BS<sub>3</sub>: Må jeg lige se? (vender computeren og åbner op for Excelarket). Det er det her ikke. Det er bare det der vi lavede / *Can I see*? (*Turns the computer around and opens the Excel sheet*). It is this one right? It is just what we did.

BS1: Så jeg skal skrive 80 og det andet? / So I have to write 80 and the other one?

BS<sub>3</sub>: Ja (peger på felterne) / Yes (points to the boxes)

BS<sub>1</sub>: Og så tager jeg et billede af den? (Tager et screenshot og indsætter i dokumentet) / And then I take a picture of it? (Makes a screenshot and inserts in the document)

#### FIGURE 18: TRANSCRIPTION OF OBSERVATION NO. 4 (APPENDIX 11)

This adidactical situation can be seen as an action phase, and the communication between the students is very limited. This can be regarded as a language related problem in the sense that BS<sub>1</sub> is having difficulties expressing what is difficult in the situation. In particular, BS<sub>1</sub> states: "*Jeg kan ikke finde ud af (stopper med at tale) / I don't know how to (stops talking)*" and thus stops talking in the middle of the sentence. BS<sub>1</sub> is therefore not capable of expressing exactly what is not understood. This aspect was also pointed out in the theoretical framework were Andersen (2004, p. 286) argued that bilingual students are in particular having difficulties when engaging in collaboration with other students where it is a perquisition that they are to interpret what others are saying. As BS<sub>1</sub> is struggling to express what is difficult it could be more difficult for the other students to provide any help. It is also unclear if BS<sub>1</sub> does not know how to use the CAS-tool or if BS<sub>1</sub> does not understand how the problem is to be solved.

In this case,  $BS_3$  is capable of helping  $BS_1$  but the communication by  $BS_3$  is also very limited in the sense that  $BS_3$  points to the places of where  $BS_1$  is supposed to enter the values. In the formulation phase between the two bilingual students there is no use of relevant technical terms such as "antal-sparameter" (number parameter) or "sandsynlighedsparameter" (probability parameter): Instead

formulations such as "*det der / that*" and "*det andet /the other thing*" and gesticulations are used to describe how the technique is applied.

By the help of  $BS_3$ ,  $BS_1$  completed part a) of problem 1 and obtained the correct illustration of the bar chart, where a reconstruction has been made in image 27 just below:



IMAGE 27: SOLUTION MADE BY WORD MAT TO PROBLEN 1A FROM WORKSHEET 3

BS<sub>1</sub> then started on a solution for part b) and quietly wrote down: "80\*0,35" in the document. BS<sub>1</sub> therefore applied the formula to calculate the mean where  $\mu = n \cdot p$  which is a correct strategy for a solution. BS<sub>1</sub> therefore chooses to calculate the mean,  $\mu$ , by hand even though the student actually already had the answer to b) contained in the screenshot from a) (see image 27). This might indicate that in solving part a) of the problem BS<sub>1</sub> has correctly identified that the first part of the problem is isomorphic to what the teacher had just reviewed, but perhaps BS<sub>1</sub> does not fully understand the method. In the conversation BS<sub>1</sub> also expresses a lack of understanding by stating: *"I don't understand"* and *"I don't know how to"* and BS<sub>3</sub> points to the exact places where BS<sub>1</sub> has to fill in the exact numbers. Combined, these things contribute to the interpretation of BS<sub>1</sub> following a method without fully understanding what is being calculated.

Another aspect of this situation is in relation to the symbolic register, as BS<sub>1</sub> does not write down any symbolics in regard to the calculations and simply choses to write down the numbers. In addition, by

calculating this by hand even though the answer is provided in the screenshot it could indicate that BS<sub>1</sub> does not fully understand the symbolics involved ( $\mu$  and  $\sigma$ ) and what they represent: Otherwise, the student probably would not have started calculating these values by hand.

The use of a CAS-tool in this case is thus an aide in the sense, that  $BS_1$  was capable of providing a solution to part a of the problem and the solution was obtained quickly and without having to put in a lot of work. However, the use of CAS in this case also made it possible to obtain a solution of a problem without fully understanding the underlying mathematical technique. In this case it was indicated that  $BS_1$  did not fully understand the process and the involved symbolics.

### A PARTIAL CONCLUSION

In the presented observations, different types of language related problems with the bilingual students occurred and it was particular in the formulation phase the problems occurred. In observation no. 1, BS<sub>2</sub> and BS<sub>3</sub> struggled to explain why the respective answers were correct and both students were incapable of arguing why they were correct. BS<sub>3</sub> struggled particularly with finding the right words to describe the intended meaning, whereas BS<sub>2</sub> kept repeating that the answer was correct. In observation no. 2 a bilingual student struggled to understand why the term random did mean that the stochastic variable had an equal probability and as mentioned this could either be a language related problem caused by semantics and syntax or an example of the equiprobability bias. In observation no. 3 the language related difficulties with BS<sub>3</sub> and BS<sub>4</sub> occurred in the formulation phase as the lack of precision in their formulations made the answers incorrect. In particular, it was also noted that BS4 did not realize that the answer provided by BS<sub>3</sub> was almost isomorphic to the answer BS<sub>4</sub> provided. In observation no. 4 BS1 and BS3 worked adidactically on a problem by the use of CAS-tools, and the conversation between the two bilingual students was very limited: This can also be reviewed as a language related problem in the sense that they both had difficulties articulating what was difficult in the situation and how it could be solved. We have thus seen several cases of language related problems in particular in the formulation phase and in the transition from natural language to formal language as the importance of precision became evident.

The data and the provided analysis will be further discussed in the next section.

# DISCUSSION

In this section there will be a discussion of the results obtained from the presented and analyzed data of the previous sections. The discussion will focus on which tendencies the data showed and if these tendencies can be seen as representatives of the language related problems occurring with learners whose first language is not the language of instruction or if the difficulties simply lie with the individual. The methodological aspects of the obtained data will also be discussed, in particular the discussion will include both quantitative and qualitative aspects of the obtained data and the formulation of the questions in the questionnaire will also be discussed. In addition, the aspect of reading in mathematics will also be involved in the discussion as the students had to read questions and problems in the obtained data. Finally, suggestions to improve bilingual students' performance in the context of probability and statistics will be presented based on the presented theoretical framework and the findings of the case studies.

# WHAT DID THE DATA SHOW?

The obtained data in the context of both statistics and probability showed several cases of language related problems with the bilingual students and some of the language related problems occurred repeatedly with several bilingual students. These language related problems included syntactical errors, semantical errors, and difficulties in the formulation phase. In the theoretical framework, it was presented that some of these difficulties were typical errors amongst all types of students and the findings of the analysis will therefore be discussed. In particular, it will be discussed what we are able to conclude from these findings.

# DIFFICULTIES IN THE FORMULATION PHASE

In the context of both probability and statistics, the data showed several incidents of difficulties in the formulation phase with the bilingual students. This was the case with BS<sub>2</sub> (in relation to "spredning" (spread)), BS<sub>5</sub> (in relation to "hyppighed" (frequency)), BS<sub>8</sub> (in relation to "hyppighed" (frequency)), in observation no. 1, and in observation no. 3. The language related problems in the formulation phase were therefore present both in relation to a written and oral media.

Observation no. 1 (situation 1, 2 and 3) and observation no. 3 were similar in the sense that the involved bilingual students all struggled in the formulation phase. The struggles in the formulation phase varied slightly: In observation 1, BS<sub>3</sub> struggled to find the right words whereas BS<sub>2</sub> had difficulties formulating why the answer was correct and instead kept repeating that the answer was correct

and in observation no. 3, BS<sub>3</sub> struggled to find the right words to explain the intended and the lack of precision in the formulation made the answer incorrect. In observation no. 3, BS4 repeated the answer from BS<sub>3</sub> perhaps due to the fact that BS<sub>3</sub> did not realize there was a difference in the two formulations. As pointed out in the theoretical framework by Andersen (2004, p. 286) these language related problems in the formulation phase can be an obstacle for the students both in didactical and adidactical situations and these observations exemplify this as both observation no.1 and no.3 showed bilingual students having difficulties when engaging in classroom discussions or collaboration with other students where it is a perquisition that they are to interpret what others are saying. This was also supported by Andersen (2004, p. 286) and the observations may confirm this hypothesis. As argued by Dunn et al (2016, p. 9-10) in the theoretical framework, some of the challenges, which students face if they do not fully grasp the concepts and the use of the related vocabulary include: The content, seeking assistance, and group work. These challenges were not exclusive to bilingual students, and another possible explanation for the struggle in the formulation phase therefore also lie in the fact that the students do not fully master the vocabulary of probability yet. More observations and similar investigations are therefore needed in order to obtain a final conclusion, as only a relatively low amount of observations was provided. By expanding the amount of observations to including monolingual students as well, it would provide a comparative foundation of the observations made. Due to the qualitative design of the study it cannot be generalized with certainty to a larger population that all bilingual students will experience these language related difficulties in the formulation phase, as other variables may affect the outcome as the data has only been obtained from one High School. However, as the theoretical framework, particularly Andersen (2004), supports this postulate, the findings may be regarded as representative.

Additionally, similar observations were also made in the written answers to the questionnaire regarding statistics where with BS<sub>2</sub> (in relation to "spredning" (spread)), BS<sub>5</sub> (in relation to "hyppighed" (frequency)) and BS<sub>8</sub> (in relation to "hyppighed" (frequency)) all formulated answers which made it difficult to extract a meaning. In particular, the formulation of the mathematical definition of spread provided by BS<sub>2</sub> contained both semantical and syntactical errors making the sentence grammatically incorrect and therefore very difficult to extract a meaning. In the case of BS<sub>8</sub> (in relation to "hyppighed" (frequency)) the formulation also contained either semantical or syntactical errors and an ambiguous mathematical definition of the word was provided. The case of BS<sub>8</sub> has some similarities with observation no. 3 where the imprecise use of language also made the answer incorrect. The data obtained in statistics therefore shows a tendency of semantical and syntactical errors in the formulation to such a degree that the formulated sentence becomes difficult to understand.

All five cases showed difficulties in the formulation and in particular the lack of precision in the formulation made the communicated incorrect. In formulation phases the shape of the statement also becomes important as reviewed in the theoretical framework, as it becomes a matter of syntax in the sense that students are expected to compose and combine different representations in a way which is formally correct (Winsløw, 2016, p. 160). These cases all exemplify this, and the monolingual student's replies did not show any similar problems where the syntax made the provided answers difficult to understand. Only in the case of  $MS_2$  (in relation to "frekvens"(relative frequency)), the formulation of the mathematical definition of the word was ambiguous in the sense that the quotient described linguistically could be interpreted in different ways. This syntactical error could have been avoided by using a mathematical representation to describe the intended quotient, thus emphasizing the importance of the formal language of mathematics.

The observations show some of the difficulties experienced by both mono- and bilingual students in the transition from everyday language to formal language. In the theoretical framework it was presented by Bitterlich and Schütte (2018, p. 36) that "...*regarding the discourse practices explanation and argumentation, that many learners do not gather enough language-based experience in family and peer groups prior to entering school. But often, such language competences in explaining and arguing are assumed"* and this aspect is particularly important in relation to the bilingual students who did not speak danish at home. It is therefore supported by the previously conducted research, that the language related difficulties in relation to semantics, syntax, and formulation is enforced by bilingualism.

# THE ASPECT OF READING

In the theoretical framework, it was presented that it seemed as if there is a high correlation between the reading skills of bilingual students and their advancement in mathematics (Dale & Cuevas, 1987, p. 24). Hence, it is also relevant to discuss the aspect of reading the questions of the questionnaire and the reading of the problems in the observations and to discuss if this factor also contributes to the language related difficulties experienced by the bilingual students in the obtained data.

In the questionnaire the students were asked to "define" the chosen lexically ambiguous words, i.e., to explain and describe the meaning and exact limits of something. In several cases, the bilingual

students, when asked to define a concept, instead chose to explain the usage of the word in the previously used sentence or relied on a new example to explain the meaning. This was for example the case with  $BS_4$  and  $BS_5$  in relation to "spredning" (spread),  $BS_7$  in relation to "hyppighed" (frequency) and  $BS_6$  in relation to "frekvens" (relative frequency). The provided definitions were not definitions per say and it would not be a correct answer to the question: The language related problems therefore also lied partially in the reading of the questions, as the content provided in the definitions were not a general definition. As a consequence of not understanding the concept of a definition, the formal language of mathematics and the required precision in the formulation were not used.

In observation no. 1 (part 1 and 2), the bilingual student BS<sub>3</sub> had chosen an incorrect technique to solve the problem, and as presented in the theoretical framework, a lot of bilingual students are having difficulties acquiring the information and as a consequence the bilingual students often are having difficulties in solving word problems in mathematics (Andersen, 2004, p. 286). Observation no. 1 (part 1 and 2) could therefore be viewed as representative of this particular problem. In addition, it was empathized in the theoretical framework that if the context or formulation of a problem is slightly changed from the isomorphic problems, it will become evident that only a small amount of semantics is behind these syntactical skills (Winsløw, 2016, p. 161). Observation no. 1 (part 1 and 2) can be viewed as a confirmation of this, as BS<sub>3</sub> read the problem as isomorphic to the previously calculated problem.

However, in the theoretical framework it was also presented that some of the common errors made by students included for example " ... over-counting and confusing situations in which order matters or does not matter " (Lockwood, Swinyard & Caughmann, 2015, p. 29). The wrongfully applied technique by BS<sub>3</sub> could also be regarded as an example of confusing a situation in which the order matters. This is a common mistake happening amongst all types of students and not just bilingual students and it can therefore be argued that this type of error could also happen to a monolingual student.

# A COMBINATION OF SEMANTICAL AND SYNTACTICAL ERRORS

In the theoretical framework it was also presented that the misunderstandings made by students is often a combination of semantics and syntax and it can be difficult to separate the semantical difficulties from the syntactical difficulties (Winsløw, 2016, p. 160). This was also seen in the obtained data for example in the case of  $BS_8$  (in relation to "hyppighed" (relative frequency)) and in observation no. 2. Observation no. 2 illustrated syntactical and semantical errors in the use of the word "tilfældigt" (random), and it could be caused by the translation made by the teacher in the beginning

of the lesson. As argued by Andersen (2004, p. 286) bilingual students are in particular having difficulties when engaging in classroom discussions where it is a perquisition that they are to interpret what others are saying as a lot of the bilingual students are having difficulties acquiring the presented information. Observation no. 2 can be viewed as a confirmation of this and the misconception by the bilingual student will then be a language related problem experienced in particular by bilingual students. However, it was also presented by Batanero & Sanchez (2013) that a tendency amongst students is to believe random events are "equiprobable" by their very nature. The equiprobability bias, i.e. "... people exhibiting this bias judge outcomes to be equally likely when their probabilities are not equal" (Batanero & Sanchez, 2013, p. 265)" is therefore a common error amongst all types of students and not just bilingual students. It is thus possible, that this misconception by the bilingual student is not a consequence of being bilingual but a consequence of the common equiprobability bias. As the previously argued, the observations did not have a control group with similar/unsimilar observations amongst monolingual students, and it can not be definitively concluded that the misconception is a consequence of being the student being bilingual.

## SEMANTICAL DIFFICULTIES

The obtained data also showed several cases of semantical difficulties in the transition from natural language to formal language, for example in the case of  $BS_1$  (in relation to "spredning" (spread)),  $BS_2$  (in relation to "spredning" (spread)),  $BS_8$  (in relation to "hyppighed" (frequency)),  $BS_1$  (in relation to "hyppighed" (frequency)),  $BS_6$  (in relation to "frekvens" (relative frequency)), observation no. 2, and observation no. 4. The data thus showed 7 cases of semantical difficulties for the bilingual students and the types of semantical difficulties were similar in some of the cases.

In the theoretical framework it was presented that students tend to struggle to learn "... *the technical definitions of lexically ambiguous words and often retain the GE definitions of these words*" (Dunn et al, 2016, p. 11) and that students tend to face statistics with "... *strongly-held, but incorrect, intuitions that are highly resistant to change*" (Kaplan et al, 2010, p. 2). The case of MS<sub>1</sub> in relation to "spredning" (spread) showed this, as the general danish definition of spread was transferred to the mathematical definition of the word. It was also argued by Dunn et al (2016) that learners whose first language was not the language of instruction would more successfully adapt the new language of statistics as they would not rely on the general danish conception of the word. The case of BS<sub>1</sub> (in relation to "spredning" (spread)) did support this claim as the wrongful general danish conception of spread was not transferred to the mathematical definition, which was, nevertheless, imprecise. In the case of  $BS_2$  (in relation to "spredning" (spread)) and  $BS_6$  (in relation to "frekvens" (relative frequency)), the students also had a wrongful general danish conception of the word in question and the wrongful danish conception of the word was transferred to the mathematical definition provided. These cases did therefore not support the claim by Dunn et al (2016). The cases showed bilingual students having a conception of the word, which were not in accordance with the lexical meaning of the word, and the same thing was found by Kazima (2006). She found that a lot of the students" "... *preconceived meanings for probability vocabulary were distant from established conventional meanings*" (Kazima, 2006, p. 169).

In observation no. 2, the bilingual student struggled with the semantics of "tilfældig" (random) as the student relied on the meaning provided by everyday language, and as previously discussed this could be regarded as either semantical problems or as a consequence of the equiprobability bias. Molnar (2018, p. 28) presented the lexical ambiguity of the term random and the lexically ambiguous words presented a challenge in the transition from natural language to formal language for all types of students. The difficulties experienced by the bilingual student in observation no. 2, either if it is regarded as a semantical problem or as the equiprobability bias, is therefore not exclusively occurring with bilingual students as similar difficulties are found with monolingual students in the theoretical framework. It could be argued, that due to the interpretation necessary in the classroom discussion, the experienced difficulties could be enforced by the student being bilingual as Andersen (2004), presented that bilingual students are having significant difficulties in this context.

In observation no. 4 the bilingual student also experienced semantical difficulties in the transition from everyday language to a formal language. In particular, it was argued that it could be difficulties with the symbolical register in the context of probability, and the presented observations show different types of semantical difficulties as they are rooted in different registers, and the semantical difficulties is therefore present in the transfer between everyday language, the mathematical register, and the symbolical register. It is important to note, as already argued in the analysis of observation no. 4, that there could be several other explanations to the actions of the bilingual student in the case than difficulties with the symbolical register.

# GENERAL DISCUSSION OF THE RESULTS

The obtained data in the context of statistics and probability shows several incidents of semantical and syntactical errors, and an adjacent question is: Are these semantical and syntactical difficulties caused by the students being bilingual? The data did show these tendencies and it was supported by the theoretical framework that some of these difficulties may be enforced by the students being bilingual. It is, however, also important to notice for example BS<sub>3</sub> in relation to statistics who generally did well in the questionnaire with precise and correct formulations which showed no indications of semantical or syntactical errors, even though BS<sub>3</sub> only relied on the mathematical meaning of the word and exclusively provided sentences and definitions in a mathematical context. In addition, the two monolingual students also displayed some semantical errors in relation to the word "frekvens" (relative frequency) and MS<sub>2</sub> generally relied on the mathematical meaning of the semantical and syntactical errors which occurred with the bilingual students therefore might not exclusively be due to the fact that they are bilingual: As presented in the theoretical framework, students generally struggle in the context of probability and statistics (Undervisningsministeriet, 2018: Batanero et al, 2006) as the subjects contains challenges in relation to both the students and the teacher. It is possible that some of the semantical and syntactical errors which occurred are simply occurring because statistics and probability is generally difficult.

In the observations made and the answers provided to the questionnaire another relevant question is: Which of the difficulties are due to language and which are due to a lack of knowledge? In the collection of the data, some of the difficulties could also be caused by a lack of knowledge in the sense that it becomes relatively more difficult to solve a mathematical problem or define a mathematical concept if you cannot remember the involved mathematics. It is therefore a possible factor which could contribute to some of the difficulties experienced by the bilingual students. As the design of the collection of the data is qualitative and the amount of data is relatively low, more research is necessary in order to determine if the difficulties are due to bilingualism or a lack of knowledge. In particular, it will be relevant to do similar investigations and observations at other danish High Schools as well. The qualitative and quantitative aspects of the data will be further discussed at p. 105.

# THE RESULTS IN STATISTICS

In the analysis of the answers to the questionnaire we saw several cases of semantical and syntactical errors in the bilingual students reply and this was seen in relation to both "spredning" (spread), "hyppighed" (frequency) and "frekvens" (relative frequency) as reviewed. The obtained data in relation to statistics was provided through a written medium as the students were asked to write down

their answers and this was done in accordance with the previously described methodology, which was chosen in relation to the studies conducted by Kaplan, Fischer and Rogness (2010) and Díaz and Batanero (2009). The collection of data through a written media has several advantages as the data is collected easily and it is possible to get a lot of responses to the same question. Additionally, it made it possible to compare the bilingual student's replies to the replies of the monolingual students. As described in the theoretical framework, it is impossible to separate the mathematical content from the mathematical representations and both are a part of the language used within mathematics (Winsløw, 2006, p. 159) and both are essential when communication through written media. The writing of mathematics is an essential part of the mathematical classroom discourse and the requirements on the formulation in the mathematical register are also to be met in written media. However, in the analysis of the obtained data it was sometimes difficult to retract a meaning of an answer: In particular when the answer had syntactical, semantical, and grammatical errors. It could be argued that in order to obtain a more thorough analysis, a follow-up interview with the students should be conducted. This would allow some of the students to further explain some of their answers to the questionnaire, such that the answers would become easier to interpret and the language related difficulties experienced by the students would be further supported through the interview. In further investigation, it is therefore suggested that the questionnaires are supported by a semi-structured qualitative interview where the students use and define the selected lexically ambiguous words in natural language and in a mathematical context.

# SUGGESTIONS TO OBTAIN BETTER ANSWERS IN RELATION TO INDEPENDENCE

As reviewed in the analysis of the questionnaire in relation to the notion of independence, neither of the students provided a mathematical definition related to statistics and the answers lied outside the scope of this thesis. For further research it is suggested that the question regarding "uafhængighed" (independence) is changed such that it is clear to the students that they are to provide a meaning of "uafhængighed" (independence) in the statistical meaning of the word. This could be done by altering question 4 of appendix 1A by adding the two following questions:

- E. Skriv en sætning med ordet "uafhængig" som du kender det fra statistik eller sandsynlighedsregning i Gymnasiet / Write a sentence with the word "uafhængig" with the meaning of statistics or probability as you know it from High School
- F. Giv en definition af ordet "uafhængig" med den mening det havde i E / Provide a definition of the word "uafhængig" maintaining the meaning of E

By adding these questions, the students would be "forced" to provide a statistical meaning of the word and it would then be possible to analyze the answers more properly than the analysis provided in this thesis. The notion of independence is essential within statistics and probability theory, and as presented by Molnar (2018) misconceptions arise in relation to this notion and one potential reason for this trouble is the lexical ambiguity as independence has different meanings in probability, everyday language, algebra and statistics. In particular the concept is symmetrical within probability, whereas in everyday language, independent and dependent are not necessarily symmetric. Within algebra, the notion of independent and dependent variables is never symmetric *"... although closer to the sometimes-symmetric everyday definition than the always-symmetric probability version"* (Molnar, 2018, p. 31). Finally, in statistics the definition of independence is "... *never symmetric, similar to mathematical variables but opposite of probability events"* (Molnar, 2018, p. 31). Being lexically ambiguous, the notion of independence therefore poses possible challenges in relation to both statistics and probability and further investigations is suggested because of this.

Another possible way of obtaining answers in relation to the students' conception of "uafhængighed" (independence) could be through a semi structured interview containing a problem in relation to independence for 2-3 students to solve. In order to investigate the possible language related problems for bilingual students, a control group with monolingual students could be interviewed as well such that a comparative analysis could be conducted.

#### DISCUSSION OF RESULTS IN PROBABILITY

The observations made in the context of probability, were all made with a focus on the bilingual students of the class and the data exclusively contained cases concerning bilingual students. As a consequence of this, there is no control group to the data in the sense that no similar observations with monolingual students were conducted in this thesis. However, as presented in the theoretical framework some of the difficulties experienced by the bilingual students, for example in observation no. 1 and no. 2, has been described in the research literature. In observation no. 2, where a bilingual student struggled to understand the term random in relation to an asymmetric probability field, this type of problem was also introduced in the theoretical framework as an equiprobability bias. It is a common problem amongst students thus it might be argued that this particular difficulty was not caused by the fact that the student was bilingual.

In some of the observations it can be difficult to conclude that the difficulties are caused exclusively by bilingualism as several other factors may contribute to the experienced difficulties. As previously argued, it can be difficult to judge if the difficulties occur due to language related problems or due to a lack of knowledge. However, as bilingual students are in particular having difficulties when engaging in classroom discussions or collaboration with other students where it is a perquisition that they are to interpret what others are saying (Andersen, 2004, p. 286), it is not unreasonable to assume that some of the difficulties may be enforced by bilingualism. In particular, in a conversation between two bilingual students where the intended meaning has to go through two "filters" of interpretation some of the original meaning might get lost or misinterpreted along the way. This argument can be supported by observation no. 1 where BS<sub>2</sub> and BS<sub>3</sub> both had difficulties communicating, why they were right. As presented in the theoretical framework, probability is in general considered a difficult subject to teach and learn. This was also argued in the report from the Ministry of Education (2019, p. 5), the students generally had difficulties in the context of probability in the summer exams of 2019. Some of the difficulties in the context of probability therefore might be a more general problem amongst students.

## QUANTITATIVE VS QUALITATIVE DATA

The data was collected in two Danish High Schools (Frederiksberg Gymnasium and H.C. Ørsted Gymnasiet, Frederiksberg) in Copenhagen. Both High Schools have a relatively high amount of bilingual students as H.C. Ørsted Gymnasium (Frederiksberg) is a part of TEC (Technical Education Copenhagen) where in 2018 50.4% of the students had an ethnic origin different than danish (Danmarks Statistik, 2018) and at Frederiksberg Gymnasium 37.4% of the students had an ethnic origin different than danish (Danmarks Statistik, 2018). The chosen High Schools for the investigation therefore have a relatively high amount of bilingual students which in particular was evident in the answering of the questionnaire where only 2 of 17 students were monolingual. Consequently, the obtained data has a low amount of monolingual student replies to provide a comparative foundation and the obtained data should be regarded with this in mind. In order to obtain a more representative foundation for comparison it would be necessary to obtain more monolingual student replies. Additionally, as the studies were only conducted in two different classes of two High Schools it is possible that other factors have contributed to the findings of the data. It is therefore suggested that similar observations and investigations are conducted at other danish High Schools to support or refute the findings.

The design of the investigation was thus mainly qualitative which makes it difficult to generalize with certainty to a larger population, as other variables may affect the outcome. However, as previously
argued, the research literature presented in the theoretical framework supports some of the findings in the data. In particular, it was argued that some of the semantical and syntactical difficulties may be enforced by bilingualism as seen in both the student conversation and the written data.

#### SUGGESTIONS TO IMPROVE BILINGUAL STUDENT PERFORMANCE IN THE CONTEXT OF STATISTICS AND PROBABILITY

Throughout the thesis, it has been argued that language and mathematics cannot be separated and the suggestions to improve bilingual student performance in the context of probability and statistics are also centered around language. In the investigation on Malawian students, whose first language was not the language of instruction, Kazima (2006, p. 169) found that a lot of the students' "... preconceived meanings for probability vocabulary were distant from established conventional meanings" (Kazima, 2006, p. 169) and some of these meanings were rooted in the students' first language. It was recommended by Kazima that the students were offered something more than teacher-spoken definitions, "instances which would help them refine their construction of meanings for the vocabulary" (Kazima, 2006, p. 187) and how the words are appropriately used. The same is recommended in the case of bilingual students in Danish High Schools, as the data also showed incidents of students who experienced semantical difficulties with some of the vocabulary both in the everyday meaning of the word and the mathematical meaning of the word.

It was also recommended by Andersen (2004, p. 289), that when a new subject is introduced, the teacher should make up a list of the words that may be new to some or all of the students and as the words appear during the lessons, they should be explained by students who have understood them. The same thing could be suggested to improve bilingual student performance in the context of probability and statistics, as this could increase the focus on lexically ambiguous words, which the data and the theoretical framework presented as difficult. An inquiry-based introduction to statistics and probability is also suggested, which could be organized by the use of TDS. The idea behind TDS is that the intellectual work of the students must be more like the scientific activity as "... knowing mathematics is not simply learning definitions and theorems in order to recognize when to use and apply them." (Brousseau, 2002, p. 21). The process of inquiry is considered essential if the student is supposed to gain knowledge of these results and convince herself/himself of their validity of them, and the mathematical knowledge is thus established from the students experience and reasoning instead of being presented as a matter of fact. This would provide a deeper understanding of the statistical and probabilistic concepts, and additionally mathematics would generally become more rational

and challenging but also a more satisfying activity in which the students are more likely to engage in (Jessen & Winsløw, 2017, p. 36).

Finally, in relation to probability teaching it has been suggested that "... we should not only present different models and show their applications, but we have to go deeper into wider questions, consisting of how to obtain knowledge from data, why a model is suitable, how to help students develop correct intuitions in this field and deal with controversial ideas, such as randomness or causality." (Batanero, Godino & Roa, 2004, p. 2). A problem is the fact that few teachers have prior experience with conducting probability experiments or simulations. It could propose a challenge to implement an experimental approach to teaching probability or teaching through statistical investigation if the teacher has no training in this area (Batanero, 2009, p. 7). A focus on proper education of teachers is therefore also suggested as this could improve the understanding of concepts within probability not only for bilingual students but for all students involved. In the aspect of teacher education, it is also essential that the teachers develop competences such that they are capable of dealing with bilingual students. This development of competences includes for example different didactical aspects of the subjects being taught, as only few of the teachers have the insight needed to become aware of and meet the academic and social needs of many bilingual students. (Danmarks Evalueringsinstitut, 2006, p. 10).

#### CONCLUSION

The case studies presented in this thesis shows several cases of semantical and syntactical difficulties experienced by the bilingual students involved. In particular, the case studies showed difficulties in the transition from natural language to formal language where the semantics caused trouble as the students relied on the general danish conception of the words "spredning" (spread), "hyppighed" (frequency) and "frekvens" (relative frequency). The bilingual students also experienced difficulties in interpreting information both in classroom discussions and in group work and as a consequence of this language related problem, the bilingual students experienced difficulties in the formulation phases, as they struggled with the language of mathematics and finding the right words to describe the intended meaning. Some of these language related problems were also experienced by monolingual students and has been supported by the research literature. The thesis does therefore not conclude that language related problems in the context of probability and statistics occur exclusively with bilingual students. However, it would appear that some of the language related difficulties are enforced due to bilingualism as argued in the discussion and supported by the research literature. The qualitative design of the study was discussed, and it was suggested that more data was provided in order to generalize the results further.

#### REFERENCES

Andersen, M. W. (2004). *Dansk som andetsprog i matematikundervisningen*. In Democracy and Participation - A Challenge for Special Needs Educations in Mathematics, Örebro University, p. 283 -300

Barton, B; Chan, R; King, C; Neville-Bartin, P; Sneddin, J. (2005). *EAL undergraduates learning mathematics*. International Journal of Mathematical Education in Science and Technology, Vol. 36, No. 7, 2005, p. 721–729

Batanero, C., Chernoff, E. J., Engel, J., Lee, H. S. & Sánchez, E. 2016. *Research on Teaching and Learning Probability*. Springer Open

Batanero, C., Godino, J. D., & Navarro-Pelayo, V. (1997). *Effect of the implicit combinatorial model on combinatorial reasoning in secondary school pupils*. Educational Studies in Mathematics, 32, 181-199.

Batanero, C., Godino, J. D. & Roa, R. (2004). *Training Teachers to Teach Probability*, Journal of Statistics Education Volume 12, Number 1

Batanero, C. & Sanchez, E. (2013). *What is the nature of High School students' conceptions and misconceptions about probability?*. In Graham A. Jones (ed.), Exploring probability in school: Challenges for teaching and learning, 260—289. 2013 Kluwer Academic Publishers

Bitterlich, E. & Schütte, M. (2018). *Language use in different situations in the mathematics classroom: Everyday academic language and mathematical discourse*. In N. Planas & Schütte, M. (Eds.), Proceedings of the Fourth ERME Topic Conference 'Classroom-based research on mathematics and language' (pp. 34-40). Dresden, Germany: Technical University of Dresden / ERME.

Brousseau, G. (2002). THEORY OF DIDACTICAL SITUATIONS IN MATHEMATICS - DI-DACTIQUE DES MATHÉMATIQUES, 1970–1990. Kluwer Academic Publishers (Edited and translated by Nicolas Balacheff, Martin Cooper, Rosamund Sutherland and Virginia Warfield)

Cramer, J., Henriksen, L.A., Kunøe, M., Larsen, E.V., Togeby, O. & Widell, P. (1996). 699 varme termer. Leksikon til sprogkundskab. Aarhus Universitetsforlag.

Christensen, L.H. & Christensen, R. Z. (2016). *Dansk grammatik*. Syddansk Universitetsforlag. 3. udgave, 2. oplag

Conference Board of the Mathematical Sciences. (2001). *The mathematical education of teachers*. Providence, Rhode Island: American Mathematical Society.

Dale, T.C. og G. J. Cuevas, (1987). *Integrating Language and Mathematics Learning*. In: J. Crandall, red: ESL through Content-Area Instruction. Engelwood Cliffs, NJ: Prentice Hall Regents.

Danmarks Evalueringsinstitut. (2006). Gymnasiernes tiltag for tosprogede elever. Vester Kopi

Damarks Statistik. (2018). *Indvandrere og efterkommere i flertal på 13 gymnasier*. Located at dst.dk, <u>https://www.dst.dk/da/Statistik/bagtal/2018/2018-01-05-indvandrere-og-efterkommere-i-flertal-paa-13-gymnasier</u> visited 30/11-2020

Díaz, C. & Batanero, C. (2009). UNIVERSITY STUDENTS' KNOWLEDGE AND BIASES IN CON-DITIONAL PROBABILITY REASONING. International Electronic Journal of Mathematics Education, Vol. 4, No. 3, October 2009, p. 131-162

Dunn, P. K; Carey, M. D.; Richardson, A. M; Mcdonald, C. (2016). *Learning the language of statistics: Challenges and teaching approaches*. Statistics Education Research Journal, 15(1), p. 8-27

Durand-Guerrier, V., Kazima, M., Libbrecht, P., Ngansop, J. N., Salekhova, L., Tuktamyshov, N. & Winsløw, C. 2016. *Challenges and Oppoertunities for Second Language Learners in undergraduate Mathematics*. In Mathematics Education and Language Diversity - the 21<sup>st</sup> ICMI Study. Springer. p. 85-102

EMU. (2020). *Matematisk formelsamling STX B-niveau*. Located at EMU.DK, <u>https://emu.dk/si-tes/default/files/2019-02/Formelsamling-Matematik-B---stx-2018.pdf</u>. Visited 10/10-2020

Farrugia, M. T. (2018). *A functional linguistics analysis of a mathematics register expressed through two languages*. In N. Planas & Schütte, M. (Eds.), Proceedings of the Fourth ERME Topic Conference 'Classroom-based research on mathematics and language' (p. 57-64). Dresden, Germany: Technical University of Dresden / ERME.

Franklin, C. & Mewborn, D. (2006). *The statistical education of PreK-12 teachers: A shared responsibility*. In G. Burrill (Ed.), NCTM 2006 Yearbook: Thinking and reasoning with data and chance (pp. 335-344). Reston, V A: NCTM Gal, I. (2002). *Adult's statistical literacy. Meanings, components, responsibilities.* International Statistical Review, 70(1), p. 1-25.

Groth, R. E. (2010). *Teachers' Construction of Learning Environments for Conditional Probability and Independence*. In International Electronic Journal of Mathematics Education, Vol.5, No.1, p. 32-55

Holm, L. & Laursen, H. P. (2019). *Dansk som andetsprog - pædagogiske og didaktiske perspektiver*. Dansklærerforeningens Forlag. 1. udgave, 6. oplag

Høines, M. J. (1998): Begynneroplæringen. Bergen: Caspar forlag.

Jessen, B. & Winsløw, C. (2017). *The theory of Didactical Situations*. In Winsløw, C. (Ed.), Practical Guide to Inquiry Based Mathematics Ed., p. 29-42.

Jones, G., Langrall, C. & Mooney, E. (2007). *Research in probability: responding to classroom realities*. In F. Lester (Ed.), Second handbook of research on mathematics teaching and learning. Greenwich, CT: Information Age Publishing, Inc. & NCTM.

Kaplan, J., Fisher, D.G. & Rogness, N.T. (2010). *Lexical Ambiguity in Statistics: How students use and define the words: association, average, confidence, random and spread*. Journal of Statistics Education, Volume 18, Number 2. p. 1-22

Kazima, M. (2006). *Malawian students' meanings for probability vocabulary*. In Educational Studies in Mathematics (2006) 64. p. 164-189

Lockwood, E., Swinyard, C. A. & Caughman, J. S. (2015). *Patterns, Sets of Outcomes, and Combinatorial Justification: Two Students' Reinvention of Counting Formulas*. In Int. J. Res. Undergrad. Math. Ed. (2015). p. 27–62

Marthinus, K., Jensen, M., Pedersen, J. S., Pedersen, N. P. & Hansen, P. (2020). *MAT B htx*. Located at matbhtx.systime.dk, <u>https://matbhtx.systime.dk/?id=p1507</u>, visited 16/09-2020

Molnar, A. (2018). *Language and Lexical Ambiguity in the Probability Register*. In Teaching and Learning Stochastics - Advances in Probability Education Research, ICME-13 Monographs. Springer international Publishing, p. 23-38

Moore, D. S., Notz, W. I., & Fligner, M. A. (2013). *The basic practice of statistics* (6th ed.). New York: W. H. Freeman.

Moschkovich, J. (2003). *What counts as mathematical discourse*? In N. Pateman, B. Dougherty, & J. Zilliox (Eds.), Proceedings of the 2003 Joint Meeting of the International Group for Psychology of Mathematics Education (PME) and the North American Chapter of PME (PME-NA) (pp. 325-331). Honolulu, HI: CRDG, College of Education, University of Hawaii.

Ordnet. (2020). Spredning. Located at ordnet.dk, <u>https://ordnet.dk/ddo/ordbog?query=spredning</u> Visited 30/09-2020

Ordnet (2). (2020). Hyppighed. Located at ordnet.dk, <u>https://ordnet.dk/ddo/ordbog?query=hyppighed</u> Visited 30/09-2020

Ordnet (3). (2020). Sandsynlighed. Located at ordnet.dk, <u>https://ordnet.dk/ddo/ordbog?query=sand-</u> <u>synlighed</u> Visited 20/11-2020

Ordnet (4). 2020. Chance. Located at ordnet.dk, <u>https://ordnet.dk/ddo/ordbog?query=chance</u> Visited 20/11-2020

Russell, S. (1990). *Issues in training teachers to teach statistics in the elementary school: A world of uncertainty* In A. Hawkins (Ed.), Training teachers to teach statistics Proceedings of the International Statistical Institute Round Table Conference (pp. 59-71). Voorburg, Netherlands: International Statistical Institute.

Schleppegrell, M. J. (2001). *Linguistic features of the language of schooling*. Linguistics and Education, 12(4), p. 431–459.

Snow, C., & Uccelli, P. (2009). *The challenge of academic language*. In D. Olson & N. Torrance (Eds.), The Cambridge handbook of literacy. New York: Cambridge University Press.

Törner, G. & Arzarello, F. (2012). *Grading Mathematics Education Research Journals*. In European Mathematical Society Newsletter, December 2012

Undervisningsministeriet. (2019). Evaluering af de skriftlige prøver i matematik på stx og hf ved sommereksamen 2019, Delrapport 1 – ny ordning. Located at

https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwjKpZX5p

<u>PsAhXDh1wKHSGlCTAQFjADegQIBBAC&url=https%3A%2F%2Fwww.uvm.dk%2F-%2Fmedia%2Ffiler%2Fuvm%2Fudd%2Fgym%2Fpdf19%2Fnov%2F191119-evaluering-ma-tematik-stx-og-hf-rapport-2019-ny-ordning.pdf&usg=AOvVaw0teZjHJEPBoD8HDoSW9pR3</u> visited 01/12-2020

Winsløw, C. (2006). *Didaktiske elementer - en indføring i matematikkens og naturfagenes didaktik.* Forlaget Biofolia. 1. udgave, 4. oplag

Undervisningsministeriet. (2019). Evaluering af de skriftlige prøver i matematik på stx og hf ved sommereksamen 2019, Delrapport 1 - ny ordning. Located at

https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwjKpZX5p \_PsAhXDh1wKHSGlCTAQFjADegQIBBAC&url=https%3A%2F%2Fwww.uvm.dk%2F-%2Fmedia%2Ffiler%2Fuvm%2Fudd%2Fgym%2Fpdf19%2Fnov%2F191119-evaluering-matematik-stx-og-hf-rapport-2019-ny-ordning.pdf&usg=AOvVaw0teZjHJEPBoD8HDoSW9pR3 visited 01/12-2020

#### LIST OF FIGURES

Figure 1: An overview of the phases of TDS and the roles of the participants (Jessen & Win	ısløw,
2017, p. 41)	12
Figure 2: Overview of different meanings of probability (Batanero & Diaz, 2007, p. 117)	25
Figure 3: Selected correct bilingual student answers to 2C and 2D	49
Figure 4: Bilingual student replies to 4C and 4D	61
Figure 5: Translation of Problem 1 of worksheet 1	69
Figure 6: The translated problem from powerpoint slide no. 3 (appendix 5)	69
Figure 7: Problem 2 from worksheet 1 (appendix 2A/2B)	72
Figure 8: The sleeping problem from PowerPoint presentation no. 3, slide 3 (appendix 7)	74
Figure 9: Problem 1 from worksheet no. 3 (translated) (appendix 4a / 4b)	76
Figure 10: Part of the transcription of observation no. 1 (appendix 8)	80
Figure 11: Part of the transcription of observation no. 1 (appendix 8)	82
Figure 12: Part of the transcription of observation no. 1 (appendix 8)	84
Figure 13: Transcription of observation no. 2 (appendix 9)	86
Figure 14: The sleeping problem (appendix 7)	88
Figure 15: Part of the transcription of observation no. 3 (appendix 10)	89
Figure 16: Part of the transcription of observation no. 3 (appendix 10)	90
Figure 17: Problem 1 of worksheet 3 (apeendix 4a/4b)	92
Figure 18: Transcription of observation no. 4 (appendix 11)	93

LIST OF IMAGES

Image 1: Answers from BS <sub>1</sub> in relation to "spredning" (spread)	37
Image 2: Answers from BS2 in relation to "spredning" (spread)	38
Image 3: Answers from BS3 in relation to "spredning" (spread)	39
Image 4: Answers from MS1 in relation to "spredning" (spread)	41
Image 5: Answers from MS <sub>2</sub> in relation to "spredning" (spread)	42
Image 6: Answers from BS <sub>8</sub> in relation to "hyppighed" (frequency) part A and B	45
Image 7: Answers from BS <sub>8</sub> in relation to "hyppighed" (frequency) part C and D	46
Image 8: Answers from BS1 in relation to "hyppighed" (frequency) part A and B	47
Image 9: Answers from BS1 in relation to "hyppighed" (frequency) part C and D	48
Image 10: Answers from BS5 in relation to "hyppighed" (frequency) part C and D	48
Image 11: answers from MS <sub>1</sub> in relation to "hyppighed" (frequency) part A and B	50
Image 12: Answers from MS <sub>1</sub> in relation to "hyppighed" (frequency) part C and D	50
Image 13: Answers from MS <sub>2</sub> in relation to "hyppighed" (frequency) part A and B	51
Image 14: Answers from MS <sub>2</sub> in relation to "hyppighed" (frequency) part C and D	51
Image 15: Answers from BS <sub>6</sub> in relation to "frekvens" (relative frequency)	54
Image 16: Answers from BS9 in relation to "frekvens" (relatie frequency)	55
Image 17: answers from BS <sub>3</sub> in relation to "frekvens" (relative frequency)	56
Image 18: Answers from MS <sub>1</sub> in relation to "frekvens" (relative frequency)	57
Image 19: Answers from MS <sub>2</sub> in relation to "frekvens" (relative frequency)	58
Image 20: Answers from MS <sub>2</sub> in relation to "uafhængighed" (independence)	62
Image 21: Screenshot of how to acces the excel sheet in relation to the binomial distribution in	n word
Mat	67
Image 22: Screenshot of the excel sheet provided by WordMat	67
Image 23: CAS-solution of problem 1, worksheet 3	77
Image 24: CAS-solution of problem 1, worksheet 3	78
Image 25: Written explanation by the teacher of "stochastic variable" on the blackboard	85
Image 26: A probability table describing the number of minutes the bus is delayed	85
Image 27: Solution made by Word Mat to problen 1a from worksheet 3	94

#### APPENDIX

#### APPENDIX 1A: LEXICAL AMBIGUITY WITHIN STATISTICS

Navn og klasse:

Hvilket sprog taler I hjemme ved dine forældre?

- 1. Spredning
  - A. Skriv en sætning hvor ordet "spredning" indgår med samme betydning, som du bruger det i til daglig
  - B. Giv en definition af hvad ordet "spredning" betyder, når det bruges som du gjorde i svaret på A
  - C. Skriv en sætning hvor ordet "spredning" indgår med den betydning du kender fra gymnasiets matematik
  - D. Giv en definition af ordet "spredning" sådan som det bruges i svaret på C
- 2. Hyppighed
  - A. Skriv en sætning hvor ordet "hyppighed" indgår med samme betydning, som du bruger det i til daglig
  - B. Giv en definition af hvad ordet "hyppighed" betyder, når det bruges som du gjorde i svaret på A
  - C. Skriv en sætning hvor ordet "hyppighed" indgår med den betydning du kender fra gymnasiets matematik
  - D. Giv en definition af ordet "hyppighed" sådan som det bruges i svaret på C
- 3. Frekvens
  - A. Skriv en sætning hvor ordet "frekvens" indgår med samme betydning, som du bruger det i til daglig
  - B. Giv en definition af hvad ordet " frekvens" betyder, når det bruges som du gjorde i svaret på A
  - C. Skriv en sætning hvor ordet " frekvens" indgår med den betydning du kender fra gymnasiets matematik
  - D. Giv en definition af ordet " frekvens" sådan som det bruges i svaret på C
- 4. Uafhængig
  - A. Skriv en sætning hvor ordet "uafhængig" indgår med samme betydning, som du bruger det i til daglig
  - B. Giv en definition af hvad ordet "uafhængig" betyder, når det bruges som du gjorde i svaret på A
  - C. Skriv en sætning hvor ordet "uafhængig" indgår med den betydning du kender fra gymnasiets matematik
  - D. Giv en definition af ordet "uafhængig" sådan som det bruges i svaret på C

#### APPENDIX 1B: LEXICAL AMBIGUITY WITHIN STATISTICS ENGLISH TRANSLATION

Name and class:

What language do you speak at home with your parents?

- 1. Spredning (spread)
  - A. Write a sentence with the word "spredning" with the meaning of everyday langauge.
  - B. Provide a definition of the word "spredning" maintaining the meaning of A
  - C. Write a sentence with the word "spredning" with the meaning of mathematics as you know it from High School
  - D. Provide a definition of the word "spredning" maintaining the meaning of C
- 2. Hyppighed (frequency)
  - A. Write a sentence with the word "hyppighed" with the meaning of everyday langauge.
  - B. Provide a definition of the word "hyppighed" maintaining the meaning of A
  - C. Write a sentence with the word "hyppighed" with the meaning of mathematics as you know it from High School
  - D. Provide a definition of the word "hyppighed" maintaining the meaning of C
- 3. Frekvens (frequence)
  - A. Write a sentence with the word "frekvens" with the meaning of everyday langauge.
  - B. Provide a definition of the word "frekvens" maintaining the meaning of A
  - C. Write a sentence with the word "frekvens" with the meaning of mathematics as you know it from High School
  - D. Provide a definition of the word "frekvens" maintaining the meaning of C
- 4. Uafhængig (independent)
  - A. Write a sentence with the word "uafhængig" with the meaning of everyday langauge.
  - B. Provide a definition of the word "uafhængig" maintaining the meaning of A
  - C. Write a sentence with the word "uafhængig" with the meaning of mathematics as you know it from High School
  - D. Provide a definition of the word "uafhængig" maintaining the meaning of C

### APPENDIX 2A: WORKSHEET 1 DANISH

# Arbejdsseddel 1: Binomialkoefficienter

#### **Opgave 1**

På et volleyballhold er der tilmeldt syv spillere. Der er altid seks på banen af gangen.

- a) På hvor mange forskellige måder kan man vælge seks spillere til at starte på banen?
- b) Hvor mange forskellige måder kan man sætte et starthold på, hvis der havde været tilmeldt otte spillere?

#### **Opgave 2**

- a) Bestem K(3, 2)
- b) Bestem K(5,3)
- c) Bestem K(7, 2)
- d) Bestem K(7,5)

Sammenlign dine svar i Opgave 2c og 2d.

# Arbejdsseddel 2:

#### **Opgave 3**

Betragt nedenstående sandsynlighedstabel for vejret i morgen. Hvad er sandsynligheden for, at der bliver tåge?

Udfald	Solskin	Regnvejr	Overskyet	Tåge
Sandsynlighed	5 %	70 %	10 %	

#### Opgave 4

Theodor, Fabian og Pernilla slår med en tisidet terning med tal fra 0 til 9. De aftaler, at hvis terningen viser 0, 1 eller 2 vinder Theodor, hvis den viser 3, 4, 5 eller 6 vinder Fabian, og hvis den viser 7, 8 eller 9 vinder Pernilla.



a) Udfyld nedenstående sandsynlighedstabel:

Udfald	Theodor vin-	Fabian vinder	Pernilla vinder
	der		

Sandsynlighed	3	4 2	3
	10	$\frac{10}{10} = \frac{1}{5}$	10

- b) Forklar, hvorfor ovenstående tabel viser, at der ikke er tale om et symmetrisk sandsynlighedsfelt.
- c) Hvad kunne Theodor, Fabian Pernilla gøre for at ændre deres spil, så sandsynlighedsfeltet bliver symmetrisk?

Til de nedenstående opgaver får du brug for formel (178) i formelsamlingen.

#### **Opgave 5**

Theodor, Fabian og Pernilla slår stadig med terningen fra Opgave 4, men har nu også fået fat i en helt almindelig sekssidet terning.

- a) Hvad er sandsynligheden for, at de slår en 7'er med den tisidede terning og en 5'er med den sekssidede terning?
- b) Hvad er sandsynligheden for, at de slår mere end 6 med den tisidede terning og mindre end 3 med den sekssidede?

Allerede færdig? Godt gået! Gå ind på ABaCus.dk (du kan finde koden på skemabrikken), og arbejd på opgaverne dér.

#### APPENDIX 2B: WORKSHEET 1 ENGLISH TRANSLATION

#### Problem 1

At a volleyball team there are seven players signed up. There is always six people at the field at once.

- a) In how many ways can one choose six players to start at the field?
- b) In how many different ways can one make a starting team if there had been eight players?

#### Problem 2

- a) Calculate K(3, 2)
- b) Calculate K(5,3)
- c) CalculateK(7, 2)
- d) Calculate K(7, 5)

Compare your answers from problem 2b and 2c.

#### Worksheet 2:

#### Problem 3

Consider the following table of probabilities for the weather tomorrow. What is the probability for it to be foggy?

Outcome	Sunshine	Rain	Cloudy	Foggy
Probability	5 %	70 %	10 %	

#### Problem 4

Theodor, Fabian, and Pernilla is throwing a ten sided dice with numbers from 0 to 9. They agree that if the dice show 0, 1 or 2, Theodor wins, if it shows 3, 4, 5, or 6, Fabian wins, and if it shows 7, 8, or 9, Pernilla wins.

a) Fill put the following table of probability:

Outcome	Theodor wins	Fabian wins	Pernilla wins
Probability			

- b) Explain why the above tabel shows that this is not a symmetrical field of probability.
- c) What could Theodor, Fabian and Pernilla do to change their game in order to make the field of probability symmetrical

For the problems below you will need formula (178) of the collection of formulas

#### Problem 5

Theodor, Fabian and Pernilla are still rolling the dice from Task 4, but have now also got hold of an ordinary six-sided dice.

- d) What is the probability that they hit a 7 with the ten-sided dice and a 5 with the six-sided dice?
- e) What is the probability that they hit more than 6 with the ten-sided dice and less than 3 with the six-sided?

Already done? Well done! Go to ABaCus.dk (you can find the code on the form factory), and work on the tasks there.

#### APPENDIX 3A: WORKSHEET 2 DANISH

# Arbejdsseddel om stokastiske variable, middelværdi og spredning

(hvis du går i stå, så brug det øverste link på skemabrikken)

#### Opgave 1:

En stokastisk variabel, X, har følgende sandsynlighedstabel:

$X = x_i$	1	2	3
$P(X = x_i)$	0,4	0,5	

- a. Bestem P(X = 3).
- b. Bestem middelværdien for *X*.
- c. Bestem spredningen for *X*.

#### **Opgave 2**

Lad Y være den stokastiske variabel for antallet af øjne for en helt almindelig, sekssidet terning.

$Y = y_i$	1	2	3	4	5	6
$P(Y = y_i)$	1	1	1	1	1	1
	6	6	6	6	6	6

- a. Bestem middelværdien for *X*
- b. Bestem spredningen for *X*

#### Opgave 3

En stokastisk variabel, Y, har følgende sandsynlighedstabel:

$Y = y_i$	1	5	x
$P(Y = y_i)$	0,1	0,4	0,5

- a) Bestem middelværdien for *Y* udtrykt ved *x*.
- b) Bestem den manglende værdi, x, således at middelværdien for Y er 8.

#### **Opgave 4**

Når DOT skal beslutte, hvor de skal lægge bøden for at køre uden billet i metroen, er de nødt til at se situationen fra en snyders perspektiv. Hvis vi antager, at:

- Sandsynligheden for at blive grebet i at køre uden billet er 10 %
- At en billet koster 16 kroner
- En bøde for at køre uden billet er på 750 kroner

kan vi opstille denne sandsynlighedstabel for den stokastiske variabel, Z, der betegner, hvor meget man sparer ved at snyde i metroen:

$Z = z_i$	16	-734
P(Z = z)		0,1

- a. Bestem P(Z = 16).
- b. Bestem middelværdien *Z*, og forklar hvorfor tallet viser, at det i gennemsnit ikke kan betale sig at snyde.
- c. (Svær) Bestem, hvor langt bøden skal ned, for at det pludselig begynder at kunne betale sig at snyde, hvis sandsynligheden for at blive grebet uden billet er 10 %.
- d. (Svær) Bestem, hvor langt sandsynligheden for at blive grebet i at snyde skal ned, for at det pludselig begynder at kunne betale sig at snyde, hvis bøden for at blive grebet uden billet er 750 kroner.

#### APPENDIX 3B: WORKSHEET 2 ENGLISH TRANSLATION

# Worksheet on stochastic variables, the mean and the spread *(if you get stuck, use the top link on the schedule)*

#### Problem 1:

A stochastic variable X, has the following table of probabilities:

$X = x_i$	1	2	3
$P(X = x_i)$	0,4	0,5	

- a. Determine P(X = 3).
- b. Determine the mean for *X*.
- c. Determine the spread for *X*.

#### Problem 2:

Let Y be the stochastic variable describing the number of eyes on a regular six sided cube

$Y = y_i$	1	2	3	4	5	6
$P(Y = y_i)$	1	1	1	1	1	1
	6	6	6	6	6	6

- a. Determine the mean for *Y*
- b. Determine the spread for Y

#### Opgave

A stochastic variable *Y* has the following table of probabilities:

$Y = y_i$	1	5	x
$P(Y = y_i)$	0,1	0,4	0,5

- a. Determine the mean of Y expressed by x.
- b. Determine the missing value, x, such that the mean of Y is 8.

#### **Opgave 4**

When DOT has to decide the price for a fine in the subway, the have to view the situation from a cheaters perspective. If we assume that:

- The probability of being caught driving without a ticket is 10%
- That a ticket costs 16 kroner
- A fine for driving without a ticket is 750 kroner

we can set up this probability table for the stochastic variable, Z, which denotes how much you save by cheating in the metro:

$Z = z_i$	16	-734
P(Z = z)		0,1

- a. Determine P(Z = 16).
- b. Determine the mean of Z, and explain why the figure shows that on average it does not pay to cheat.
- c. (Difficult) Determine how far the fine must go down in order for it to suddenly start to pay to cheat if the probability of being caught without a ticket is 10%.
- d. (Difficult) Decide how far the probability of being caught cheating must go down, so that it suddenly starts to pay to cheat if the fine for being caught without a ticket is 750 kroner.

#### APPENDIX 4A: WORKSHEET 3 DANISH

# Arbejdsseddel den 21. oktober - binomialfordeling

#### **Opgave 1 (med hjælpemidler)**

X er en binomialfordelt stokastisk variabel med antalsparameter 80 og sandsynlighedsparameter p = 0,35.

- a) Tegn et søjlediagram for sandsynlighedsfordelingen for X.
- b) Bestem middelværdien og spredningen for *X*.
- c) Bestem P(X = 28).

#### **Opgave 2 (med hjælpemidler)**

En stokastisk variabel Y er binomialfordelt med antalsparameter 30 og sandsynlighedsparameter 0,43.

- a) Bestem P(Y = 12)
- b) Bestem  $P(Y \le 8)$

#### Opgave 3 (med hjælpemidler)

Theodor ringer rundt og spørger 400 tilfældigt udvalgte vælgere i amerikanske delstat New Hampshire, om de har tænkt sig at stemme på Donald Trump til præsidentvalget.

Antallet af Trump-vælgere, Theodor snakker med, er en binomialfordelt stokastisk variabel Z med sandsynlighedsparameteren 0,50.

- a) Bestem sandsynligheden for, at Theodor snakker med netop 200 Trump-vælgere.
- b) Bestem sandsynligheden for, at Theodor snakker med højest 200 Trump-vælgere.
- c) Bestem sandsynligheden for, at Theodor snakker med netop 184 Trump-vælgere.
- d) Bestem sandsynligheden for, at Theodor snakker med højest 184 Trump-vælgere.

Det viser sig, at 175 ud af de 400 vælgere, som Theodor har snakker med, har tænkt sig at stemme på Donald Trump.

e) Kommentér sandsynligheden for dette udfald. Inddrag sandsynlighedsparameteren i dit svar.

# APPENDIX 4B: WORKSHEET 3 ENGLISH TRANSLATION Worksheet for October 21st - binomial distribution

#### Problem 1 (aids allowed)

X is a binomially distributed stochastic variable with number parameter 80 and probability parameter p = 0,35.

- f) Make a bar chart of the probability distribution for *X*.
- g) Determine the mean ands spread/standard deviation of *X*.
- h) Dertermine P(X = 28).

#### Problem 2 (aids allowed)

A stochastic variable Y is binomially distributed with number parameter 30 and probability parameter 0,43

- a) Determine P(Y = 12)
- b) Determine  $P(Y \le 8)$

#### Problem 3 (aids allowed)

Theodor is calling 400 randomly selected voters in the American state New Hampshire, and aks if they are going to vote for Donald Trump at the presidential elections.

The amount of Trump voters, Theodor talks to, is a binomially distributed stochastic variable Z with a probability parameter of 0,50.

- a. Determine the probability of Theodor talking to exactly 200 Trump voters
- b. Determine the probability of Theodor talking to at most 200 Trump voters
- c. Determine the probability of Theodor talking to exactly 184 Trump voters
- d. Determine the probability of Theodor talking to at most 184 Trump voters

It turns out that 175 of the 400 voters, Theodor has spoken to are going to vote for Donald Trump.

e. Comment on the probability for this outcome. Include the probability parameter in your answer.

#### Appendix 5: PowerPoint presentation from module 1

Dagsorden • Et par opvarmningsopgaver • Arbejdsseddel: Binomialkoefficienten • Opsamling • Case: Theodor vrøvler! • Lidt nyt: Symmetriske sandsynlighedsfelter • Mere arbejdsseddel • Mere opsamling 1













Sandsynlighed for forskellige hændelser Udfald Solskin Regnvejr **Overskyet** Tåge Sandsynlighed 5 % 70 % 10 % 15 % • Hvad er sandsynligheden for, at der bliver enten regn eller tåge? Udfald Plat Krone Sandsynlighed 50 % 50 % • Hvad er sandsynligheden for, at der bliver regn, og at vi slår plat med mønt? Sandsynlighed ved (178) $P(\text{både } A \text{ og } B) = P(A) \cdot P(B)$ kombination af uafhængige hændelser A og B Fortsæt med arbejdssedlen Sandsynlighed ved (179)P(A eller B) = P(A) + P(B)kombination af hændelser A og B, som ikke har noget fælles udfald

#### Appendix 6: PowerPoint presentation from module 2

Dagsorden • Et par opvarmningsopgaver • Lidt om prøven på onsdag • Noget nyt: Stokastiske variable • Case: Theodor kommer måske for sent! Middelværdi og spredning Arbejdsseddel • Opsamling 1



Opvarmn I Theodors klasse skal fremlægge. S	ingsopg skal de lave fre andsynligheder	aver: S mlæggelser ne kan beste	nak med samfundsfag. D emmes ud fra fø	d sidemakke et afgøres ved lodtræk gende sandsynlighedst	ren ning, hvem der abel:
Udfald	Theodor skal fremlægge	Fabian skal fremlægge	Pernilla skal fremlægge	Hverken Theodor, Fabia fremlægge	an eller Pernilla skal
Sandsynlighed	4 %	4 %	4 %	p	)
<ol> <li>Bestem sands</li> <li>Der er en 5 % san</li> <li>Bestem sandsy</li> </ol>	synligheden for, dsynlighed for, a nligheden for, a	at enten The at læreren e t læreren rei	eodor eller Fabia r syg, og module nt faktisk er rask	n skal fremlægge. t bliver aflyst. , <u>og</u> det er Theodor, de	r skal fremlægge.
Sandsynlighed ved kombination af uafh hændelser A og B	(i nængige	178) <i>P</i> (båd	e A og B) = P(A)	P(B)	
Sandsynlighed ved kombination af hæn og <i>B</i> , som ikke har i fælles udfald	(delser A noget	179) P(A el	$\operatorname{ler} B) = P(A) + F$	( <i>B</i> )	



Tabellen viser værdierne for en stokastisk variabel X sammen med nogle af de tilhørende sandsynligheder.

$X = x_i$	3	7	11	20
$P(X=x_i)$	0,2	0,125	0,5	

a) Bestem P(X=20).





Middelværdi af en stokastisk variabel X		(182) $\mu = E(X) = \sum_{i=1}^{n} x_i \cdot P(X = x_i)$ $= x_1 \cdot p_1 + x_2 \cdot p_2 + x_3 \cdot p_3 + \dots + x_n \cdot p_n$		
Bussen kommer til tiden	Bussen er et minut forsinket	Bussen er to minutter forsinket	Bussen er 5 minutter forsin	iket
0,45	0,25	0,2	0,1	
	okastisk Bussen kommer til tiden 0,45	okastisk (182) $\mu =$ = Bussen kommer til tiden 0,45 0,25	okastisk(182) $\mu = E(X) = \sum_{i=1}^{n} x_i \cdot X_i = x_1 \cdot p_1 + x_2 \cdot p_2 = 0$ Bussen kommer till tidenBussen er et minut forsinketBussen er to minutter forsinket0,450,250,2	okastisk(182) $\mu = E(X) = \sum_{i=1}^{n} x_i \cdot P(X = x_i)$ $= x_1 \cdot p_1 + x_2 \cdot p_2 + x_3 \cdot p_3 + \dots + x_n \cdot p_n$ Bussen kommer til tidenBussen er et minut forsinketBussen er to minutter forsinketBussen er 5 minutter forsin to 0,250,450,250,20,1

Middelværdi af en stokastisk variabel X

(182) 
$$\mu = E(X) = \sum_{i=1}^{n} x_i \cdot P(X = x_i)$$
$$= x_1 \cdot p_1 + x_2 \cdot p_2 + x_3 \cdot p_3 + \dots + x_n \cdot p_n$$



9

Middelværdi af en stokastisk variabel X  $(182) \quad \mu = E(X) = \sum_{i=1}^{n} x_i \cdot P(X = x_i)$   $= x_1 \cdot p_1 + x_2 \cdot p_2 + x_3 \cdot p_3 + \dots + x_n \cdot p_n$ Varians af en stokastisk variabel X  $(183) \quad \operatorname{Var}(X) = \sum_{i=1}^{n} (x_i - \mu)^2 \cdot P(X = x_i)$   $= (x_1 - \mu)^2 \cdot p_1 + \dots + (x_n - \mu)^2 \cdot p_n$ Spredning af en stokastisk variabel X  $(184) \quad \sigma = \sigma(X) = \sqrt{\operatorname{Var}(X)}$ 



#### APPENDIX 7: POWERPOINT PRESENTATION FROM MODULE 3

Dagsorden • Et par opvarmningsopgaver: Binomialfordeling • Et kig på Aflevering 4 Binomialfordelingens formel Arbejdsseddel Opsamling Binomialfordelingen med hjælpemidler Arbejdsseddel Opsamling 1



















#### APPENDIX 8: TRANSCRIPTION OF OBSERVATION NO. 1

BS<sub>3</sub>: Vent hvad er .. / Wait what is..

BS<sub>1</sub>: Men der er syv spillere der er tilmeldt .. / But there are seven players assigned..

BS<sub>2</sub>: Har I lavet opgave et? / Did you finish problem one?

BS<sub>3</sub>: Jeg tror det er sådan her.. syv parentes gange fem gange fire gange tre gange to gange en. Jeg tror, jeg ved det ikke, det er sådan det er mig der har svært ved at sige det. / *I think it is like this*.. Seven parenthesis times five times four times three times two times one. I think, I don't know, it is like this, it's me who is having difficulties saying it.

BS1: Er det for sjov eller siger du din mening? / Are you kidding or do you really mean this?

BS<sub>2</sub>: Ved du hvad tallet bliver, det bliver sådan femten millioner / *Do you know what the number will be, it will be like fifteen millions* 

BS3: Det bliver bare fem tusind / It will just be five thousand

BS<sub>2</sub>: Ja som om det er meget bedre. / Yeah like that is much better

BS<sub>1</sub>: Neeeej.. / Nooo..

BS<sub>2</sub>: Der kan ikke være fem tusind forskellige muligheder for syv spillere. Det passer ikke. / *There can not be five thousand different options for seven players. That is not true.* 

T: I får brug for den her formel til opgaverne (skriver formlen på tavlen). Det er formel 169 i formelsamlingen / You will need this formula for the problems (writes the formula on the blackboard). It is formula 169 in the book of formulas.

BS4: Ej det her det kan ikke være rigtigt der er noget galt. / This can not be true something is wrong

BS<sub>2</sub>: Ja ikke? Der er et eller andet galt / Yeah right? Something is wrong

BS3: Altså vi kan lave den næste opgave / Well we can do the next problem

 $BS_1$ ,  $BS_3$  and  $BS_4$  continues to work on problem 2 instead which is going well. The students solve the problems quietly and without any obvious problems.  $BS_2$  looks at the formula 169, and the turns to another student.
BS<sub>2</sub>: Kig kig kig. Den her formel. Jeg forstår den ikke. Prøv og kig hundrede ni og tres (peger på formlen i bogen)/ Look look. This formula. I don't get it. Have a look one hundred and sixty nine (points to the formula)

BS<sub>5</sub>: Man siger der er syv spillere ikke? Og der er 6 spillere. Så siger du 7 gange 6 gange 5 gange .../ You say there are seven players right? And there are six players. So you say seven times six times five times ...

BS<sub>2</sub>: Det er det hun lavede og der gav det fem tusind et eller andet / *That is what she did and then it became five thousand something* 

BS5: Præcis / Exactly

A sixth student asks BS<sub>5</sub> why you need to multiply the numbers and BS<sub>5</sub> explains this. Then BS<sub>2</sub> asks BS<sub>3</sub> again:

S<sub>2</sub>: Det giver ikke mening med det her. Svaret kan ikke give fem tusind. Der kan ikke være fem tusind forskellige måder at komme syv spillere ind på en bane / *It does not make sense with this. The answer can not be five thousand. There can not be five thousand differint ways to put seven players into a field.* 

S<sub>3</sub>: Jo / Yes

S<sub>2</sub>: Nej, tror I der er fire hundrede spillere eller hvad? / *No do you think there are four hundred players or what*?

BS3: Nej, det er der ikke. Nej fordi at .. / No there is not. No because ..

BS2: Ja så giver det syv. / Yes so it is seven

BS<sub>3</sub>: Nej fordi ... Nej.. Det kan ikke være syv. Der er seks.. / *No because ... No. It can not be seven. There are six ..* 

BS<sub>2</sub>: Ja så giver det syv. Der er en der er ude ad gangen. / Yes so it is seven. There is one person who is out at the time.

BS<sub>3</sub>: Nej, fordi enhver af de der.. Nej, fordi på forskellige måder.. / *No because each of these ... No because in different ways*...

BS2: Der kan kun være syv forskellige måder / There can only be seven different ways

 $BS_1$ ,  $BS_3$  and  $BS_4$  looks at the assignment again, point to the text and mumble.  $BS_2$  then seeks assistance from the teacher who walks by:

BS<sub>2</sub>: T er der ikke kun syv forskellige måder? / T isn't there only seven different ways?

T: Hvorfor det? / Why is that?

BS<sub>2</sub>: Fordi øh (griner). Du kan kun have nummer et ude én gang. Du kan kun have nummer to ude én gang. Du kan kun.. / Because eh (laughs). You can only have number one out one time. You can only have number two out one time. You can only have ..

T: Vi kan jo godt vælge en vi sætter ude. Det giver glimrende mening. / *We can choose one whom we set outside. That makes excellent sense.* 

## APPENDIX 9: TRANSCRIPTION OF OBSERVATION NO. 2

BS: Hvis det er tilfældigt hvorfor er det så ikke 25% chance for at .../ *If it is random why isn't there a 25% chance to.*.

T: Det er fordi det ikke er et symmetrisk sandsynlighedsfelt / *That is because this is not a symmetrical probability field* 

S: Jeg forstår bare ikke hvordan der kan være større chance for at der sker noget som er tilfældigt / *I just don't understand how there can be a greater chance of something happening which is random* 

T: Det er en meget meget øhm.. Det er en meget mærkelig terning vi er ude i. Det var også ligesom før, sandsynligheden for at læreren var syg og læreren var rask var ikke det samme. Selvom noget er tilfældigt kan det godt ske, så der er noget der er mere sandsynligt end noget andet. Sandsynligheden for at vinde i lotteri og for ikke at vinde i lotteri vil også sjældent være det samme / *It is a very uhm.*. *It's a very weird dice we have. It is like before, the probability of the teacher being sick and the teacher being healthy was not the same. Even though something is random it can happen that something is more likely that something else. The probability of winning the lottery and not winning the lottery will rarely be the same.* 

BS: Nåh ja ja / Yeah right

T: Så det er egentlig det at der er nogle forskellige sandsynligheder og hvis vi skulle gætte, hvis bare vi skulle komme med et bud på, hvad tror vi at den her variabel bliver. Så vil vi jo gætte på det er 11 for det er der jo trods alt 50% sandsynlighed for. / So, it is really the fact that there are different probabilities and if we had to guess, just give an estimate on what we believe this variable will be. Then we will guess on 11 because after all there is a 50% probability of this.

BS: Hvorfor er der det? Det er det jeg ikke forstår / Why is that? That is what I don't understand

T: Og det er det.. Det er det der er lidt svært ved det her for det er der ikke noget forklaring på. / *And this is it.. This is what is a little difficult about this as there is no explanation to this.* 

Another student suggests an example in the context of gambling on soccer games and the number of corner kicks and another conversation about this begins.

## APPENDIX 10: TRANSCRIPTION OF OBSERVATION NO. 3

T: Så skulle vi bestemme middelværdien og der har jeg jo været så venlig at give jer en formel, nemlig formel 189 der kan hjælpe jer med der her. / *Now we were to determine the mean and for this I have been so kind as to give you a formula, formula 189, to help you with this* 

BS<sub>1</sub>: Øhm vi skal gange tyve med nul komma nul fem. / *Ehm we should multiply twenty with zero point zero five* 

T: Ja vi skal. Hvor ved du det fra? / Yes we should. How do you know?

BS1: Du skal bruge formel 189 i den der / You are to use formula 189 in the ...

T: Lige præcis, formlen fortæller os at middelværdien, det der græske bogstav der /*Exactly the for*mula tells us that the mean, the greek letter there

 $S_2: my \ / \ mu$ 

T: Det hedder nemlig my ja det var mit næste spørgsmål. Det er antalsparameteren gange sandssynlighedsparameteren så det må jo bare være 20 .. gange .. 0,05. Og hvad giver det? / *It is called mu yes that was my next question. That is the number parameter times the probability parameter so it just has to be twenty ... times .. zero point zero five. And what is this?* 

BS<sub>1</sub>: En / One

T: Ja det giver en. Lige præcis. Og hvordan skal det her tal så fortolkes, det er det næste. Nu ved vi det her, men hvad fortæller det os så om? / Yes it is one. Exactly. And how is this number to be interpreted, that is the next part. Now we know this, but what does it tell us?

BS<sub>3</sub>: Er det ikke bare at der en der har mulighed ... At der er en der kommer til at sove over sig. / *Is it not just that there is one who has the possibility* .. *That there is one who is going to oversleep* 

T: Der er en der kommer til at sove over sig? Ved vi det? At der er en der kommer til at sove over sig? / *There is one who is going to oversleep? Do we know this? That there is one who is going to oversleep?* 

BS<sub>3</sub>: Nej men sandsynligheden er (stopper med at tale)/ No but the probability is (stops talking)

T: Sandsynligheden er at der er en der kommer til at sove over sig? Vi nærmer os. Vi er tæt på. / *The probability is that there is one who is going to oversleep? We are getting closer. We are close.* 

BS<sub>4</sub>: Jeg er ikke sikker men det er en ud af 20, der sover over sig, er det ikke? / *I am not sure but it is one out of twenty who oversleeps is it not*?

T: Jamen problemet er at det kan vi jo ikke vide. Det her er tilfældigt. Der er bare hver af dem der har en sandsynlighed på 0,05 for at sove over sig. Denne her variable den kan ende i alle mulige [] det eneste vi ved er at den lander et sted mellem 0 og 20. Der er et sted mellem nul og tyve elever der sover over sig. / Well problem is that we can not know this. This is random. It is simply just each of them who has a probability of 0,05 to oversleep. This variable it can end in all sorts of [] the only thing we know is that it will end somewhere between zero and twenty. There is somewhere between zero and twenty students who oversleeps.

S<sub>5</sub>: Middelværdien siger at der gennemsnitligt vil være en der sover over sig / *The mean is saying that averagely one will oversleep* 

T: Der vil gennemsnitligt være en der sover over sig. Vi kan ikke vide om der er en der kommer til at sove over sig, men der vil gennemsnitligt være en der sover over sig. / Averagely there will be one who oversleeps. We can not know if there will be one who oversleeps but averagely there will be one who oversleeps.

## APPENDIX 11: TRANSCRIPTION OF OBSERVATION NO. 4

BS<sub>1</sub>: Hallo kan vi ikke lave noget, wallah, jeg forstår det ikke / *Hallo can we not get started, wallah, I don't understand* 

BS<sub>2</sub>: Hvad er den? / What is it?

BS<sub>1</sub>: Jeg kan ikke finde ud af (stopper med at tale) / I don't know how to (stops talking)

Eleverne taler om andre ting / The students discuss other things

BS<sub>2</sub>: Hvordan gør vi det? Okay vent. Skal man have wordmat? / *How do we do it? Okay wait. Do I need wordmat*?

BS<sub>3</sub>: Må jeg lige se? (vender computeren og åbner op for Excelarket). Det er det her ikke. Det er bare det der vi lavede / *Can I see*? (*Turns the computer around and opens the Excel sheet*). It is this one right? It is just what we did.

BS1: Så jeg skal skrive 80 og det andet? / So I have to write 80 and the other one?

BS<sub>3</sub>: Ja (peger på felterne) / Yes (points to the boxes)

BS<sub>1</sub>: Og så tager jeg et billede af den? (Tager et screenshot og indsætter i dokumentet) / And then I take a picture of it? (Makes a screenshot and inserts in the document)